Chapter 5

SPECIAL RANDOM VARIABLES

Certain types of random variables occur over and over again in applications. In this chapter, we will study a variety of them.

5.1 THE BERNOULLI AND BINOMIAL RANDOM VARIABLES

Suppose that a trial, or an experiment, whose outcome can be classified as either a "success" or as a "failure" is performed. If we let $X = 1$ when the outcome is a success and $X = 0$ when it is a failure, then the probability mass function of X is given by

$$\begin{aligned} P\{X = 0\} &= 1 - p \\ P\{X = 1\} &= p \end{aligned} \tag{5.1.1}$$

where p, $0 \leq p \leq 1$, is the probability that the trial is a "success."

A random variable X is said to be a Bernoulli random variable (after the Swiss mathematician James Bernoulli) if its probability mass function is given by Equations 5.1.1 for some $p \in (0, 1)$. Its expected value is

$$E[X] = 1 \cdot P\{X = 1\} + 0 \cdot P\{X = 0\} = p$$

That is, the expectation of a Bernoulli random variable is the probability that the random variable equals 1.

Suppose now that n independent trials, each of which results in a "success" with probability p and in a "failure" with probability $1 - p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a *binomial* random variable with parameters (n, p).

The probability mass function of a binomial random variable with parameters n and p is given by

$$P\{X = i\} = \binom{n}{i} p^i(1-p)^{n-i}, \quad i = 0, 1, \ldots, n \tag{5.1.2}$$

where $\binom{n}{i} = n!/[i!(n-i)!]$ is the number of different groups of i objects that can be chosen from a set of n objects. The validity of Equation 5.1.2 may be verified by first noting that the probability of any particular sequence of the n outcomes containing i successes and $n - i$ failures is, by the assumed independence of trials, $p^i(1-p)^{n-i}$. Equation 5.1.2 then follows since there are $\binom{n}{i}$ different sequences of the n outcomes leading to i successes and $n-i$ failures — which can perhaps most easily be seen by noting that there are $\binom{n}{i}$ different selections of the i trials that result in successes. For instance, if $n = 5$, $i = 2$, then there are $\binom{5}{2}$ choices of the two trials that are to result in successes — namely, any of the outcomes

$$\begin{array}{lll}
(s,s,f,f,f) & (f,s,s,f,f) & (f,f,s,f,s) \\
(s,f,s,f,f) & (f,s,f,s,f) & \\
(s,f,f,s,f) & (f,s,f,f,s) & (f,f,f,s,s) \\
(s,f,f,f,s) & (f,f,s,s,f) &
\end{array}$$

where the outcome (f,s,f,s,f) means, for instance, that the two successes appeared on trials 2 and 4. Since each of the $\binom{5}{2}$ outcomes has probability $p^2(1-p)^3$, we see that the probability of a total of 2 successes in 5 independent trials is $\binom{5}{2}p^2(1-p)^3$. As a check, note that, by the binomial theorem, the probabilities sum to 1; that is,

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{n} \binom{n}{i} p^i(1-p)^{n-i} = [p + (1-p)]^n = 1$$

The probability mass function of three binomial random variables with respective parameters (10, .5), (10, .3), and (10, .6) are presented in Figure 5.1. The first of these is symmetric about the value .5, whereas the second is somewhat weighted, or *skewed*, to lower values and the third to higher values.

EXAMPLE 5.1a It is known that disks produced by a certain company will be defective with probability .01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective. What proportion of packages is returned? If someone buys three packages, what is the probability that exactly one of them will be returned?

SOLUTION If X is the number of defective disks in a package, then assuming that customers always take advantage of the guarantee, it follows that X is a binomial random variable

