Chapter 6

Expected Value and Variance

6.1 Expected Value of Discrete Random Variables

When a large collection of numbers is assembled, as in a census, we are usually interested not in the individual numbers, but rather in certain descriptive quantities such as the average or the median. In general, the same is true for the probability distribution of a numerically-valued random variable. In this and in the next section, we shall discuss two such descriptive quantities: the *expected value* and the *variance.* Both of these quantities apply only to numerically-valued random variables, and so we assume, in these sections, that all random variables have numerical values. To give some intuitive justification for our definition, we consider the following game.

Average Value

A die is rolled. If an odd number turns up, we win an amount equal to this number; if an even number turns up, we lose an amount equal to this number. For example, if a two turns up we lose 2, and if a three comes up we win 3. We want to decide if this is a reasonable game to play. We first try simulation. The program **Die** carries out this simulation.

The program prints the frequency and the relative frequency with which each outcome occurs. It also calculates the average winnings. We have run the program twice. The results are shown in Table 6.1.

In the first run we have played the game 100 times. In this run our average gain is $-.57$. It looks as if the game is unfavorable, and we wonder how unfavorable it really is. To get a better idea, we have played the game 10,000 times. In this case our average gain is $-.4949$.

We note that the relative frequency of each of the six possible outcomes is quite close to the probability $1/6$ for this outcome. This corresponds to our frequency interpretation of probability. It also suggests that for very large numbers of plays, our average gain should be

$$\mu = 1\Big(\frac{1}{6}\Big) - 2\Big(\frac{1}{6}\Big) + 3\Big(\frac{1}{6}\Big) - 4\Big(\frac{1}{6}\Big) + 5\Big(\frac{1}{6}\Big) - 6\Big(\frac{1}{6}\Big)$$

Winning	n = 100		n = 10000	
	Frequency	Relative Frequency	Frequency	Relative Frequency
1	17	.17	1681	.1681
-2	17	.17	1678	.1678
3	16	.16	1626	.1626
-4	18	.18	1696	.1696
5	16	.16	1686	.1686
-6	16	.16	1633	.1633

Table 6.1: Frequencies for dice game.

$$= \frac{9}{6} - \frac{12}{6} = -\frac{3}{6} = -.5 \ .$$

This agrees quite well with our average gain for 10,000 plays.

We note that the value we have chosen for the average gain is obtained by taking the possible outcomes, multiplying by the probability, and adding the results. This suggests the following definition for the expected outcome of an experiment.

Expected Value

Definition 6.1 Let X be a numerically-valued discrete random variable with sample space Ω and distribution function $m(x)$. The *expected value* $E(X)$ is defined by

$$E(X) = \sum_{x \in \Omega} x m(x) \ ,$$

provided this sum converges absolutely. We often refer to the expected value as the *mean,* and denote $E(X)$ by μ for short. If the above sum does not converge absolutely, then we say that X does not have an expected value. □

Example 6.1 Let an experiment consist of tossing a fair coin three times. Let X denote the number of heads which appear. Then the possible values of X are $0, 1, 2$ and 3. The corresponding probabilities are $1/8, 3/8, 3/8,$ and $1/8$. Thus, the expected value of X equals

$$0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{3}{2} \ .$$

Later in this section we shall see a quicker way to compute this expected value, based on the fact that X can be written as a sum of simpler random variables. □

Example 6.2 Suppose that we toss a fair coin until a head first comes up, and let X represent the number of tosses which were made. Then the possible values of X are $1, 2, \ldots,$ and the distribution function of X is defined by

$$m(i) = \frac{1}{2^i} \ .$$

(This is just the geometric distribution with parameter 1/2.) Thus, we have

$$\begin{aligned} E(X) &= \sum_{i=1}^{\infty} i \frac{1}{2^i} \\ &= \sum_{i=1}^{\infty} \frac{1}{2^i} + \sum_{i=2}^{\infty} \frac{1}{2^i} + \cdots \\ &= 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots \\ &= 2 \ . \end{aligned}$$

□

Example 6.3 (Example 6.2 continued) Suppose that we flip a coin until a head first appears, and if the number of tosses equals n, then we are paid 2^n dollars. What is the expected value of the payment?

We let Y represent the payment. Then,

$$P(Y = 2^n) = \frac{1}{2^n} \ ,$$

for $n \geq 1$. Thus,

$$E(Y) = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} \ ,$$

which is a divergent sum. Thus, Y has no expectation. This example is called the *St. Petersburg Paradox*. The fact that the above sum is infinite suggests that a player should be willing to pay any fixed amount per game for the privilege of playing this game. The reader is asked to consider how much he or she would be willing to pay for this privilege. It is unlikely that the reader's answer is more than 10 dollars; therein lies the paradox.

In the early history of probability, various mathematicians gave ways to resolve this paradox. One idea (due to G. Cramer) consists of assuming that the amount of money in the world is finite. He thus assumes that there is some fixed value of n such that if the number of tosses equals or exceeds n, the payment is 2^n dollars. The reader is asked to show in Exercise 20 that the expected value of the payment is now finite.

