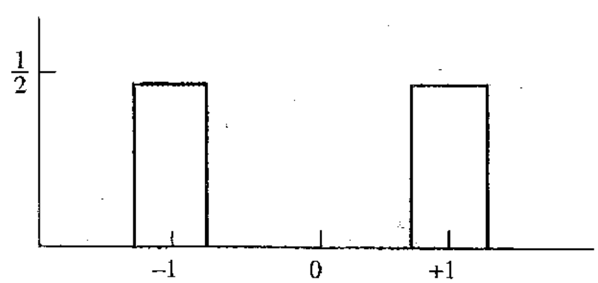
3.2 Expected Values of Random Variables

Because a probability can be thought of as the long-run relative frequency of occurrence for an event, a probability distribution can be interpreted as showing the long-run relative frequency of occurrence for numerical outcomes associated with an experiment. Suppose, for example, that you and a friend are matching balanced coins. Each of you tosses a coin. If the upper faces match, you win \$1.00; if they do not match, you lose \$1.00 (your friend wins \$1.00). The probability of a match is 0.5 and, in the long run, you should win about half of the time. Thus, a relative frequency distribution of your winnings should look like the one shown in Figure 3.3. The −1 under the leftmost bar indicates a loss by you.

FIGURE 3.3
Relative frequency of winnings.



On the average, how much will you win per game over the long run? If Figure 3.3 presents a correct display of your winnings, you win -1 half of the time and $+1$ half of the time, for an average of

$$(-1)\left(\frac{1}{2}\right) + (1)\left(-\frac{1}{2}\right) = 0$$

This average is sometimes called your expected winnings per game, or the *expected value* of your winnings. (An expected value of 0 indicates that this is a fair game.) The general definition of expected value is given in Definition 3.4.

DEFINITION 3.4

The **expected value*** of a discrete random variable X with probability distribution $p(x)$ is given by

$$E(X) = \sum_x x\, p(x)$$

(The sum is over all values of x for which $p(x)>0$.)
We sometimes use the notation

$$E(X) = \mu$$

for this equivalence.

Now payday has arrived, and you and your friend up the stakes to \$10 per game of matching coins. You now win -10 or $+10$ with equal probability. Your expected winnings per game is

$$(-10)\left(\frac{1}{2}\right) + (10)\left(\frac{1}{2}\right) = 0$$

and the game is still fair. The new stakes can be thought of as a function of the old in the sense that, if X represents your winnings per game when you were playing for \$1.00, then $10X$ represents your winnings per game when you play for \$10.00. Such functions of random variables arise often. The extension of the definition of expected value to cover these cases is given in Theorem 3.1.

THEOREM 3.1

If X is a discrete random variable with probability distribution $p(x)$ and if $g(x)$ is any real-valued function of X, then

$$E[g(X)] = \sum_x g(x)\, p(x)$$

(The proof of this theorem will not be given.)

*We assume absolute convergence when the range of X is countable; we talk about an *expectation* only when it is assumed to exist.

You and your friend decide to complicate the payoff picture in the coin-matching game by agreeing to let you win \$1 if the match is tails and \$2 if the match is heads. You still lose \$1 if the coins do not match. Quickly you see that this is not a fair game, because your expected winnings are

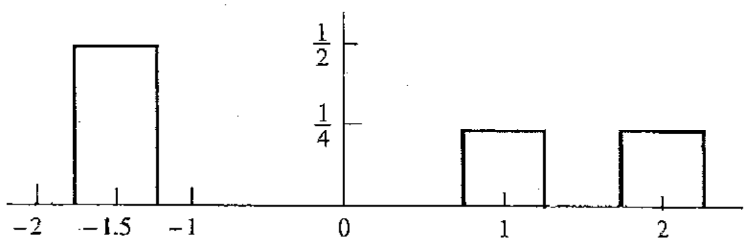
$$(-1)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{4}\right) + (2)\left(\frac{1}{4}\right) = 0.25$$

You compensate for this by agreeing to pay your friend \$1.50 if the coins do not match. Then, your expected winnings per game are

$$(-1.5)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{4}\right) + (2)\left(\frac{1}{4}\right) = 0$$

and the game is again fair. What is the difference between this game and the original one, in which all payoffs were \$1? The difference certainly cannot be explained by the expected value, since both games are fair. You can win more but also lose more with the new payoffs, and the difference between the two games can be explained to some extent by the increased variability of your winnings across many games. This increased variability can be seen in Figure 3.4, which displays the relative frequency for your winnings in the new game; the winnings are more spread out than they were in Figure 3.3. Formally, variation is often measured by the *variance* and by a related quantity called the *standard deviation.*

FIGURE 3.4
Relative frequency of winnings.



