2

Discrete Probability Models

2.1 Random Variables

While the concept of a sample space provides a comprehensive description of the possible outcomes of a random experiment, it turns out that, in a host of applications, the sample space provides more information than we need or want. A simple example of this overly informative potential of the sample space is the 36-point space describing the possible outcomes when one rolls a pair of dice. It's certainly true that once a probability has been assigned to each simple event, we are in a position to compute the probability of any and all compound events that may be relevant to a particular discussion. Now suppose that the discussion *du jour* happens to be about the game of craps. The classic wager in the game of craps is based solely on the sum of the two digits facing upwards when two dice are rolled. When making this wager, we don't really care or need to know how a particular sum arose. Only the value of this sum matters. The sum of the digits facing up in two rolled dice is an example of a "random variable" in which we might be interested. This sum can only take on one of eleven possible values, so concentrating on the sum reduces the number of events of interest by more than 2/3. We proceed with our treatment of random variables with an appropriate definition and a collection of examples of the concept. For now, we will restrict attention to the discrete case, that is, to the case in which the sample space is either finite or countably infinite.

Definition 2.1.1. (Random Variables) A *random variable* is a function whose domain is the sample space of a random experiment and whose range is a subset of the real line.

This definition simply states that a random variable is a "mapping" that associates a real number with each simple event in a given sample space. Since different simple events can map onto the same real number, each value of the random variable actually corresponds to a compound event, namely, the set of all simple events which map on to this same number. If you were into linguistics, you would probably notice that the phrase "random variable" isn't an especially good name for the object just defined. You might insist that a random variable X is neither random nor is it a variable. It's actually just a function which maps the sample space S onto the real line R. In a mathematics book, you would find X described by the favored notation for "mappings" in that discipline, i.e., $X : S \to R$. Be that as it may, the language introduced above is traditional in Probability and Statistics, and we will stick with it. Before you know it, the term "random variable" will be rolling off your tongue and you will know precisely what it means simply because you will encounter and use the term with great frequency.

The values that a random variable X may take on will depend on exactly how it was defined. We will of course be interested in the values that X can take, but we will also be interested in the probabilities associated with those values. As mentioned above, each possible value of X corresponds to a compound event. It is thus natural for a particular value of X to inherit the probability carried by the compound event that it goes with. Suppose that A is the compound event consisting of every simple event "a" for which $X = k$, that is, suppose that $X = k$ if and only if $a \in A$. In mathematical function notation, we would write $X(a) = k$ if and only if $a \in A$. This correspondence allows us to identify the probability that must logically be associated with any given value of X. Thus, the probability model for the random experiment with which we started leads to a probability model for that random variable of interest. We often refer to this latter model as the probability distribution of X. In "small problems," this distribution is often displayed as a table listing the possible values of X and their corresponding probabilities. The following examples display the probability distributions of several random variables which arise in simple everyday problems.

Example 2.1.1. Suppose you toss a pair of *fair* coins and you observe a head or a tail on each of the coins. We may write the sample space of this random experiment as $S = \{HH, HT, TH, TT\}$. Let X be the random variable which keeps track of the number of heads obtained in the two tosses. Then clearly, $X \in \{0, 1, 2\}$. Since the coin is fair, each of the four simple events in S is equally likely, and the probability distribution of X may be displayed as

$X = x$	0	1	2
$p(x)$	1/4	1/2	1/4

Table 2.1.1. The probability distribution of X, the number of heads in two tosses of a fair coin.

■

The function p(x) in Table 2.1.1 is usually referred to as the *probability mass function* (or pmf) of X. We will use that phrase when describing the probabilities associated with any discrete random variable. The word "mass" refers to the weight (or probability) assigned to given values of X.

Example 2.1.2. Suppose that two *balanced* (6-sided) dice are rolled and we are interested in the digits facing upwards. If X is equal to the sum of the digits facing up, then it is clear that $X \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. The probability distribution of X is given by

$X = x$	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Table 2.1.2. The pmf of X, the sum of the digits facing up when two balanced dice are rolled.