Daniel Bernoulli and Cramer also considered another way to assign value to the payment. Their idea was that the value of a payment is some function of the payment; such a function is now called a utility function. Examples of reasonable utility functions might include the square-root function or the logarithm function. In both cases, the value of $2n$ dollars is less than twice the value of n dollars. It can easily be shown that in both cases, the expected utility of the payment is finite (see Exercise 20). □

Example 6.4 Let T be the time for the first success in a Bernoulli trials process. Then we take as sample space Ω the integers 1, 2, ... and assign the geometric distribution

$$m(j) = P(T = j) = q^{j-1}p \ .$$

Thus,

$$\begin{aligned} E(T) &= 1 \cdot p + 2qp + 3q^2p + \cdots \\ &= p(1 + 2q + 3q^2 + \cdots) \ . \end{aligned}$$

Now if $|x| < 1$, then

$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x} \ .$$

Differentiating this formula, we get

$$1 + 2x + 3x^2 + \cdots = \frac{1}{(1-x)^2} \ ,$$

so

$$E(T) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p} \ .$$

In particular, we see that if we toss a fair coin a sequence of times, the expected time until the first heads is $1/(1/2) = 2$. If we roll a die a sequence of times, the expected number of rolls until the first six is $1/(1/6) = 6$. □

Interpretation of Expected Value

In statistics, one is frequently concerned with the average value of a set of data. The following example shows that the ideas of average value and expected value are very closely related.

Example 6.5 The heights, in inches, of the women on the Swarthmore basketball team are 5' 9", 5' 9", 5' 6", 5' 8", 5' 11", 5' 5", 5' 7", 5' 6", 5' 6", 5' 7", 5' 10", and 6' 0".

A statistician would compute the average height (in inches) as follows:

$$\frac{69 + 69 + 66 + 68 + 71 + 65 + 67 + 66 + 66 + 67 + 70 + 72}{12} = 67.9 \ .$$

One can also interpret this number as the expected value of a random variable. To see this, let an experiment consist of choosing one of the women at random, and let X denote her height. Then the expected value of X equals 67.9. □

Of course, just as with the frequency interpretation of probability, to interpret expected value as an average outcome requires further justification. We know that for any finite experiment the average of the outcomes is not predictable. However, we shall eventually prove that the average will usually be close to $E(X)$ if we repeat the experiment a large number of times. We first need to develop some properties of the expected value. Using these properties, and those of the concept of the variance

X	Y
HHH	1
HHT	2
HTH	3
HTT	2
THH	2
THT	3
TTH	2
TTT	1

Table 6.2: Tossing a coin three times.

to be introduced in the next section, we shall be able to prove the *Law of Large Numbers.* This theorem will justify mathematically both our frequency concept of probability and the interpretation of expected value as the average value to be expected in a large number of experiments.

Expectation of a Function of a Random Variable

Suppose that X is a discrete random variable with sample space Ω, and $\phi(x)$ is a real-valued function with domain Ω. Then $\phi(X)$ is a real-valued random variable. One way to determine the expected value of $\phi(X)$ is to first determine the distribution function of this random variable, and then use the definition of expectation. However, there is a better way to compute the expected value of $\phi(X)$, as demonstrated in the next example.

Example 6.6 Suppose a coin is tossed 9 times, with the result

$$HHHTTTTHT\ .$$

The first set of three heads is called a *run*. There are three more runs in this sequence, namely the next four tails, the next head, and the next tail. We do not consider the first two tosses to constitute a run, since the third toss has the same value as the first two.

Now suppose an experiment consists of tossing a fair coin three times. Find the expected number of runs. It will be helpful to think of two random variables, X and Y, associated with this experiment. We let X denote the sequence of heads and tails that results when the experiment is performed, and Y denote the number of runs in the outcome X. The possible outcomes of X and the corresponding values of Y are shown in Table 6.2.

To calculate $E(Y)$ using the definition of expectation, we first must find the distribution function $m(y)$ of Y i.e., we group together those values of X with a common value of Y and add their probabilities. In this case, we calculate that the distribution function of Y is: $m(1) = 1/4,\ m(2) = 1/2$, and $m(3) = 1/4$. One easily finds that $E(Y) = 2$.

Now suppose we didn't group the values of X with a common Y-value, but instead, for each X-value x, we multiply the probability of x and the corresponding value of Y, and add the results. We obtain

$$1\left(\frac{1}{8}\right) + 2\left(\frac{1}{8}\right) + 3\left(\frac{1}{8}\right) + 2\left(\frac{1}{8}\right) + 2\left(\frac{1}{8}\right) + 3\left(\frac{1}{8}\right) + 2\left(\frac{1}{8}\right) + 1\left(\frac{1}{8}\right) ,$$

which equals 2.