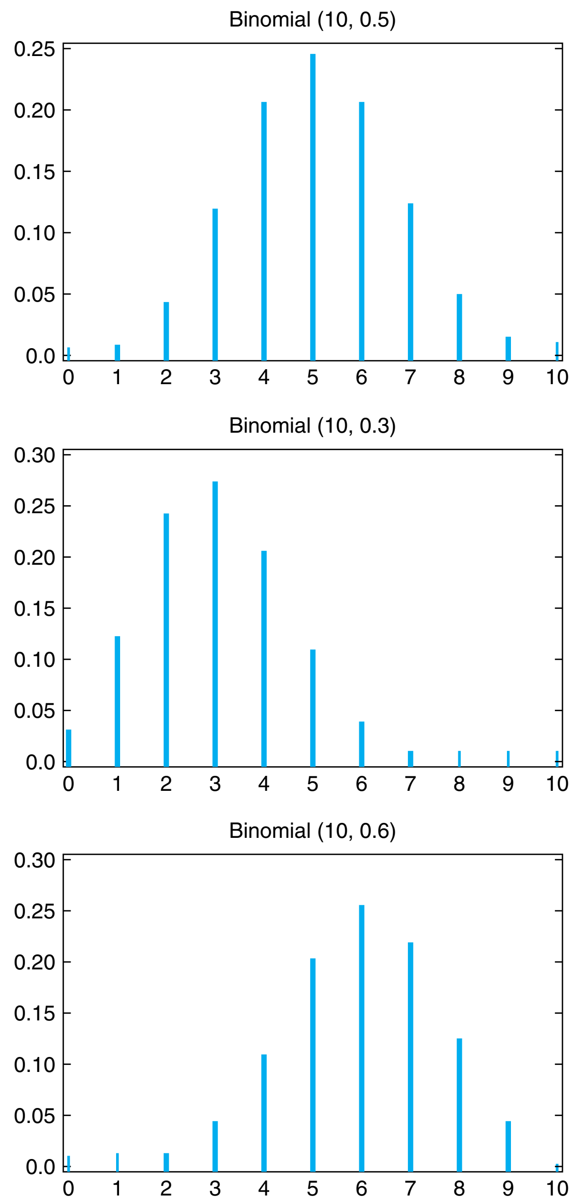


FIGURE 5.1 *Binomial probability mass functions.*

with parameters (10, .01). Hence the probability that a package will have to be replaced is

$$
\begin{aligned}
P\{X > 1\} &= 1 - P\{X = 0\} - P\{X = 1\} \\
&= 1 - \binom{10}{0}(.01)^0(.99)^{10} - \binom{10}{1}(.01)^1(.99)^9 \approx .005
\end{aligned}
$$

Because each package will, independently, have to be replaced with probability .005, it follows from the law of large numbers that in the long run .5 percent of the packages will have to be replaced.

It follows from the foregoing that the number of packages that will be returned by a buyer of three packages is a binomial random variable with parameters $n = 3$ and $p = .005$. Therefore, the probability that exactly one of the three packages will be returned is $\binom{3}{1}(.005)(.995)^2 = .015$. ■

EXAMPLE 5.1b The color of one's eyes is determined by a single pair of genes, with the gene for brown eyes being dominant over the one for blue eyes. This means that an individual having two blue-eyed genes will have blue eyes, while one having either two brown-eyed genes or one brown-eyed and one blue-eyed gene will have brown eyes. When two people mate, the resulting offspring receives one randomly chosen gene from each of its parents' gene pair. If the eldest child of a pair of brown-eyed parents has blue eyes, what is the probability that exactly two of the four other children (none of whom is a twin) of this couple also have blue eyes?

SOLUTION To begin, note that since the eldest child has blue eyes, it follows that both parents must have one blue-eyed and one brown-eyed gene. (For if either had two brown-eyed genes, then each child would receive at least one brown-eyed gene and would thus have brown eyes.) The probability that an offspring of this couple will have blue eyes is equal to the probability that it receives the blue-eyed gene from both parents, which is $\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{4}$. Hence, because each of the other four children will have blue eyes with probability $\frac{1}{4}$, it follows that the probability that exactly two of them have this eye color is

$$\binom{4}{2}(1/4)^2(3/4)^2 = 27/128 \quad ■$$

EXAMPLE 5.1c A communications system consists of n components, each of which will, independently, function with probability p. The total system will be able to operate effectively if at least one-half of its components function.

(a) For what values of p is a 5-component system more likely to operate effectively than a 3-component system?

(b) In general, when is a $2k + 1$ component system better than a $2k - 1$ component system?

SOLUTION

(a) Because the number of functioning components is a binomial random variable with parameters (n, p), it follows that the probability that a 5-component system will be effective is

$$\binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + p^5$$

whereas the corresponding probability for a 3-component system is

$$\binom{3}{2} p^2(1-p) + p^3$$

Hence, the 5-component system is better if

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 \geq 3p^2(1-p) + p^3$$

which reduces to

$$3(p-1)^2(2p-1) \geq 0$$

or

$$p \geq \frac{1}{2}$$

(b) In general, a system with $2k+1$ components will be better than one with $2k-1$ components if (and only if) $p \geq \frac{1}{2}$. To prove this, consider a system of $2k+1$ components and let X denote the number of the first $2k-1$ that function. Then

$$P_{2k+1}(\text{effective}) = P\{X \geq k+1\} + P\{X = k\}(1-(1-p)^2) + P\{X = k-1\}\,p^2$$

which follows since the $2k+1$ component system will be effective if either

(1) $X \geq k+1$;
(2) $X = k$ and at least one of the remaining 2 components function; or
(3) $X = k-1$ and both of the next 2 function.