■

Example 2.1.3. Let us suppose that five cards are drawn at random, without replacement, from a 52-card deck. Let X be the number of spades obtained. Clearly, X can only take the values in the set $\{0, 1, 2, 3, 4, 5\}$. The probability of each of the six possible values of X can be computed and then displayed in a table as in the two examples above. However, having studied some standard counting formulas in Chapter 1, we may turn to these formulas to write a general expression of the probability of any observed value of X. Recall that, when cards are sampled without replacement, all possible five-card poker hands are equally likely. In addition, note that there are

a) $\binom{52}{5}$ ways of picking 5 cards without replacement from the 52 cards,

b) $\binom{13}{x}$ ways of picking x spades without replacement from the 13 spades,

c) $\binom{39}{5-x}$ ways of picking $5-x$ cards without replacement from the 39 non-spades.

It thus follows from the basic rule of counting that the probability that $X=x$ is given by

$$P(X=x) = \frac{\binom{13}{x}\binom{39}{5-x}}{\binom{52}{5}} \text{ for } x = 0,1,2,3,4,5.$$

This is the first occurrence of a very useful discrete probability model called the "hypergeometric" distribution. We will treat this distribution in some detail later in this chapter. ■

A probability histogram (PH) is a useful visual tool that may be employed in tandem with the table or formula that defines a probability mass function. The probability histograms associated with the pmfs displayed in the Examples 2.1.1–2.1.3 are shown in the figures below. The x-axis shows the possible values of a random variable X, and the graph shows a set of rectangles whose heights are the probabilities of the respective values of X.

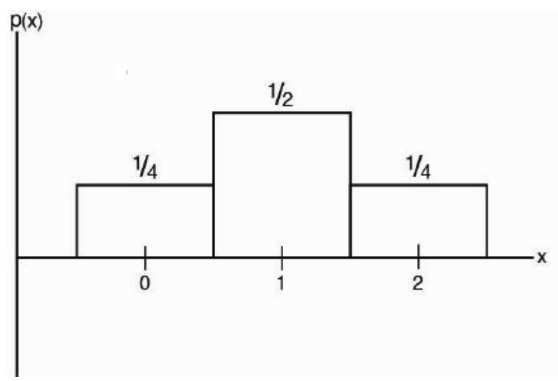

Figure 2.1.1. The PH for X = number of heads in 2 tosses of a fair coin.

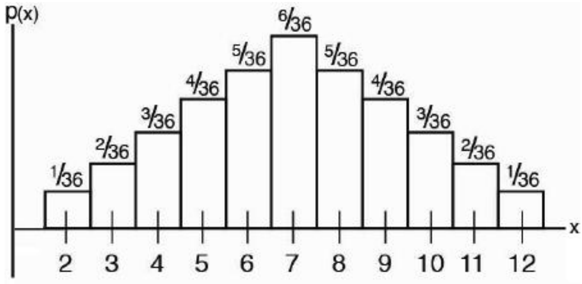

Figure 2.1.2. The PH for X = sum of digits facing up if 2 balanced dice are rolled.

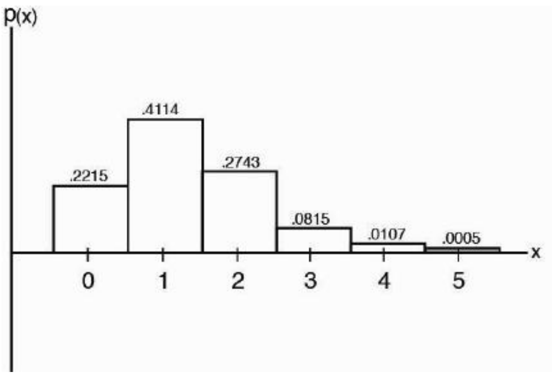

Figure 2.1.3. The PH for X = the number of spades in a random 5-card poker hand.

Exercises 2.1.