DEFINITION 3.5

The **variance** of a random variable X with expected value μ is given by

$$V(X) = E\left[(X-\mu)^2\right]$$

We sometimes use this notation

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

for this equivalence.

The smallest value that σ^2 can assume is zero; it occurs when all the probability is concentrated at a single point (that is, when X takes on a constant value with probability 1). The variance becomes larger as the points with positive probability spread out more.

The variance squares the units in which we are measuring. A measure of variation that maintains the original units is the standard deviation.

DEFINITION 3.6

The **standard deviation** of a random variable X is the square root of the Variance and is given by

$$\sigma = \sqrt{\sigma^2} = \sqrt{E\left[(X-\mu)^2\right]}$$

For the game represented in Figure 3.3, the variance of your winnings (with $\mu = 0$) is

$$\begin{aligned}\sigma^2 &= E(X-\mu)^2 \\ &= (-1)^2\left(\frac{1}{2}\right) + (1)^2\left(\frac{1}{2}\right) = 1\end{aligned}$$

It follows that $\sigma = 1$, as well. For the game represented in Figure 3.4, the variance of your winnings is

$$\begin{aligned}\sigma^2 &= (-1.5)^2\left(\frac{1}{2}\right) + (1)^2\left(\frac{1}{4}\right) + (2)^2\left(\frac{1}{4}\right) \\ &= 2.375\end{aligned}$$

and the standard deviation is

$$\sigma = 1.54$$

Which game would you rather play?

The standard deviation can be thought of as the size of a "typical" deviation between an observed outcome and the expected value. For the situation described in Figure 3.3, each outcome (-1 or $+1$) deviates by precisely one standard deviation from the expected value. For the situation described in Figure 3.4, the positive values average 1.5 units from the expected value of 0 (as do the negative values), and so 1.5 units is approximately one standard deviation here.

The mean and the standard deviation often yield a useful summary of the probability distribution for a random variable that can assume many values. An illustration is provided by the age distribution of the U.S. population for 1990 and 2050 (projected, as shown in Table 3.1).

Age is actually a continuous measurement, but since it is reported in categories, we can treat it as a discrete random variable for purposes of approximating its key functions. To move from continuous age intervals to discrete age classes, we assign each interval the value of its midpoint (rounded). Thus, the data in Table 3.1 are interpreted as showing that 7.6% of the 1990 population were around 3 years of age and that 22.5% of the 2050 population is anticipated to be around 55 years of age. (The open intervals at the upper end were stopped at 100 for convenience.)

TABLE 3.1
Age Distribution of U.S. Population (in percents)

Age Interval	Age Midpoint	1990	2050
Under 5	3	7.6%	6.4%
5–13	9	12.8	11.6
14–17	16	5.3	5.2
18–24	21	10.8	9.0
25–34	30	17.3	12.5
35–44	40	15.1	12.2
45–64	55	18.6	22.5
65–84	75	11.3	16.0
85 and over	92	1.2	4.6

Source: U.S. Bureau of the Census.

Interpreting the percentages as probabilities, we see that the mean age for 1990 is approximated by

$$\begin{aligned}\mu &= \sum xp(x) \\ &= 3(0.076) + 9(0.128) + \cdots + 92(0.012) \\ &= 35.5\end{aligned}$$

(How does this compare with the median age for 1990, as approximated from Table 3.1?) For 2050, the mean age is approximated by

$$\begin{aligned}\mu &= \sum xp(x) \\ &= 3(0.064) + 9(0.116) + \cdots + 92(0.046) \\ &= 41.2\end{aligned}$$

Over the projected period, the mean age increases rather markedly (as does the median age).

The variations in the two age distributions can be approximated by the standard deviations. For 1990, this is

$$\begin{aligned}\sigma &= \sqrt{\sum (x-\mu)^2 p(x)} \\ &= \sqrt{(3-35.5)^2(0.76) + \cdots + (92-35.5)^2(0.012)} \\ &= 22.5\end{aligned}$$

A similar calculation for the 2050 data yields $\sigma = 25.4$. These results are summarized in Table 3.2.

TABLE 3.2
Age Distribution of U.S. Population Summary

Statistic	1990	2050
Mean	35.5	41.2
Standard deviation	22.5	25.4

Not only is the population getting older, on the average, but its variability is increasing. What are some of the implications of these trends?

We now provide other examples and extensions of these basic results.