Because

$$\begin{aligned} P_{2k-1}(\text{effective}) &= P\{X \geq k\} \\ &= P\{X = k\} + P\{X \geq k+1\} \end{aligned}$$

we obtain that

$$\begin{aligned} &P_{2k+1}(\text{effective}) - P_{2k-1}(\text{effective}) \\ &\quad = P\{X = k-1\}\,p^2 - (1-p)^2 P\{X = k\} \\ &\quad = \binom{2k-1}{k-1} p^{k-1}(1-p)^k p^2 - (1-p)^2 \binom{2k-1}{k} p^k (1-p)^{k-1} \\ &\quad = \binom{2k-1}{k} p^k (1-p)^k [p - (1-p)] \qquad \text{since } \binom{2k-1}{k-1} = \binom{2k-1}{k} \\ &\quad \geq 0 \Leftrightarrow p \geq \frac{1}{2} \quad \blacksquare \end{aligned}$$

EXAMPLE 5.1d Suppose that 10 percent of the chips produced by a computer hardware manufacturer are defective. If we order 100 such chips, will X, the number of defective ones we receive, be a binomial random variable?

SOLUTION The random variable X will be a binomial random variable with parameters (100, .1) if each chip has probability .9 of being functional and if the functioning of successive chips is independent. Whether this is a reasonable assumption when we know that 10 percent of the chips produced are defective depends on additional factors. For instance, suppose that all the chips produced on a given day are always either functional or defective (with 90 percent of the days resulting in functional chips). In this case, if we know that all of our 100 chips were manufactured on the same day, then X will not be a binomial random variable. This is so since the independence of successive chips is not valid. In fact, in this case, we would have

$$P\{X = 100\} = .1$$
$$P\{X = 0\} = .9 \quad \blacksquare$$

Since a binomial random variable X, with parameters n and p, represents the number of successes in n independent trials, each having success probability p, we can represent X as follows:

$$X = \sum_{i=1}^{n} X_i \tag{5.1.3}$$

where

$$X_i = \begin{cases} 1 & \text{if the } i\text{th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

Because the X_i, $i = 1, \ldots, n$ are independent Bernoulli random variables, we have that

$$\begin{aligned} E[X_i] &= P\{X_i = 1\} = p \\ \text{Var}(X_i) &= E[X_i^2] - p^2 \\ &= p(1-p) \end{aligned}$$

where the last equality follows since $X_i^2 = X_i$, and so $E[X_i^2] = E[X_i] = p$.

Using the representation Equation 5.1.3, it is now an easy matter to compute the mean and variance of X:

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} E[X_i] \\ &= np \end{aligned}$$

$$\begin{aligned}\text{Var}(X) &= \sum_{i=1}^{n} \text{Var}(X_i) \quad \text{since the } X_i \text{ are independent} \\ &= np(1-p)\end{aligned}$$

If X_1 and X_2 are independent binomial random variables having respective parameters (n_i, p), $i = 1, 2$, then their sum is binomial with parameters $(n_1 + n_2, p)$. This can most easily be seen by noting that because X_i, $i = 1, 2$, represents the number of successes in n_i independent trials each of which is a success with probability p, then $X_1 + X_2$ represents the number of successes in $n_1 + n_2$ independent trials each of which is a success with probability p. Therefore, $X_1 + X_2$ is binomial with parameters $(n_1 + n_2, p)$.

5.1.1 Computing the Binomial Distribution Function

Suppose that X is binomial with parameters (n, p). The key to computing its distribution function

$$P\{X \le i\} = \sum_{k=0}^{i} \binom{n}{k} p^k (1-p)^{n-k}, \qquad i = 0, 1, \ldots, n$$

is to utilize the following relationship between $P\{X = k+1\}$ and $P\{X = k\}$:

$$P\{X = k+1\} = \frac{p}{1-p}\frac{n-k}{k+1}P\{X = k\} \tag{5.1.4}$$

The proof of this equation is left as an exercise.

EXAMPLE 5.1e Let X be a binomial random variable with parameters $n = 6, p = .4$. Then, starting with $P\{X = 0\} = (.6)^6$ and recursively employing Equation 5.1.4, we obtain

$$\begin{aligned}
P\{X = 0\} &= (.6)^6 = .0467 \\
P\{X = 1\} &= \tfrac{4}{6}\tfrac{6}{1}P\{X = 0\} = .1866 \\
P\{X = 2\} &= \tfrac{4}{6}\tfrac{5}{2}P\{X = 1\} = .3110 \\
P\{X = 3\} &= \tfrac{4}{6}\tfrac{4}{3}P\{X = 2\} = .2765 \\
P\{X = 4\} &= \tfrac{4}{6}\tfrac{3}{4}P\{X = 3\} = .1382 \\
P\{X = 5\} &= \tfrac{4}{6}\tfrac{2}{5}P\{X = 4\} = .0369 \\
P\{X = 6\} &= \tfrac{4}{6}\tfrac{1}{6}P\{X = 5\} = .0041 \quad ■
\end{aligned}$$

The text disk uses Equation 5.1.4 to compute binomial probabilities. In using it, one enters the binomial parameters n and p and a value i and the program computes the probabilities that a binomial (n, p) random variable is equal to and is less than or equal to i.