1. Suppose two balanced dice are rolled. Let X be the largest digit facing up. (If the digits facing up are both equal to the value c, then $X = c$.) Obtain the probability mass function of X.

2. Suppose a fair coin is tossed until the outcome "heads" (H) is obtained. Let X be the number of "tails" (T) obtained before the first H occurs, and let Y be the number of trials needed to get the first H. Find the probability mass functions $p_X(x)$ of X and $p_Y(y)$ of Y. What is the algebraic relationship between X and Y.

3. The Board of Directors of the MacArthur Foundation consists of 6 males and 4 females. Three members are selected at random (obviously without replacement) to serve on the Foundation's annual spring picnic committee. Let X be the number of females selected to serve on the committee. Obtain the p. m. f. of the random variable X.

4. Suppose that a fair coin is tossed four times. Let X be the number of heads obtained. Find the pmf of X. (Hint: Start with a probability tree.)

2.2 Mathematical Expectation

One numerical summary of the distribution of a random variable X that is quite meaningful and very widely used is the *mathematical expectation* of X. Often, this numerical summary is referred to as the *expected value* of X and is denoted by $E(X)$ or simply by EX. As the name suggests, this number represents the value we would "expect" X to be if the random experiment from which X is derived were to be carried out. We recognize, of course, that the actual value of the X that we observe might be larger or smaller than EX, but EX is meant to be interpreted as what we would expect X to be, "on the average," in a typical trial of the experiment. The expected value of any discrete random variable X may be computed by the following formula.

Definition 2.2.1. If X is a discrete random variable with probability mass function $p(x)$, then the expected value of X is given by

$$EX = \sum_{\text{all } x} xp(x). \tag{2.1}$$

Technically, the summation in (2.1) is taken over all x for which $p(x) > 0$, though the summation is also correct as written since the summand is zero for any x with $p(x) = 0$. The expected value of a random variable X is often called its mean and denoted by μ_X or simply by μ. Let us compute the mean or expected value of the variable X in Example 2.1.1. Recall that X is the number of heads obtained in two tosses of a fair coin. We thus have

$$EX = 0(1/4) + 1(1/2) + 2(1/4) = 1.$$

Intuitively, this answer seems appropriate since if we did do this experiment many times, we'd expect that get, on average, about 1 head per trial (that is, one head from the two coin tosses). This initial computation already suggests that we might be able to obtain EX "by inspection," that is, without any arithmetic, in certain kinds of problems. It is worth noting that the value taken on by EX need not be a value that the variable X can actually take on. For example, if a coin with $P(\text{heads}) = .4$ was tossed twice, it is easy to confirm that the expected value of X, the number of heads obtained, is $EX = .8$.

There are two common interpretations of EX that can occasionally be of help in computing EX and can also lend insight in certain types of applications (notably, situations involving games of chance). The first of these is the "center of gravity" interpretation. EX is supposed to be a sort of average value and, as such, should fall somewhere near the center of the values that the random variable X can take on. But as the formula in Equation (2.1) shows, it's not a pure average of the possible values of X. Rather, it is a weighted average of those values, with the weights based on how likely the various possible values of X are. Now, picture the probability histogram in Figure 2.1.1 as an object made out of wood, and ask where the balancing point would be if you were to try to lift the entire object with one finger. It turns out that if $EX < \infty$, then it will always be that balancing point. For the histogram in Figure 2.1.1, you could lift it with one finger if that finger was placed at the value 1 on the x-axis. If $EX < \infty$, the center of gravity of the probability histogram of a random variable X will be at the value EX. This interpretation (and property) of EX tells you something useful about computing EX. If the probability histogram is perfectly symmetric about a central point and if $EX < \infty$, then that central point is EX. The idea of the mean value of a random variable being infinite may seem difficult to imagine, but we'll see soon that this can happen. If it does, the center of gravity interpretation of EX goes out the window. Fortunately, this circumstance occurs fairly rarely.