This illustrates the following general principle. If X and Y are two random variables, and Y can be written as a function of X, then one can compute the expected value of Y using the distribution function of X. □

Theorem 6.1 If X is a discrete random variable with sample space Ω and distribution function $m(x)$, and if $\phi : \Omega \to \mathrm{R}$ is a function, then

$$E(\phi(X)) = \sum_{x \in \Omega} \phi(x) m(x) ,$$

provided the series converges absolutely. □

The proof of this theorem is straightforward, involving nothing more than grouping values of X with a common Y-value, as in Example 6.6.

The Sum of Two Random Variables

Many important results in probability theory concern sums of random variables. We first consider what it means to add two random variables.

Example 6.7 We flip a coin and let X have the value 1 if the coin comes up heads and 0 if the coin comes up tails. Then, we roll a die and let Y denote the face that comes up. What does $X + Y$ mean, and what is its distribution? This question is easily answered in this case, by considering, as we did in Chapter 4, the joint random variable $Z = (X, Y)$, whose outcomes are ordered pairs of the form (x, y), where $0 \le x \le 1$ and $1 \le y \le 6$. The description of the experiment makes it reasonable to assume that X and Y are independent, so the distribution function of Z is uniform, with $1/12$ assigned to each outcome. Now it is an easy matter to find the set of outcomes of $X + Y$, and its distribution function. □

In Example 6.1, the random variable X denoted the number of heads which occur when a fair coin is tossed three times. It is natural to think of X as the sum of the random variables X_1, X_2, X_3, where X_i is defined to be 1 if the ith toss comes up heads, and 0 if the ith toss comes up tails. The expected values of the X_i's are extremely easy to compute. It turns out that the expected value of X can be obtained by simply adding the expected values of the X_i's. This fact is stated in the following theorem.

Theorem 6.2 Let X and Y be random variables with finite expected values. Then

$$E(X+Y) = E(X) + E(Y) \ ,$$

and if c is any constant, then

$$E(cX) = cE(X) \ .$$

Proof. Let the sample spaces of X and Y be denoted by Ω_X and Ω_Y, and suppose that

$$\Omega_X = \{x_1, x_2, \ldots\}$$

and

$$\Omega_Y = \{y_1, y_2, \ldots\} \ .$$

Then we can consider the random variable $X+Y$ to be the result of applying the function $\phi(x,y) = x+y$ to the joint random variable (X,Y). Then, by Theorem 6.1, we have

$$\begin{aligned} E(X+Y) &= \sum_j \sum_k (x_j + y_k) P(X = x_j,\ Y = y_k) \\ &= \sum_j \sum_k x_j P(X = x_j,\ Y = y_k) + \sum_j \sum_k y_k P(X = x_j,\ Y = y_k) \\ &= \sum_j x_j P(X = x_j) + \sum_k y_k P(Y = y_k) \ . \end{aligned}$$

The last equality follows from the fact that

$$\sum_k P(X = x_j,\ Y = y_k) = P(X = x_j)$$

and

$$\sum_j P(X = x_j,\ Y = y_k) = P(Y = y_k) \ .$$

Thus,

$$E(X+Y) = E(X) + E(Y) \ .$$

If c is any constant,

$$\begin{aligned} E(cX) &= \sum_j c x_j P(X = x_j) \\ &= c \sum_j x_j P(X = x_j) \\ &= cE(X) \ . \end{aligned}$$

□

X			Y
a	b	c	3
a	c	b	1
b	a	c	1
b	c	a	0
c	a	b	0
c	b	a	1

Table 6.3: Number of fixed points.

It is easy to prove by mathematical induction that *the expected value of the sum of any finite number of random variables is the sum of the expected values of the individual random variables.*

It is important to note that mutual independence of the summands was not needed as a hypothesis in the Theorem 6.2 and its generalization. The fact that expectations add, whether or not the summands are mutually independent, is sometimes referred to as the First Fundamental Mystery of Probability.

Example 6.8 Let Y be the number of fixed points in a random permutation of the set $\{a, b, c\}$. To find the expected value of Y, it is helpful to consider the basic random variable associated with this experiment, namely the random variable X which represents the random permutation. There are six possible outcomes of X, and we assign to each of them the probability $1/6$ see Table 6.3. Then we can calculate $E(Y)$ using Theorem 6.1, as

$$3\Bigl(\frac{1}{6}\Bigr) + 1\Bigl(\frac{1}{6}\Bigr) + 1\Bigl(\frac{1}{6}\Bigr) + 0\Bigl(\frac{1}{6}\Bigr) + 0\Bigl(\frac{1}{6}\Bigr) + 1\Bigl(\frac{1}{6}\Bigr) = 1\ .$$