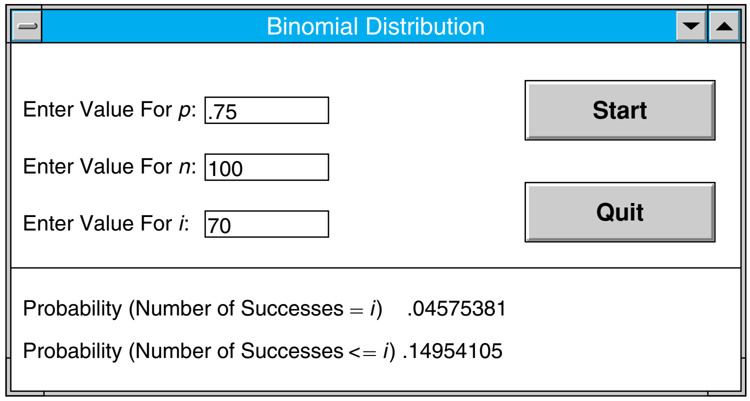


FIGURE 5.2

EXAMPLE 5.1f If X is a binomial random variable with parameters $n = 100$ and $p = .75$, find $P\{X = 70\}$ and $P\{X \leq 70\}$.

SOLUTION The text disk gives the answers shown in Figure 5.2. ■

5.2 THE POISSON RANDOM VARIABLE

A random variable X, taking on one of the values 0, 1, 2, ..., is said to be a Poisson random variable with parameter $\lambda, \lambda > 0$, if its probability mass function is given by

$$P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \qquad i = 0, 1, \ldots \tag{5.2.1}$$

The symbol e stands for a constant approximately equal to 2.7183. It is a famous constant in mathematics, named after the Swiss mathematician L. Euler, and it is also the base of the so-called natural logarithm.

Equation 5.2.1 defines a probability mass function, since

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda}\sum_{i=0}^{\infty} \lambda^i/i! = e^{-\lambda}e^{\lambda} = 1$$

A graph of this mass function when $\lambda = 4$ is given in Figure 5.3.

The Poisson probability distribution was introduced by S. D. Poisson in a book he wrote dealing with the application of probability theory to lawsuits, criminal trials, and the like. This book, published in 1837, was entitled *Recherches sur la probabilité des jugements en matière criminelle et en matière civile.*

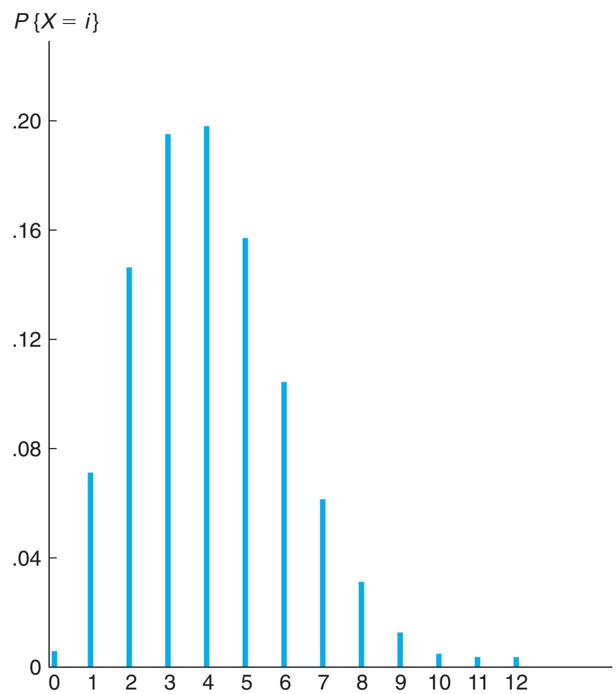


FIGURE 5.3 *The Poisson probability mass function with* $\lambda = 4$.

As a prelude to determining the mean and variance of a Poisson random variable, let us first determine its moment generating function.

$$\begin{aligned}
\phi(t) &= E[e^{tX}] \\
&= \sum_{i=0}^{\infty} e^{ti} e^{-\lambda} \lambda^i / i! \\
&= e^{-\lambda} \sum_{i=0}^{\infty} (\lambda e^t)^i / i! \\
&= e^{-\lambda} e^{\lambda e^t} \\
&= \exp\{\lambda(e^t - 1)\}
\end{aligned}$$

Differentiation yields

$$\begin{aligned}
\phi'(t) &= \lambda e^t \exp\{\lambda(e^t - 1)\} \\
\phi''(t) &= (\lambda e^t)^2 \exp\{\lambda(e^t - 1)\} + \lambda e^t \exp\{\lambda(e^t - 1)\}
\end{aligned}$$

Evaluating at $t = 0$ gives that

$$
\begin{aligned}
E[X] &= \phi'(0) = \lambda \\
\mathrm{Var}(X) &= \phi''(0) - (E[X])^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda
\end{aligned}
$$

Thus both the mean and the variance of a Poisson random variable are equal to the parameter λ.

