2.3.2 Three Useful Distributions

This section examines three distributions which are useful in simulations: the uniform, normal, and binomial.

The Uniform Distribution. There are many situations in which the outcomes of an experiment are equally likely. Let us consider the following two scenarios:

1. Consider the interval $[0, 1]$. If you drop a pin on this interval, the point x at which the tip of the pin lands could be any value between 0 and 1. Further, since we can not predict where the tip of the pin lands, each of these possible values of x would have the same chance of being selected, namely zero, since there are an uncountable number of numbers in that interval. This scenario is the same as describing the "at random" choice of a number between 0 and 1.

2. A bank opens at 8:00 am. The first customer arrives at random within the first 10 minutes that the bank is open. How can we describe the possible arrival times of the first customer? Since the customer could arrive at any time between 8:00 and 8:10 am, our description must include all these values (i.e., the 10-minute interval $[0, 10]$).

The frequency distribution for a continuous random variable where intervals of equal length are equally likely is known as the *uniform distribution*. A graph of a uniform distribution is shown in Figure 2.8. The general form of the uniform density function is

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{for } \alpha < x < \beta, \\ 0 & \text{elsewhere} \end{cases} \tag{27}$$

Notice that the non-zero part of this function is a horizontal line segment, and the area under this segment is one. Since intervals of equal length are equally probable, the mean of the uniform distribution on (α, β) is the midpoint of the interval, which is $(\alpha + \beta)/2$. In Exercise 7 it is shown that the variance is $(\beta - \alpha)^2/12$.

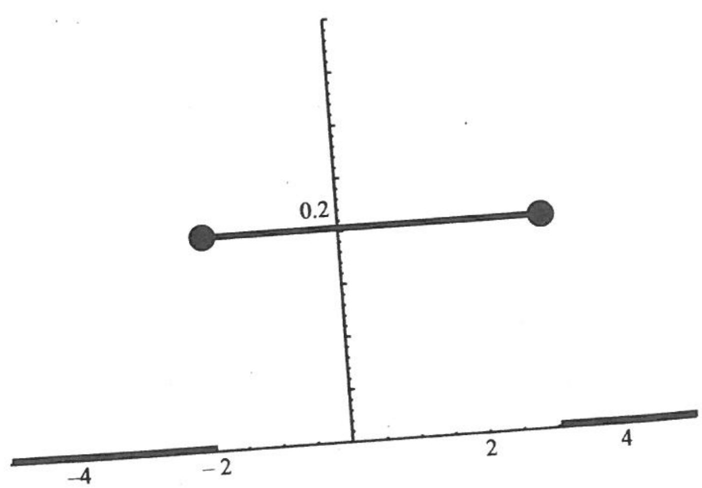

FIGURE 2.8 The Uniform Distribution on $[-2, 3]$.

Example: Suppose a random variable is uniformly distributed on the interval $[-2, 3]$. The graph of the density function is shown in Figure 2.8.

To compute a probability for a uniform random variable, we need only to find the area of the appropriate rectangle. For example, to compute the probability that the value of x is between -1 and 1.5, written $P(-1 < X < 1.5)$, we simply compute the area of the enclosed rectangle; that is,

$$P(-1 < X < 1.5) = (1.5 + 1)\frac{1}{(3+2)} = 0.5.$$

Similarly, to find the probability of values between 0 and 2.5,

$$P(0 < X < 2.5) = (2.5 - 0)\frac{1}{(3+2)} = 0.5.$$

Notice that these intervals $(-1, 1.5)$ and $(0, 2.5)$ are the same lengths and, hence, have the same probabilities.

The Normal Distribution. We have previously seen bell-shaped frequency distributions in connection with empirically obtained data. Frequency distributions and histograms that are bell-shaped are often represented by a *normal distribution.* It is a fact that measurements from many random variables appear to have been generated from frequency distributions that are closely approximated by a normal probability distribution. For this reason the normal distribution is considered the most important probability distribution. The normal distribution has the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for } \sigma > 0, \; -\infty < \mu < \infty, \; -\infty < x < \infty. \tag{28}$$

The normal distribution function depends upon the location and shape parameters μ and σ. The choice of symbols for these parameters is not an accident. It is an exercise to show that μ and σ^2 are the mean and variance of a normal random variable. The graph of a normal distribution in Figure 2.9 shows the characteristic bell-shape.