We now give a very quick way to calculate the average number of fixed points in a random permutation of the set $\{1, 2, 3, \ldots, n\}$. Let Z denote the random permutation. For each i, $1 \le i \le n$, let X_i equal 1 if Z fixes i, and 0 otherwise. So if we let F denote the number of fixed points in Z, then

$$F = X_1 + X_2 + \cdots + X_n\ .$$

Therefore, Theorem 6.2 implies that

$$E(F) = E(X_1) + E(X_2) + \cdots + E(X_n)\ .$$

But it is easy to see that for each i,

$$E(X_i) = \frac{1}{n}\ ,$$

so

$$E(F) = 1\ .$$

This method of calculation of the expected value is frequently very useful. It applies whenever the random variable in question can be written as a sum of simpler random variables. We emphasize again that it is not necessary that the summands be mutually independent. □

Bernoulli Trials

Theorem 6.3 Let S_n be the number of successes in n Bernoulli trials with probability p for success on each trial. Then the expected number of successes is np. That is,

$$E(S_n) = np .$$

Proof. Let X_j be a random variable which has the value 1 if the jth outcome is a success and 0 if it is a failure. Then, for each X_j,

$$E(X_j) = 0 \cdot (1 - p) + 1 \cdot p = p .$$

Since

$$S_n = X_1 + X_2 + \cdots + X_n ,$$

and the expected value of the sum is the sum of the expected values, we have

$$\begin{aligned} E(S_n) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= np . \end{aligned}$$

□

Poisson Distribution

Recall that the Poisson distribution with parameter λ was obtained as a limit of binomial distributions with parameters n and p, where it was assumed that $np = \lambda$, and $n \to \infty$. Since for each n, the corresponding binomial distribution has expected value λ, it is reasonable to guess that the expected value of a Poisson distribution with parameter λ also has expectation equal to λ. This is in fact the case, and the reader is invited to show this (see Exercise 21).

Independence

If X and Y are two random variables, it is not true in general that $E(X \cdot Y) = E(X)E(Y)$. However, this is true if X and Y are *independent.*

Theorem 6.4 If X and Y are independent random variables, then

$$E(X \cdot Y) = E(X)E(Y) .$$

Proof. Suppose that

$$\Omega_X = \{x_1, x_2, \ldots\}$$

and

$$\Omega_Y = \{y_1, y_2, \ldots\}$$

are the sample spaces of X and Y, respectively. Using Theorem 6.1, we have

$$E(X \cdot Y) = \sum_j \sum_k x_j y_k P(X = x_j,\ Y = y_k)\ .$$

But if X and Y are independent,

$$P(X = x_j, Y = y_k) = P(X = x_j)P(Y = y_k)\ .$$

Thus,

$$\begin{aligned} E(X \cdot Y) &= \sum_j \sum_k x_j y_k P(X = x_j)P(Y = y_k) \\ &= \left(\sum_j x_j P(X = x_j)\right)\left(\sum_k y_k P(Y = y_k)\right) \\ &= E(X)E(Y)\ . \end{aligned}$$

□

Example 6.9 A coin is tossed twice. $X_i = 1$ if the ith toss is heads and 0 otherwise. We know that X_1 and X_2 are independent. They each have expected value 1/2. Thus $E(X_1 \cdot X_2) = E(X_1)E(X_2) = (1/2)(1/2) = 1/4$. □

We next give a simple example to show that the expected values need not multiply if the random variables are not independent.

Example 6.10 Consider a single toss of a coin. We define the random variable X to be 1 if heads turns up and 0 if tails turns up, and we set $Y = 1 - X$. Then $E(X) = E(Y) = 1/2$. But $X \cdot Y = 0$ for either outcome. Hence, $E(X \cdot Y) = 0 \neq E(X)E(Y)$. □

We return to our records example of Section 3.1 for another application of the result that the expected value of the sum of random variables is the sum of the expected values of the individual random variables.

Records

Example 6.11 We start keeping snowfall records this year and want to find the expected number of records that will occur in the next n years. The first year is necessarily a record. The second year will be a record if the snowfall in the second year is greater than that in the first year. By symmetry, this probability is 1/2. More generally, let X_j be 1 if the jth year is a record and 0 otherwise. To find $E(X_j)$, we need only find the probability that the jth year is a record. But the record snowfall for the first j years is equally likely to fall in any one of these years, so $E(X_j) = 1/j$. Therefore, if S_n is the total number of records observed in the first n years,

$$E(S_n) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \; .$$

This is the famous *divergent harmonic series.* It is easy to show that

$$E(S_n) \sim \log n$$

as $n \to \infty$. A more accurate approximation to $E(S_n)$ is given by the expression

$$\log n + \gamma + \frac{1}{2n} \; ,$$

where γ denotes Euler's constant, and is approximately equal to .5772.

Therefore, in ten years the expected number of records is approximately 2.9298; the exact value is the sum of the first ten terms of the harmonic series which is 2.9290. □

Craps

Example 6.12 In the game of craps, the player makes a bet and rolls a pair of dice. If the sum of the numbers is 7 or 11 the player wins, if it is 2, 3, or 12 the player loses. If any other number results, say r, then r becomes the player's point and he continues to roll until either r or 7 occurs. If r comes up first he wins, and if 7 comes up first he loses. The program **Craps** simulates playing this game a number of times.