The Poisson random variable has a wide range of applications in a variety of areas because it may be used as an approximation for a binomial random variable with parameters (n, p) when n is large and p is small. To see this, suppose that X is a binomial random variable with parameters (n, p) and let $\lambda = np$. Then

$$
\begin{aligned}
P\{X = i\} &= \frac{n!}{(n-1)!i!} p^i (1-p)^{n-i} \\
&= \frac{n!}{(n-1)!i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1)\ldots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}
\end{aligned}
$$

Now, for n large and p small,

$$
\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda} \qquad \frac{n(n-1)\ldots(n-i+1)}{n^i} \approx 1 \qquad \left(1 - \frac{\lambda}{n}\right)^i \approx 1
$$

Hence, for n large and p small,

$$
P\{X = i\} \approx e^{-\lambda} \frac{\lambda^i}{i!}
$$

In other words, if n independent trials, each of which results in a "success" with probability p, are performed, then when n is large and p small, the number of successes occurring is approximately a Poisson random variable with mean $\lambda = np$.

Some examples of random variables that usually obey, to a good approximation, the Poisson probability law (that is, they usually obey Equation 5.2.1 for some value of λ) are:

1. The number of misprints on a page (or a group of pages) of a book.
2. The number of people in a community living to 100 years of age.
3. The number of wrong telephone numbers that are dialed in a day.
4. The number of transistors that fail on their first day of use.
5. The number of customers entering a post office on a given day.

6. The number of α-particles discharged in a fixed period of time from some radioactive particle.

Each of the foregoing, and numerous other random variables, is approximately Poisson for the same reason — namely, because of the Poisson approximation to the binomial. For instance, we can suppose that there is a small probability p that each letter typed on a page will be misprinted, and so the number of misprints on a given page will be approximately Poisson with mean $\lambda = np$ where n is the (presumably) large number of letters on that page. Similarly, we can suppose that each person in a given community, independently, has a small probability p of reaching the age 100, and so the number of people that do will have approximately a Poisson distribution with mean np where n is the large number of people in the community. We leave it for the reader to reason out why the remaining random variables in examples 3 through 6 should have approximately a Poisson distribution.

EXAMPLE 5.2a Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the probability that there is at least one accident this week.

SOLUTION Let X denote the number of accidents occurring on the stretch of highway in question during this week. Because it is reasonable to suppose that there are a large number of cars passing along that stretch, each having a small probability of being involved in an accident, the number of such accidents should be approximately Poisson distributed. Hence,

$$\begin{aligned} P\{X \geq 1\} &= 1 - P\{X = 0\} \\ &= 1 - e^{-3}\frac{3^0}{0!} \\ &= 1 - e^{-3} \\ &\approx .9502 \quad \blacksquare \end{aligned}$$

EXAMPLE 5.2b Suppose the probability that an item produced by a certain machine will be defective is .1. Find the probability that a sample of 10 items will contain at most one defective item. Assume that the quality of successive items is independent.

SOLUTION The desired probability is $\binom{10}{0}(.1)^0(.9)^{10} + \binom{10}{1}(.1)^1(.9)^9 = .7361$, whereas the Poisson approximation yields the value

$$e^{-1}\frac{1^0}{0!} + e^{-1}\frac{1^1}{1!} = 2e^{-1} \approx .7358 \quad \blacksquare$$

EXAMPLE 5.2c Consider an experiment that consists of counting the number of α particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on the average, 3.2 such α-particles are given off, what is a good approximation to the probability that no more than 2 α-particles will appear?

SOLUTION If we think of the gram of radioactive material as consisting of a large number n of atoms each of which has probability $3.2/n$ of disintegrating and sending off an α-particle during the second considered, then we see that, to a very close approximation, the number of α-particles given off will be a Poisson random variable with parameter $\lambda = 3.2$. Hence the desired probability is

$$\begin{aligned} P\{X \leq 2\} &= e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2} \\ &= .382 \end{aligned}$$ ■

EXAMPLE 5.2d If the average number of claims handled daily by an insurance company is 5, what proportion of days have less than 3 claims? What is the probability that there will be 4 claims in exactly 3 of the next 5 days? Assume that the number of claims on different days is independent.