This curve also shows other important features. Note that $f(x)$ obtains its maximum at the point $x = \mu$, and that the curve is symmetric about $x = \mu$. The function is decreasing as x moves away from μ. It is a worthwhile exercise (see Exercise 8) to show that the

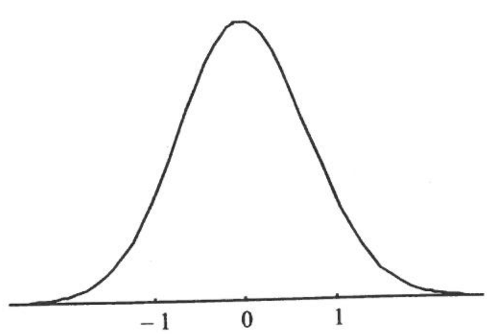

FIGURE 2.9 The Normal Distribution with $\mu = 0$ and $\sigma^2 = 1$.

normal distribution (28) obtains its maximum at $x = \mu$ and that it has inflection points at $x = \mu \pm \sigma$.

An often cited rule of thumb for determining if a data set is from a normal distribution is the "68 - 95 - 99" rule. It states that for a normal distribution, approximately 68% of the data (area) is within one standard deviation of the mean; that is,

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827. \tag{29}$$

Similarly, two standard deviations from the mean, $\mu \pm 2\sigma$, contains approximately 95% of the data. Three standard deviations from the mean, $\mu \pm 3\sigma$, contains approximately 99% of the data. A large data set is approximately normally distributed if one, two, and three standard deviations about the sample mean contain approximately these percentages of data.

The Binomial Distribution. Some experiments consist of the observation of a trial which results in one of two possible outcomes. For instance, a coin flip lands either heads up or tails up, a manufactured item is either defective or non-defective, or a free throw either hits or misses the basket. These are just some examples of a Bernoulli random variable which results when an experiment has only two outcomes, often labeled "success" or "failure."

More specifically, a random variable X that assumes only the values 0 or 1 is known as a *Bernoulli variable*, and the performance of an experiment with only two possible outcomes is called a *Bernoulli trial*. A Bernoulli random variable X is a discrete random variable since it has only two possible values. The Bernoulli distribution is given by

$$f(x) = p^x q^{1-x} \quad \text{for } x = 0, 1 \tag{30}$$

where $p = P[X = 1]$ is the probability of a "success" and $q = 1 - p = P[X = 0]$ is the probability of a "failure."

Notice that $E(X) = 0 \cdot q + 1 \cdot p = p$ and $E(X^2) = 0^2 \cdot q + 1^2 \cdot p = p$, so that the variance of X is $V(X) = p - p^2 = p(1 - p) = pq$.

An important assumption often made when performing Bernoulli trials is that successive trials are *independent*; that is, the probability of the occurrence of a success or failure on a trial does not depend upon the outcome of the previous trial (or trials).

When a sequence of n independent Bernoulli trials are conducted, an important distribution arises from counting the number of successes that occur. Let Y be the number of successes that occur on n independent Bernoulli trials. Since successes or failures occur randomly, Y is a random variable with possible values $0, 1, \ldots, n$. The distribution of Y is known as the *binomial* distribution and has the form

$$f(k) = \binom{n}{k} p^k q^{n-k} \text{ for } k = 0, 1, \ldots, n. \tag{31}$$

The expression $\binom{n}{k}$ is known as the *binomial coefficient* or the *combination number*, and is defined by

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}. \tag{32}$$

The symbol "!" represents the *factorial*, and is defined for each positive integer n by

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1. \tag{33}$$

As a convention, we define

$$0! = 1. \tag{34}$$

Note that the binomial distribution is completely determined by the parameters n and p and that a sum of n independent Bernoulli random variables is binomially distributed. See Exercise 9.