We have run the program for 1000 plays in which the player bets 1 dollar each time. The player's average winnings were $-.006$. The game of craps would seem to be only slightly unfavorable. Let us calculate the expected winnings on a single play and see if this is the case. We construct a two-stage tree measure as shown in Figure 6.1.

The first stage represents the possible sums for his first roll. The second stage represents the possible outcomes for the game if it has not ended on the first roll. In this stage we are representing the possible outcomes of a sequence of rolls required to determine the final outcome. The branch probabilities for the first stage are computed in the usual way assuming all 36 possibilites for outcomes for the pair of dice are equally likely. For the second stage we assume that the game will eventually end, and we compute the conditional probabilities for obtaining either the point or a 7. For example, assume that the player's point is 6. Then the game will end when one of the eleven pairs, $(1,5)$, $(2,4)$, $(3,3)$, $(4,2)$, $(5,1)$, $(1,6)$, $(2,5)$, $(3,4)$, $(4,3)$, $(5,2)$, $(6,1)$, occurs. We assume that each of these possible pairs has the same probability. Then the player wins in the first five cases and loses in the last six. Thus the probability of winning is $5/11$ and the probability of losing is $6/11$. From the path probabilities, we can find the probability that the player wins 1 dollar; it is $244/495$. The probability of losing is then $251/495$. Thus if X is his winning for

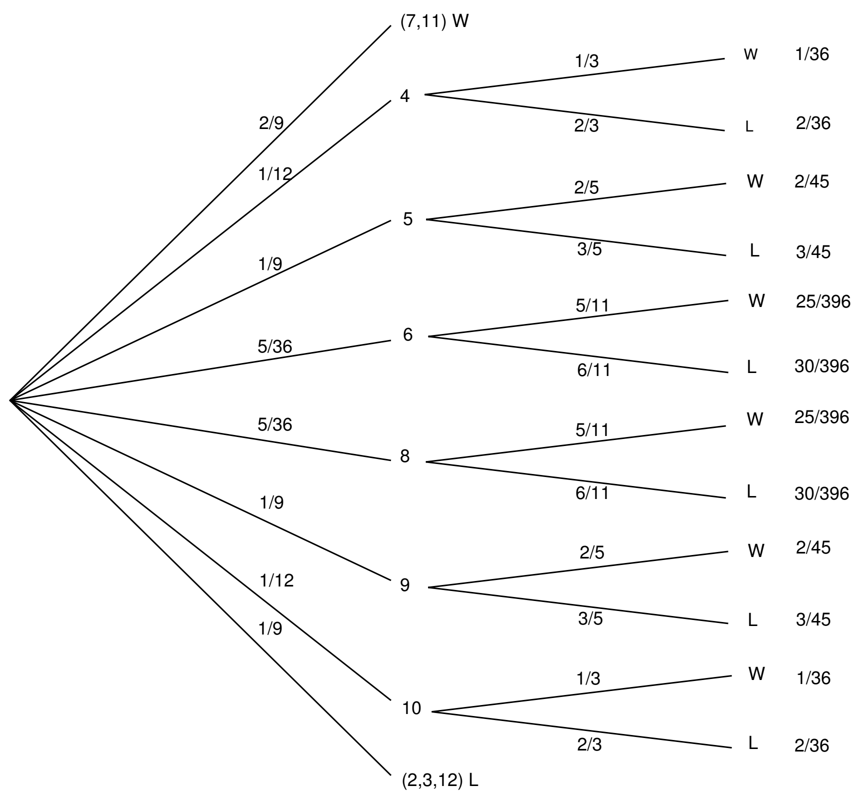

Figure 6.1: Tree measure for craps.

a dollar bet,

$$\begin{aligned} E(X) &= 1\Big(\frac{244}{495}\Big) + (-1)\Big(\frac{251}{495}\Big) \\ &= -\frac{7}{495} \approx -.0141 \ . \end{aligned}$$

The game is unfavorable, but only slightly. The player's expected gain in n plays is $-n(.0141)$. If n is not large, this is a small expected loss for the player. The casino makes a large number of plays and so can afford a small average gain per play and still expect a large profit. □

Roulette

Example 6.13 In Las Vegas, a roulette wheel has 38 slots numbered 0, 00, 1, 2, ..., 36. The 0 and 00 slots are green, and half of the remaining 36 slots are red and half are black. A croupier spins the wheel and throws an ivory ball. If you bet 1 dollar on red, you win 1 dollar if the ball stops in a red slot, and otherwise you lose a dollar. We wish to calculate the expected value of your winnings, if you bet 1 dollar on red.

Let X be the random variable which denotes your winnings in a 1 dollar bet on red in Las Vegas roulette. Then the distribution of X is given by

$$m_X = \begin{pmatrix} -1 & 1 \\ 20/38 & 18/38 \end{pmatrix},$$

and one can easily calculate (see Exercise 5) that

$$E(X) \approx -.0526 \ .$$

We now consider the roulette game in Monte Carlo, and follow the treatment of Sagan.[1] In the roulette game in Monte Carlo there is only one 0. If you bet 1 franc on red and a 0 turns up, then, depending upon the casino, one or more of the following options may be offered:

(a) You get 1/2 of your bet back, and the casino gets the other half of your bet.