SOLUTION Because the company probably insures a large number of clients, each having a small probability of making a claim on any given day, it is reasonable to suppose that the number of claims handled daily, call it X, is a Poisson random variable. Since $E(X) = 5$, the probability that there will be fewer than 3 claims on any given day is

$$\begin{aligned} P\{X \leq 3\} &= P\{X = 0\} + P\{X = 1\} + P\{X = 2\} \\ &= e^{-5} + e^{-5}\frac{5^1}{1!} + e^{-5}\frac{5^2}{2!} \\ &= \frac{37}{2}e^{-5} \\ &\approx .1247 \end{aligned}$$

Since any given day will have fewer than 3 claims with probability .125, it follows, from the law of large numbers, that over the long run 12.5 percent of days will have fewer than 3 claims.

It follows from the assumed independence of the number of claims over successive days that the number of days in a 5-day span that have exactly 4 claims is a binomial random variable with parameters 5 and $P\{X = 4\}$. Because

$$P\{X = 4\} = e^{-5}\frac{5^4}{4!} \approx .1755$$

it follows that the probability that 3 of the next 5 days will have 4 claims is equal to

$$\binom{5}{3}(.1755)^3(.8245)^2 \approx .0367$$ ■

The Poisson approximation result can be shown to be valid under even more general conditions than those so far mentioned. For instance, suppose that n independent trials are to be performed, with the ith trial resulting in a success with probability p_i, $i = 1, \ldots, n$. Then it can be shown that if n is large and each p_i is small, then the number of successful trials is approximately Poisson distributed with mean equal to $\sum_{i=1}^{n} p_i$. In fact, this result will sometimes remain true even when the trials are not independent, provided that their dependence is "weak." For instance, consider the following example.

EXAMPLE 5.2e At a party n people put their hats in the center of a room, where the hats are mixed together. Each person then randomly chooses a hat. If X denotes the number of people who select their own hat, then, for large n, it can be shown that X has approximately a Poisson distribution with mean 1. To see why this might be true, let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th person selects his or her own hat} \\ 0 & \text{otherwise} \end{cases}$$

Then we can express X as

$$X = X_1 + \cdots + X_n$$

and so X can be regarded as representing the number of "successes" in n "trials" where trial i is said to be a success if the ith person chooses her own hat. Now, since the ith person is equally likely to end up with any of the n hats, one of which is her own, it follows that

$$P\{X_i = 1\} = \frac{1}{n} \tag{5.2.2}$$

Suppose now that $i \neq j$ and consider the conditional probability that the ith person chooses her own hat given that the jth person does — that is, consider $P\{X_i = 1|X_j = 1\}$. Now given that the jth person indeed selects her own hat, it follows that the ith individual is equally likely to end up with any of the remaining $n - 1$, one of which is her own. Hence, it follows that

$$P\{X_i = 1|X_j = 1\} = \frac{1}{n-1} \tag{5.2.3}$$

Thus, we see from Equations 5.2.2 and 5.2.3 that whereas the trials are not independent, their dependence is rather weak [since, if the above conditional probability were equal to $1/n$ rather than $1/(n-1)$, then trials i and j would be independent]; and thus it is not at all surprising that X has approximately a Poisson distribution. The fact that $E[X] = 1$ follows since

$$\begin{aligned} E[X] &= E[X_1 + \cdots + X_n] \\ &= E[X_1] + \cdots + E[X_n] \\ &= n\left(\frac{1}{n}\right) = 1 \end{aligned}$$

The last equality follows since, from Equation 5.2.2,

$$E[X_i] = P\{X_i = 1\} = \frac{1}{n} \quad ■$$

The Poisson distribution possesses the reproductive property that the sum of independent Poisson random variables is also a Poisson random variable. To see this, suppose that X_1 and X_2 are independent Poisson random variables having respective means λ_1 and λ_2. Then the moment generating function of $X_1 + X_2$ is as follows:

$$\begin{aligned} E[e^{t(X_1+X_2)}] &= E[e^{tX_1} e^{tX_2}] \\ &= E[e^{tX_1}]E[e^{tX_2}] \qquad \text{by independence} \\ &= \exp\{\lambda_1(e^t - 1)\} \ \exp\{\lambda_2(e^t - 1)\} \\ &= \exp\{(\lambda_1 + \lambda_2)(e^t - 1)\} \end{aligned}$$

Because $\exp\{(\lambda_1 + \lambda_2)(e^t - 1)\}$ is the moment generating function of a Poisson random variable having mean $\lambda_1+\lambda_2$, we may conclude, from the fact that the moment generating function uniquely specifies the distribution, that $X_1 + X_2$ is Poisson with mean $\lambda_1 + \lambda_2$.

EXAMPLE 5.2f It has been established that the number of defective stereos produced daily at a certain plant is Poisson distributed with mean 4. Over a 2-day span, what is the probability that the number of defective stereos does not exceed 3?