If you look at the probability histogram in Figure 2.1.2, you can see that the mean EX of X is equal to 7 without actually computing it. You could verify this by applying the formula in (2.1) to the pmf given in Table 2.1.2. However, the symmetry of this pmf, together with the fact that $2 \leq X \leq 12$ is enough to assure you that EX is finite and that it is equal to the central value 7. You'll note, though, that a symmetry argument does not apply to the computation of EX for the random variable X in Example 2.1.3. There, you'll simply have to grind out the fact that $EX = 1.25$. (In Section 2.3, we'll derive a general formula for obtaining EX in probability models of this type.)

I mentioned above that there is a second interpretation of EX which has special utility. Consider a random experiment which yields a random variable X of interest, and suppose that this experiment is performed repeatedly (say, n times). Let $X_1, X_2, \ldots, X_n$ be the observed values of the random variable X in the n repetitions of the experiment. Denote the average of these n values of X as

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{2.2}$$

As n grows to ∞, the average or "sample mean" $\bar{X}_n$ will converge to the value EX. The sense in which $\bar{X}_n$ converges to EX can be made explicit and mathematically rigorous (with terms like "convergence in probability" or "convergence almost surely") and indeed this is done in Chapter 5, but for now, we will treat this convergence intuitively, simply writing $\bar{X}_n \to EX$ as $n \to \infty$, and interpreting this convergence as signifying, simply, that $\bar{X}_n$ gets closer and closer to EX as n grows. Interestingly, this holds true even when EX is infinite, in which case, the values of $\bar{X}_n$will grow without bound as $n \to \infty$.

This second interpretation of the expected value of X has some interesting implications when one considers "games of chance." Many games we play have a random element. For example, most board games rely on the spinning of a wheel or the roll of a die. The outcomes in the game of poker depend on random draws from a deck of cards. Indeed, virtually all the games one finds in Las Vegas (or all other) casinos depend on some random mechanism. Let us focus, for a moment, on games of chance in which a wager is offered. Let us refer to the two actors involved in the wager as the player and the house. A game involving a wager will typically involve some investment on the player's part (in dollars, let us say) and depending on the outcome of a chance mechanism, the player will either retain his/her investment and receive a payment from the house or will lose the investment to the house. In the end, a given play of this game will result in a random variable W that we will refer to as the player's winnings. The variable W may be *positive or negative*, with negative winnings representing losses to the player. Given this framework, we will evaluate the game by means of computing and interpreting EW. The expectation EW represents the long-run average winnings for the player after many repetitions of the game. (Technically, it is the limiting average winnings as the number of repetitions n tends to ∞.)

Definition 2.2.2. Consider a game between two adversaries—the player and the house—and suppose that the random variable W represents the player's winnings on a given play of the game. Then, the game is said to be a *fair game* if $EW = 0$. If $EW > 0$, the game is said to be *favorable to the player* and if $EW < 0$, the game is said to be *unfavorable to the player*.

Example 2.2.1. Years ago, I was sitting at home one evening watching the 11 o'clock news. My shoes had been kicked off a couple of hours earlier, and I was enjoying a glass of wine as the evening softly wound its way toward its inevitable end. Suddenly, the phone rang. It was a young man—a former student at Davis—who had a probability question for me. I asked him if he was a former student of mine in introductory statistics. He was impressively honest, admitting that actually, he had taken the course from another instructor and had called her first, but she wasn't answering her phone. Here was his dilemma. He was in a casino in South Lake Tahoe. He had been playing at the craps table for a while, and he was winning! He felt that he had discovered a bet that was favorable to the player, but before he invested too heavily in this wager, he thought he should verify his findings by talking to someone like me. As it happens, one of the wagers available at the craps table is *The Field Bet*. Now there are different versions of this bet out there, but most casinos use the version he described to me that night. The "field" is a particular collection of outcomes for the sum of the digits facing up when two dice are rolled. Specifically, the field F is equal to the set of outcomes $\{5,6,7,8\}$. In the field bet, you would put down a bet (say \$1.00) and would lose your bet if the outcome X was in the field. Otherwise, you would get to keep your bet and would also receive a payoff. Assuming a \$1.00 bet, your possible winnings, and the events associated with each, are shown below. As you can see, the house would sweeten the pot if you rolled a 2 or a 12.