(b) Your bet is put "in prison," which we will denote by P_1. If red comes up on the next turn, you get your bet back (but you don't win any money). If black or 0 comes up, you lose your bet.

(c) Your bet is put in prison P_1, as before. If red comes up on the next turn, you get your bet back, and if black comes up on the next turn, then you lose your bet. If a 0 comes up on the next turn, then your bet is put into double prison, which we will denote by P_2. If your bet is in double prison, and if red comes up on the next turn, then your bet is moved back to prison P_1 and the game proceeds as before. If your bet is in double prison, and if black or 0 come up on the next turn, then you lose your bet. We refer the reader to Figure 6.2, where a tree for this option is shown. In this figure, S is the starting position, W means that you win your bet, L means that you lose your bet, and E means that you break even.

[1] H. Sagan, *Markov Chains in Monte Carlo,* Math. Mag., vol. 54, no. 1 (1981), pp. 3-10.

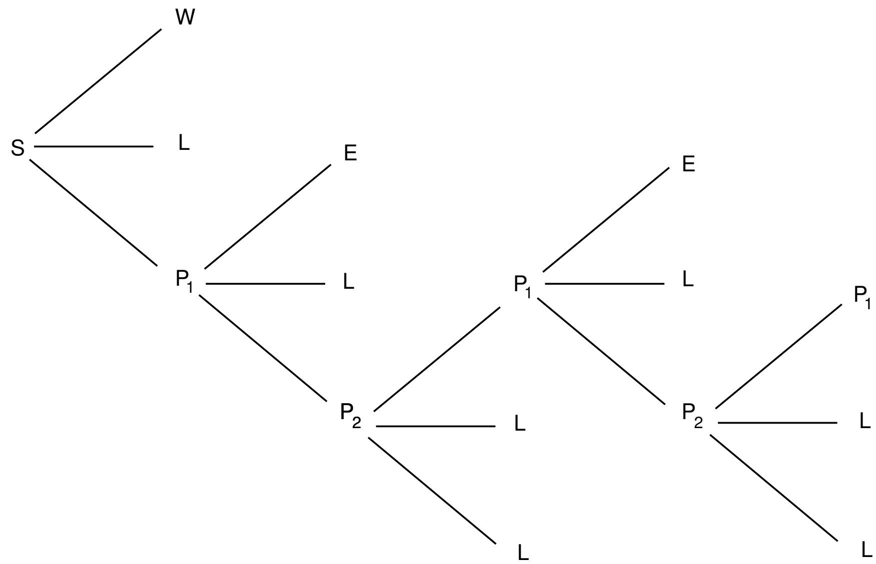

Figure 6.2: Tree for 2-prison Monte Carlo roulette.

It is interesting to compare the expected winnings of a 1 franc bet on red, under each of these three options. We leave the first two calculations as an exercise (see Exercise 37). Suppose that you choose to play alternative (c). The calculation for this case illustrates the way that the early French probabilists worked problems like this.

Suppose you bet on red, you choose alternative (c), and a 0 comes up. Your possible future outcomes are shown in the tree diagram in Figure 6.3. Assume that your money is in the first prison and let x be the probability that you lose your franc. From the tree diagram we see that

$$x = \frac{18}{37} + \frac{1}{37} P(\text{you lose your franc} \mid \text{your franc is in } P_2)\ .$$

Also,

$$P(\text{you lose your franc} \mid \text{your franc is in } P_2) = \frac{19}{37} + \frac{18}{37} x\ .$$

So, we have

$$x = \frac{18}{37} + \frac{1}{37}\Bigl(\frac{19}{37} + \frac{18}{37} x\Bigr)\ .$$

Solving for x, we obtain $x = 685/1351$. Thus, starting at S, the probability that you lose your bet equals

$$\frac{18}{37} + \frac{1}{37} x = \frac{25003}{49987}\ .$$

To find the probability that you win when you bet on red, note that you can only win if red comes up on the first turn, and this happens with probability 18/37. Thus your expected winnings are

$$1 \cdot \frac{18}{37} - 1 \cdot \frac{25003}{49987} = -\frac{687}{49987} \approx -.0137\ .$$

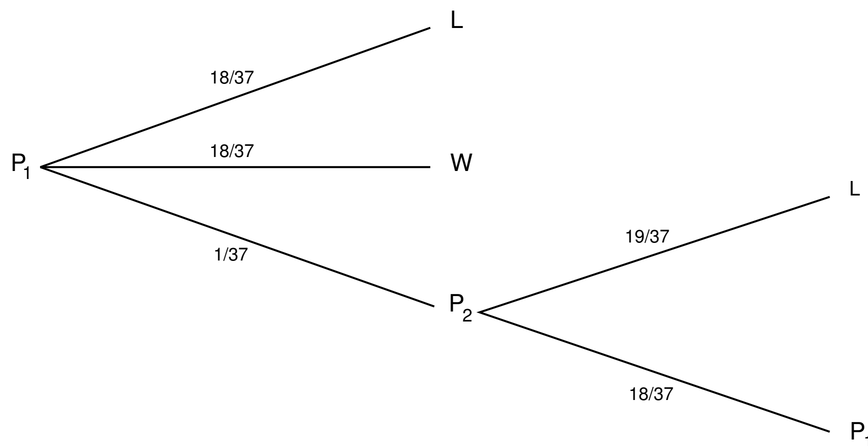

Figure 6.3: Your money is put in prison.