SOLUTION Assuming that X_1, the number of defectives produced during the first day, is independent of X_2, the number produced during the second day, then $X_1 + X_2$ is Poisson with mean 8. Hence,

$$P\{X_1 + X_2 \leq 3\} = \sum_{i=0}^{3} e^{-8} \frac{8^i}{i!} = .04238 \quad ■$$

Consider now a situation in which a random number, call it N, of events will occur, and suppose that each of these events will independently be a type 1 event with probability p or a type 2 event with probability $1 - p$. Let N_1 and N_2 denote, respectively, the numbers of type 1 and type 2 events that occur. (So $N = N_1 + N_2$.) If N is Poisson distributed with mean λ, then the joint probability mass function of N_1 and N_2 is obtained as follows.

$$\begin{aligned} P\{N_1 = n, N_2 = m\} &= P\{N_1 = n, N_2 = m, N = n + m\} \\ &= P\{N_1 = n, N_2 = m | N = n + m\}P\{N = n + m\} \\ &= P\{N_1 = n, N_2 = m | N = n + m\}e^{-\lambda} \frac{\lambda^{n+m}}{(n + m)!} \end{aligned}$$

Now, given a total of $n + m$ events, because each one of these events is independently type 1 with probability p, it follows that the conditional probability that there are exactly n type 1 events (and m type 2 events) is the probability that a binomial $(n+m, p)$ random variable is equal to n. Consequently,

$$P\{N_1 = n, N_2 = m\} = \frac{(n+m)!}{n!m!}p^n(1-p)^m e^{-\lambda}\frac{\lambda^{n+m}}{(n+m)!}$$
$$= e^{-\lambda p}\frac{(\lambda p)^n}{n!}e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!} \tag{5.2.4}$$

The probability mass function of N_1 is thus

$$P\{N_1 = n\} = \sum_{m=0}^{\infty} P\{N_1 = n, N_2 = m\}$$
$$= e^{-\lambda p}\frac{(\lambda p)^n}{n!}\sum_{m=0}^{\infty} e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!}$$
$$= e^{-\lambda p}\frac{(\lambda p)^n}{n!} \tag{5.2.5}$$

Similarly,

$$P\{N_2 = m\} = \sum_{n=0}^{\infty} P\{N_1 = n, N_2 = m\} = e^{-\lambda(1-p)}\frac{(\lambda(1-p))^m}{m!} \tag{5.2.6}$$

It now follows from Equations 5.2.4, 5.2.5, and 5.2.6, that N_1 and N_2 are independent Poisson random variables with respective means λp and $\lambda(1-p)$.

The preceding result generalizes when each of the Poisson number of events can be classified into any of r categories, to yield the following important property of the Poisson distribution: *If each of a Poisson number of events having mean λ is independently classified as being of one of the types $1, \ldots, r$, with respective probabilities $p_1, \ldots, p_r$, $\sum_{i=1}^{r} p_i = 1$, then the numbers of type $1, \ldots, r$ events are independent Poisson random variables with respective means $\lambda p_1, \ldots, \lambda p_r$.*

5.2.1 Computing the Poisson Distribution Function

If X is Poisson with mean λ, then

$$\frac{P\{X = i+1\}}{P\{X = i\}} = \frac{e^{-\lambda}\lambda^{i+1}/(i+1)!}{e^{-\lambda}\lambda^i/i!} = \frac{\lambda}{i+1} \tag{5.2.7}$$

Starting with $P\{X = 0\} = e^{-\lambda}$, we can use Equation 5.2.7 to successively compute

$$\begin{aligned} P\{X = 1\} &= \lambda P\{X = 0\} \\ P\{X = 2\} &= \frac{\lambda}{2} P\{X = 1\} \\ &\vdots \\ P\{X = i + 1\} &= \frac{\lambda}{i+1} P\{X = i\} \end{aligned}$$

The text disk includes a program that uses Equation 5.2.7 to compute Poisson probabilities.

5.3 THE HYPERGEOMETRIC RANDOM VARIABLE

A bin contains $N + M$ batteries, of which N are of acceptable quality and the other M are defective. A sample of size n is to be randomly chosen (without replacements) in the sense that the set of sampled batteries is equally likely to be any of the $\binom{N+M}{n}$ subsets of size n. If we let X denote the number of acceptable batteries in the sample, then

$$P\{X = i\} = \frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}, \qquad i = 0, 1, \ldots, \min(N, n)^* \tag{5.3.1}$$

Any random variable X whose probability mass function is given by Equation 5.3.1 is said to be a *hypergeometric* random variable with parameters N, M, n.

EXAMPLE 5.3a The components of a 6-component system are to be randomly chosen from a bin of 20 used components. The resulting system will be functional if at least 4 of its 6 components are in working condition. If 15 of the 20 components in the bin are in working condition, what is the probability that the resulting system will be functional?