Win W	-1	1	2	3
if	$X \in F$	$X \in \{3,4,9,10,11\}$	$X = 2$	$X = 12$

Table 2.2.1. The possible values of W in the Field Bet when \$1.00 is wagered.

The resulting probability distribution of W, assuming the dice used are balanced, is

$W = w$	-1	1	2	3
$p(w)$	20/36	14/36	1/36	1/36

Table 2.2.2. The probability distribution of W in the Field Bet.

After obtaining this distribution for my young friend, I calculated EW for him as

$$EW = \frac{20}{36}(-1) + \frac{14}{36}(1) + \frac{1}{36}(2) + \frac{1}{36}(3) = -\frac{1}{36}.$$

This finding was shocking and disappointing to the young man, especially because when he did this computation in his head earlier in the evening, he had gotten a different answer. But he had made an arithmetic mistake. In Table 2.2.2, he had mistakenly calculated $p(-1)$ as 19/36 and $p(1)$ as 15/ 36. As a result, what he thought of as a favorable bet was actually unfavorable to him. But there was an obvious silver lining. First, as luck would have it, he was ahead at the time. Secondly, he now knew that if he continued to make this bet repeatedly, he would lose, on average, about 3 cents per play and would eventually lose all his money. I felt good about saving him from this grim fate. He thanked me for my time, and no doubt resumed his determined search for a bet that would actually favor him. ■

It is perhaps worth mentioning that there are bets at all standard casinos that actually favor the player over the house. One example is the game of blackjack. As it happens, how well you do depends on your playing strategy. It is possible for a player using "the optimal strategy" to win, on the average, that is, to turn blackjack into a favorable game for the player. Good strategies in blackjack generally involve "card counting," that is, some form of accounting for the cards that have been seen as dealing proceeds from a shuffled deck of cards (or from a "boot" from which cards from several shuffled decks are dealt). It also pays to vary the size of your bets according to whether or not the current status of the deck favors you. You might wonder why casinos would offer a wager that is stacked against them. The answer is simple. Rather few players use anything close to the optimal strategy. Most players lose at blackjack, and casinos make tons of money from this game. So they don't mind losing a bit to the occasional "card counter," though most casinos reserve (and exercise) the right to ask you to leave the premises if they think you are "card counting" or otherwise take a dislike to you.

Since it's not my purpose here to turn my readers into gamblers, I'll fight off the temptation to tell you more. Instead, I'll give another example—a famous and rather intriguing example of a game in which a player's expected winnings are infinite, yet people who want to play the game are hard to find.

Example 2.2.2. (The St. Petersburg Paradox). Consider a game in which a fair coin is tossed repeatedly until it comes up heads. Let X be the number of trials that it takes to get the first heads. It is easy to confirm that the pmf of X is given by

$$P(X = x) = (1/2)^x \text{ for } x = 1, 2, 3, \dots \tag{2.3}$$

Suppose that the house agrees to pay you $\$2^x$ if $X = x$ is the outcome in your play at the game. The probability distribution of your payoff Y in this game is:

$Y = y$	2	4	8	16	32	64	128	256	...
$p(y)$	1/2	1/4	1/8	1/16	1/32	1/64	1/128	1/256	...

Table 2.2.3. The distribution of the payoff Y in the St. Petersburg game.