It is interesting to note that the more romantic option (c) is less favorable than option (a) (see Exercise 37).

If you bet 1 dollar on the number 17, then the distribution function for your winnings X is

$$P_X = \begin{pmatrix} -1 & 35 \\ 36/37 & 1/37 \end{pmatrix} ,$$

and the expected winnings are

$$-1 \cdot \frac{36}{37} + 35 \cdot \frac{1}{37} = -\frac{1}{37} \approx -.027 .$$

Thus, at Monte Carlo different bets have different expected values. In Las Vegas almost all bets have the same expected value of $-2/38 = -.0526$ (see Exercises 4 and 5). □

Conditional Expectation

Definition 6.2 If F is any event and X is a random variable with sample space $\Omega = \{x_1, x_2, \ldots\}$, then the *conditional expectation given* F is defined by

$$E(X|F) = \sum_j x_j P(X = x_j|F) .$$

Conditional expectation is used most often in the form provided by the following theorem. □

Theorem 6.5 Let X be a random variable with sample space Ω. If $F_1, F_2, \ldots, F_r$ are events such that $F_i \cap F_j = \emptyset$ for $i \neq j$ and $\Omega = \cup_j F_j$, then

$$E(X) = \sum_j E(X|F_j)P(F_j) .$$

Proof. We have

$$\begin{aligned}
\sum_j E(X|F_j)P(F_j) &= \sum_j \sum_k x_k P(X = x_k|F_j)P(F_j) \\
&= \sum_j \sum_k x_k P(X = x_k \text{ and } F_j \text{ occurs}) \\
&= \sum_k \sum_j x_k P(X = x_k \text{ and } F_j \text{ occurs}) \\
&= \sum_k x_k P(X = x_k) \\
&= E(X)\ .
\end{aligned}$$

□

Example 6.14 (Example 6.12 continued) Let T be the number of rolls in a single play of craps. We can think of a single play as a two-stage process. The first stage consists of a single roll of a pair of dice. The play is over if this roll is a 2, 3, 7, 11, or 12. Otherwise, the player's point is established, and the second stage begins. This second stage consists of a sequence of rolls which ends when either the player's point or a 7 is rolled. We record the outcomes of this two-stage experiment using the random variables X and S, where X denotes the first roll, and S denotes the number of rolls in the second stage of the experiment (of course, S is sometimes equal to 0). Note that $T = S + 1$. Then by Theorem 6.5

$$E(T) = \sum_{j=2}^{12} E(T|X = j)P(X = j)\ .$$

If $j = 7$, 11 or 2, 3, 12, then $E(T|X = j) = 1$. If $j = 4, 5, 6, 8, 9,$ or 10, we can use Example 6.4 to calculate the expected value of S. In each of these cases, we continue rolling until we get either a j or a 7. Thus, S is geometrically distributed with parameter p, which depends upon j. If $j = 4$, for example, the value of p is $3/36 + 6/36 = 1/4$. Thus, in this case, the expected number of additional rolls is $1/p = 4$, so $E(T|X = 4) = 1 + 4 = 5$. Carrying out the corresponding calculations for the other possible values of j and using Theorem 6.5 gives

$$\begin{aligned}
E(T) &= 1\Bigl(\frac{12}{36}\Bigr) + \Bigl(1 + \frac{36}{3+6}\Bigr)\Bigl(\frac{3}{36}\Bigr) + \Bigl(1 + \frac{36}{4+6}\Bigr)\Bigl(\frac{4}{36}\Bigr) \\
&\quad + \Bigl(1 + \frac{36}{5+6}\Bigr)\Bigl(\frac{5}{36}\Bigr) + \Bigl(1 + \frac{36}{5+6}\Bigr)\Bigl(\frac{5}{36}\Bigr) \\
&\quad + \Bigl(1 + \frac{36}{4+6}\Bigr)\Bigl(\frac{4}{36}\Bigr) + \Bigl(1 + \frac{36}{3+6}\Bigr)\Bigl(\frac{3}{36}\Bigr) \\
&= \frac{557}{165} \\
&\approx 3.375\ldots\ .
\end{aligned}$$

□

Martingales

We can extend the notion of fairness to a player playing a sequence of games by using the concept of conditional expectation.