To understand the binomial distribution, we first consider the binomial coefficient. The binomial coefficient is a device for counting the number of ways k items can be chosen from n items where order is unimportant. For example, $\binom{52}{5}$ is the number of 5 card hands that can be dealt from a 52-card deck.

As another example, the number of ways that two heads can occur on four coin tosses is

$$\binom{4}{2} = \frac{4!}{(4-2)!2!} = 6.$$

We easily verify this value by listing all of the possible arrangements of two heads on four tosses. If we let "H" denote heads and "T" denote tails, then two heads occur on four coin tosses in the following six ways: HHTT, HTHT, HTTH, THTH, TTHH, and THHT.

Continuing this example, we compute the probability of two heads's occurring on four coin tosses. Since a (fair) coin has probability $p = 1/2$ of landing heads and $q = 1/2$ of landing tails, each of the six outcomes listed above has probability

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$$

of occurrence. Since these six outcomes are the only ways to have two heads occur on four coin tosses, we have

$$P[\text{ Two heads on four coin tosses }] = 6 \times \frac{1}{16}.$$

Now this is the same value we obtain from the binomial distribution, for if Y represents the number of heads which occur on $n = 4$ coin tosses, then

$$P[Y = 2] = \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2} = \frac{6}{16}.$$

The binomial distribution does not require that $p = q$ as is the case with fair coin flips. As another example, suppose that a manufacturing process produces 5% defective light bulbs. A box of $n = 20$ light bulbs will contain two defective bulbs with probability

$$\binom{20}{2} (0.05)^2 (0.95)^{18} = 0.1887.$$

In light of the above discussion, it is useful to note the form of this expression. The coefficient $\binom{20}{2}$, which is the number of ways to find two defective bulbs in a box of 20, is multiplied by the probability $(0.05)^2$ of two defectives and then multiplied by $(0.95)^{18}$, the probability of 18 nondefective bulbs.

In this last example, if a there is a 5% chance that a light bulb will be defective, then it is reasonable to expect that a box of 20 light bulbs would contain $20 \times 0.05 = 1$ defective bulb. This is indeed the case. Exercise 9 asks the reader to show that the mean of a binomial random variable Y is

$$E(Y) = np, \tag{35}$$

and the variance of Y is

$$V(Y) = npq. \tag{36}$$

2.3.3 Environmental versus Demographic Stochasticity

We began this chapter with an example of a hypothetical declining squirrel population. We assumed that the growth rate of -0.5 is actually an average growth rate for that population. These considerations led us to a brief study of some of the ideas of probability and statistics. In this section, we model a population using the tools we developed.

There are many reasons for variations in birth or death rates. Variations resulting from slight changes in the population's behavior or structure are called variations due to *demographic* stochasticity. Variations from environmental conditions (flood, drought, fire, etc.) are called variations due to *environmental* stochasticity.

Since each year the numerical values of the birth and death rates vary about some central value and values far from this central value are rare, we frequently assume that demographic stochasticity follows a normal distribution. Environmental stochasticity can take many forms. We will consider it to be a catastrophe that randomly affects the population. For instance, a flood may occur on average once every 25 years or 4% of the time. We incorporate environmental stochasticity as a Bernoulli random variable; either a catastrophe occurs, or it does not. Most software has uniform and normal random numbers built in, but a Bernoulli random number generator is rare. We can, however, construct a Bernoulli random number from a uniform random number chosen between $[0, 1]$. Suppose we want a Bernoulli random number which is a success 4% of the time. We pick a uniform random number between 0 and 1. If the number is less than 0.04, then we consider it a success; if it is greater than 0.04, we consider it a failure.

Stochasticity and the Sandhill Crane. We have previously modeled the population of the Florida sandhill crane. We considered the population under the three different growth rates determined by the best, medium, and worst environmental conditions. Each of these conditions gave a fixed average value of the growth rate r. In this example, we use both birth and death rates for the cranes. The average reproduction rate of the sandhill crane is 0.5, while the average death rate is 0.1. Demographic stochasticity affects the birth and death rates. Here we assume that the birth and death rates are normally distributed with means of 0.5 and 0.1 respectively. The standard deviations of these birth and death rates are 0.03 and 0.08 respectively. We further assume that a catastrophe will occur on