The expected payoff from this game is thus

$$EY = \frac{1}{2}(2) + \frac{1}{4}(4) + \frac{1}{8}(8) + \frac{1}{16}(16) + \cdots = 1 + 1 + 1 + 1 + \cdots = \infty.$$

Given that the expected payoff is infinity, you should be willing to wager virtually all your money to play this game. Suppose that the house is willing to let you play for the piddling sum of \$1000.00, and that you happen to have access to that amount of cash. The house gets to keep your \$1000, whatever your outcome Y is. Would you take the bet? Remember that $\infty - 1000$ is still ∞, so your expected winnings would be infinite. Here's the problem, and the paradox. If you pay \$1000 to play this game, there is a very high probability that you'll lose money. Indeed the probability that the payoff will be larger than what you paid to play is $1/2^{10}$, or put another way, the chances that you lose money on your wager is $1023/1024 > .999$. Hey, what's up with that? ■

The idea of mathematical expectation is quite a bit more general than the ideas we have discussed thus far. The expected value of a random variable X tells us where the distribution of X is centered, but there is a good deal more about that distribution that we might be curious about. One thing we might ask about is the "spread" of the distribution—is it distributed quite tightly about EX or is the distribution more spread out. The term *dispersion* is often used in discussions about how spread out a probability distribution is. Also, one often says that one distribution is less diffuse than another if the first is more concentrated around its center than the second. In studying the dispersion of a random variable or of its distribution, we will consider the expectation of variables other than X itself; for example, we will be interested in the size of the expected value of X^2. At first glance, it would appear that calculating EX^2 requires that we obtain the probability distribution of the variable $Y = X^2$. If we have $p(y)$ in hand, then we can compute

$$EX^2 = EY = \sum_{\text{all y}} yp(y).$$

But this expectation can actually be computed more simply. It can be proven that one can obtain this expectation without actually identifying the distribution of X^2, that is, we may compute EX^2 as

$$EX^2 = \sum_{\text{all x}} x^2 p(x).$$

The theorem that justifies this simpler computation (for discrete random variables) is stated below without proof.

Theorem 2.2.1. Let X be a discrete random variable, and let $p(x)$ be its pmf. Let $y = g(x)$. If $Eg(X) < \infty$, then this expectation may be calculated as

$$Eg(X) = \sum_{\text{all x}} g(x)p(x).$$

We now define a number of other expected values of interest. The first involves the expectation of a useful class of random variables associated with a given random variable X—the class of linear functions of X. The following simple result will often prove useful.

Theorem 2.2.2. Let X be a discrete random variable with finite mean EX. Let a and b be real numbers, and let $Y = aX + b$. Then $EY = aEX + b$.

Proof. Using Theorem 2.2.1 to evaluate EY, we have

$$\begin{aligned}
EY &= \sum_{\text{all x}} (ax+b)p(x) \\
&= \sum_{\text{all x}} axp(x) + \sum_{\text{all x}} bp(x) \\
&= a\sum_{\text{all x}} xp(x) + b\sum_{\text{all x}} p(x) \\
&= aEX + b.
\end{aligned}$$

■

Definition 2.2.3. Let X be a discrete random variable with mean μ. The variance of X, denoted interchangeably by $V(X)$, σ_X^2 or when there can be no confusion, σ^2, is defined as the expected value

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

The variance of X (or of the distribution of X) is a measure of how concentrated X is about its mean. Since the variance measures the expected value of the squared distance of a variable X from its mean, large distances from the mean are magnified and will tend to result in a large value for the variance. Let us illustrate the computation of the variance using the simple random variable X treated in Example 2.2.1, that is, for $X =$ the number of heads in two tosses of a fair coin. For this particular variable X,

$$\sigma_X^2 = \sigma^2 = E(X-\mu)^2 = (0-1)^2\frac{1}{4} + (1-1)^2\frac{1}{2} + (2-1)^2\frac{1}{4} = \frac{1}{2}. \tag{2.4}$$

Besides the variance, which informs us about the spread of a probability distribution, there are a number of other expected values that provide meaningful information about the character or shape of a distribution. One class of such measures are called *moments* and are defined as follows:

Definition 2.2.4. For positive integers $k = 1,2,3\ldots$ the kth moment of a random variable X is the expected value

$$EX^k = \sum_{\text{all x}} x^k p(x).$$

As we discuss in more detail later, the moments of a random variable have a close connection to the shape of its distribution. The variance of a variable X can be obtained from its first two moments, as shown in the following result.