Example 6.15 Let S_1, S_2, ..., S_n be Peter's accumulated fortune in playing heads or tails (see Example 1.4). Then

$$E(S_n|S_{n-1} = a, \ldots, S_1 = r) = \frac{1}{2}(a+1) + \frac{1}{2}(a-1) = a \ .$$

We note that Peter's expected fortune after the next play is equal to his present fortune. When this occurs, we say the game is *fair.* A fair game is also called a *martingale.* If the coin is biased and comes up heads with probability p and tails with probability $q = 1 - p$, then

$$E(S_n|S_{n-1} = a, \ldots, S_1 = r) = p(a+1) + q(a-1) = a + p - q \ .$$

Thus, if $p < q$, this game is unfavorable, and if $p > q$, it is favorable. □

If you are in a casino, you will see players adopting elaborate *systems* of play to try to make unfavorable games favorable. Two such systems, the martingale doubling system and the more conservative Labouchere system, were described in Exercises 1.1.9 and 1.1.10. Unfortunately, such systems cannot change even a fair game into a favorable game.

Even so, it is a favorite pastime of many people to develop systems of play for gambling games and for other games such as the stock market. We close this section with a simple illustration of such a system.

Stock Prices

Example 6.16 Let us assume that a stock increases or decreases in value each day by 1 dollar, each with probability 1/2. Then we can identify this simplified model with our familiar game of heads or tails. We assume that a buyer, Mr. Ace, adopts the following strategy. He buys the stock on the first day at its price V. He then waits until the price of the stock increases by one to $V + 1$ and sells. He then continues to watch the stock until its price falls back to V. He buys again and waits until it goes up to $V + 1$ and sells. Thus he holds the stock in intervals during which it increases by 1 dollar. In each such interval, he makes a profit of 1 dollar. However, we assume that he can do this only for a finite number of trading days. Thus he can lose if, in the last interval that he holds the stock, it does not get back up to $V + 1$; and this is the only way he can lose. In Figure 6.4 we illustrate a typical history if Mr. Ace must stop in twenty days. Mr. Ace holds the stock under his system during the days indicated by broken lines. We note that for the history shown in Figure 6.4, his system nets him a gain of 4 dollars.

We have written a program **StockSystem** to simulate the fortune of Mr. Ace if he uses his sytem over an n-day period. If one runs this program a large number

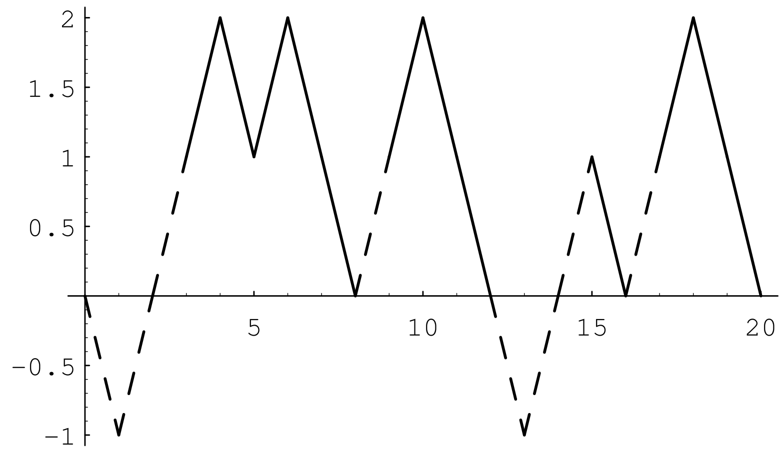

Figure 6.4: Mr. Ace's system.

of times, for $n = 20$, say, one finds that his expected winnings are very close to 0, but the probability that he is ahead after 20 days is significantly greater than 1/2. For small values of n, the exact distribution of winnings can be calculated. The distribution for the case $n = 20$ is shown in Figure 6.5. Using this distribution, it is easy to calculate that the expected value of his winnings is exactly 0. This is another instance of the fact that a fair game (a martingale) remains fair under quite general systems of play.

Although the expected value of his winnings is 0, the probability that Mr. Ace is ahead after 20 days is about .610. Thus, he would be able to tell his friends that his system gives him a better chance of being ahead than that of someone who simply buys the stock and holds it, if our simple random model is correct. There have been a number of studies to determine how random the stock market is. □

Historical Remarks

With the Law of Large Numbers to bolster the frequency interpretation of probability, we find it natural to justify the definition of expected value in terms of the average outcome over a large number of repetitions of the experiment. The concept of expected value was used before it was formally defined; and when it was used, it was considered not as an average value but rather as the appropriate value for a gamble. For example recall, from the Historical Remarks section of Chapter 1, Section 1.2, Pascal's way of finding the value of a three-game series that had to be called off before it is finished.

Pascal first observed that if each player has only one game to win, then the stake of 64 pistoles should be divided evenly. Then he considered the case where one player has won two games and the other one.

> Then consider, Sir, if the first man wins, he gets 64 pistoles, if he loses he gets 32. Thus if they do not wish to risk this last game, but wish