EXAMPLE 3.2 The manager of a stockroom in a factory knows from his study of records that the daily demand (number of times used) for a certain tool has the following probability distribution:

Demand	0	1	2
Probability	0.1	0.5	0.4

(In other words, 50% of the daily records show that the tool was used one time.) Letting X denote the daily demand, find $E(X)$ and $V(X)$.

Solution From Definition 3.4, we see that

$$\begin{aligned} E(X) &= \sum_x xp(x) \\ &= 0(0.1) + 1(0.5) + 2(0.4) = 1.3 \end{aligned}$$

The tool is used an average of 1.3 times per day.

From Definition 3.5, we see that

$$\begin{aligned} V(X) &= E(X - f\mu)^2 \\ &= \sum_x (x-\mu)^2 p(x) \\ &= (0-1.3)^2(0.1) + (1-1.3)^2(0.5) + (2-1.3)^2(0.4) \\ &= (1.69)(0.1) + (0.09)(0.5) + (0.49)(0.4) \\ &= 0.410 \quad \blacksquare \end{aligned}$$

Our work in manipulating expected values can be greatly facilitated by making use of the two results of Theorem 3.2. Often, $g(X)$ is a linear function; and when this is the case, the calculations of expected value and variance are especially simple.

THEOREM 3.2

For any random variable X and constants a and b

1) $E(aX+b) = aE(x)+b$

2) $V(aX+b) = a^2V(x)$

Proof By Theorem 3.1,

$$E(aX+b) = \sum_x (ax+b)\,p(x)$$

$$
\begin{aligned}
E(aX+b) &= \sum_x (ax+b)\,p(x) \\
&= \sum_x [(ax)p(x)+b\,p(x)] \\
&= \sum_x axp(x)+\sum_x bp(x) \\
&= a\sum_x xp(x)+b\sum_x p(x) \\
&= aE(x)+b
\end{aligned}
$$

Notice that $\sum_x p(x)=1$. Also, by definition 3.5,

$$
\begin{aligned}
V(aX+b) &= E[(aX+b)-E(aX+b]^2 \\
&= E[aX+b-(aE(X)+b)]^2 \\
&= E[aX-aE(X)]^2 \\
&= E[a^2(X-E(X))^2] \\
&= a^2\ E[(X-E(X))^2] \\
&= a^2V(x)
\end{aligned}
$$

An important special case of Theorem 3.2 involves establishing a "standardized" variable. If X has mean μ and standard deviation σ, then the "standardized" form of X is given by

$$Y = \frac{X-\mu}{\sigma}$$

Employing Theorem 3.2, one can easily show that $E(Y)=0$ and $V(Y)=1$. This idea will be used often in later chapters.

We illustrate the use of these results in the following example.

EXAMPLE 3.3 In Example 3.2, suppose that it costs the factory \$10 each time the tool is used. Find the mean and the variance of the daily costs of using this tool.

Solution Recall that the X of Example 3.2 is the daily demand. The daily cost of using this tool is the $10X$. By Theorem 3.2, we have

$$
\begin{aligned}
E(10X) &= 10E(X) \\
&= 10(1.3) \\
&= 13
\end{aligned}
$$

Thus, the factory should budget \$13 per day to cover the cost of using the tool.

Also, by Theorem 3.2,

$$
\begin{aligned}
V(10X) &= (10)^2V(X) \\
&= 100(0.410) \\
&= 41
\end{aligned}
$$

We will make use of this value in a later example. ▪

Theorem 3.2 leads us to a more efficient computational formula for variance, as given in Theorem 3.3.

THEOREM 3.3

If X is a random variable with mean (expected value) μ, then

$$V(X) = E(X^2) - \mu^2$$

Proof

Starting with the definition of variance, we have

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

EXAMPLE 3.4 Use the result of Theorem 3.3 to compute the variance of X as given in Example 3.2.

Solution In Example 3.2, X had a probability distribution given by

x	0	1	2
$p(x)$	0.1	0.5	0.4

and we saw that $E(X) = 1.3$. Now,

$$\begin{aligned} E(X^2) &= \sum_x x^2 p(x) \\ &= (0)^2(0.1) + (1)^2(0.5) + (2)^2(0.4) \\ &= 0 + 0.5 + 1.6 \\ &= 2.1 \end{aligned}$$

By Theorem 3.3,

$$\begin{aligned} V(X) &= E(X^2) - \mu^2 \\ &= 2.1 - (1.3)^2 = 0.41 \end{aligned}$$ ▪

We have computed means and variances for a number of probability distributions and noted that these two quantities give us some useful information on the center