Theorem 2.2.3. For any discrete variable X for which $\sigma_X^2 < \infty$, the variance of X may be computed as

$$\sigma_X^2 = EX^2 - (EX)^2.$$

Proof. Write $EX = \mu$. Then

$$\begin{aligned}
\sigma_X^2 &= \sum_{\text{all x}} (x-\mu)^2 p(x) \\
&= \sum_{\text{all x}} (x^2 - 2\mu x + \mu^2) p(x) \\
&= \sum_{\text{all x}} x^2 p(x) - 2\mu \sum_{\text{all x}} x p(x) + \mu^2 \sum_{\text{all x}} p(x) \\
&= EX^2 - 2\mu(\mu) + \mu^2 \\
&= EX^2 - \mu^2.
\end{aligned}$$

■

The alternative formula given in Theorem 2.2.2 is useful when computing $V(X)$ using a hand-held calculator, as it requires fewer arithmetic operations than the original formula. One further note regarding the spread of a distribution is that the *standard deviation* σ_X, which is the square root of the variance, is the most frequently used measure of spread or dispersion. Its main virtue is that σ_X is a measure that is in the same units of measurement as the X values themselves. For example, if the Xs are measured in feet, then the standard deviation is measured in feet. Because variables are squared in the computation of the variance, this measure is not in the same units of measurement as the data themselves. For example, when we wish, later on, to discuss intervals of real numbers that are thought to contain the value of an unknown mean of a random variable with a certain specified probability, we will find it natural and convenient to use intervals of the form $X \pm k\sigma_X$; for example, if X is measured in feet, then such an interval says that the mean of X is thought to be within $k\sigma_X$ feet of the value X feet.

There are other expected values that are used to investigate specific aspects of a probability distribution. For example, one is sometimes interested in measuring the amount of asymmetry in a distribution. The term "skewness" is used to describe asymmetry. A discrete distribution is said to be *skewed to the right* if the right-hand tail of its probability histogram is longer and heavier than its left-hand tail, with the opposite circumstance described as skewness to the left. The third moment of a distribution is closely related to its skewness. The population skewness is defined $E(X-\mu)^3/\sigma^3$. The intuition behind this measure is that when the distribution of X is skewed to the right, say, the large values to the right of the mean receive considerable weight because they are cubed, increasing their size, and receive more weight because the set of values to the right of the mean have higher probability of occurrence than the values to the left of the mean, a consequence of the fact that the median of a distribution that is skewed to the right will be larger than the distribution's mean. In a similar vein, the fourth moment of a random variable X provides some indication of the amount of "peakedness" of a distribution (the sharpness of the peak near its mean). A commonly used measure of peakedness is the so-called population "kurtosis" defined as $\kappa = E(X-\mu)^4/\sigma^4$.

Exercises 2.2.

1. Two balanced dice are rolled. Let X be the value facing up on the green die and Y be the value facing up on the red die. Let $Z = \max\{X, Y\}$, that is, the largest of these two values. Compute EZ and Var(Z).

2. The Dixon Stamp Collecting Club is throwing a fund raiser, and is offering eight door prizes—2 rare stamps worth \$20 each and 6 common stamps worth \$1 each. The number on your ticket has qualified you to pick two of the prizes at random. (You lucky dog!)

   a) Calculate the probability that the prizes you win are worth \$2, \$21, or \$40.
   b) Let X be the total value of your two prizes. Calculate $E(X)$.

3. NBA player Hunk Halloran is a great rebounder but a pretty average free-throw shooter. Suppose that X represents the number of free throws that Hunk will make out of 5 attempts. The probability distribution of X is given by:

X	0	1	2	3	4	5
$P(X)$	1/21	2/21	3/21	4/21	5/21	6/21

   Calculate the mean and variance of the random variable X.

4. You are dealt two cards without replacement from a standard 52-card deck. The dealer will pay you \$20 if the cards match, \$5 if the second card is larger than the first (with ace taken as "high"), and \$1 if the second card is smaller than the first. What should you be asked to pay to play this game if the game is intended to be a "fair game"?