SOLUTION If X is the number of working components chosen, then X is hypergeometric with parameters 15, 5, 6. The probability that the system will be functional is

$$\begin{aligned} P\{X \geq 4\} &= \sum_{i=4}^{6} P\{X = i\} \\ &= \frac{\binom{15}{4}\binom{5}{2} + \binom{15}{5}\binom{5}{1} + \binom{15}{6}\binom{5}{0}}{\binom{20}{6}} \\ &\approx .8687 \quad \blacksquare \end{aligned}$$

* We are following the convention that $\binom{m}{r} = 0$ if $r > m$ or if $r < 0$.

To compute the mean and variance of a hypergeometric random variable whose probability mass function is given by Equation 5.3.1, imagine that the batteries are drawn sequentially and let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th selection is acceptable} \\ 0 & \text{otherwise} \end{cases}$$

Now, since the ith selection is equally likely to be any of the $N + M$ batteries, of which N are acceptable, it follows that

$$P\{X_i = 1\} = \frac{N}{N + M} \tag{5.3.2}$$

Also, for $i \neq j$,

$$\begin{aligned} P\{X_i = 1, X_j = 1\} &= P\{X_i = 1\}P\{X_j = 1|X_i = 1\} \\ &= \frac{N}{N+M}\frac{N-1}{N+M-1} \end{aligned} \tag{5.3.3}$$

which follows since, given that the ith selection is acceptable, the jth selection is equally likely to be any of the other $N + M - 1$ batteries of which $N - 1$ are acceptable.

To compute the mean and variance of X, the number of acceptable batteries in the sample of size n, use the representation

$$X = \sum_{i=1}^{n} X_i$$

This gives

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P\{X_i = 1\} = \frac{nN}{N+M} \tag{5.3.4}$$

Also, Corollary 4.7.3 for the variance of a sum of random variables gives

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum\sum_{1 \le i < j \le n} \text{Cov}(X_i, X_j) \tag{5.3.5}$$

Now, X_i is a Bernoulli random variable and so

$$\text{Var}(X_i) = P\{X_i = 1\}(1 - P\{X_i = 1\}) = \frac{N}{N+M}\frac{M}{N+M} \tag{5.3.6}$$

Also, for $i < j$,

$$\text{Cov}(X_i, X_j) = E[X_iX_j] - E[X_i]E[X_j]$$

Now, because both X_i and X_j are Bernoulli (that is, $0 - 1$) random variables, it follows that X_iX_j is a Bernoulli random variable, and so

$$\begin{aligned} E[X_iX_j] &= P\{X_iX_j = 1\} \\ &= P\{X_i = 1, X_j = 1\} \\ &= \frac{N(N-1)}{(N+M)(N+M-1)} \quad \text{from Equation 5.3.3} \end{aligned} \tag{5.3.7}$$

So from Equation 5.3.2 and the foregoing we see that for $i \neq j$,

$$\begin{aligned} \text{Cov}(X_i, X_j) &= \frac{N(N-1)}{(N+M)(N+M-1)} - \left(\frac{N}{N+M}\right)^2 \\ &= \frac{-NM}{(N+M)^2(N+M-1)} \end{aligned}$$

Hence, since there are $\binom{n}{2}$ terms in the second sum on the right side of Equation 5.3.5, we obtain from Equation 5.3.6

$$\begin{aligned} \text{Var}(X) &= \frac{nNM}{(N+M)^2} - \frac{n(n-1)NM}{(N+M)^2(N+M-1)} \\ &= \frac{nNM}{(N+M)^2}\left(1 - \frac{n-1}{N+M-1}\right) \end{aligned} \tag{5.3.8}$$

If we let $p = N/(N+M)$ denote the proportion of batteries in the bin that are acceptable, we can rewrite Equations 5.3.4 and 5.3.8 as follows.

$$\begin{aligned} E(X) &= np \\ \text{Var}(X) &= np(1-p)\left[1 - \frac{n-1}{N+M-1}\right] \end{aligned}$$

It should be noted that, for fixed p, as $N + M$ increases to ∞, Var(X) converges to $np(1-p)$, which is the variance of a binomial random variable with parameters (n, p). (Why was this to be expected?)

EXAMPLE 5.3b An unknown number, say N, of animals inhabit a certain region. To obtain some information about the population size, ecologists often perform the following experiment: They first catch a number, say r, of these animals, mark them in some manner, and release them. After allowing the marked animals time to disperse throughout the region, a new catch of size, say, n is made. Let X denote the number of marked animals in this second capture. If we assume that the population of animals in the region remained fixed between the time of the two catches and that each time an animal was caught it was equally likely to be any of the remaining uncaught animals, it follows that X is a hypergeometric random variable such that

$$P\{X = i\} = \frac{\binom{r}{i}\binom{N-r}{n-i}}{\binom{N}{n}} \equiv P_i(N)$$

Suppose now that X is observed to equal i. That is, the fraction i/n of the animals in the second catch were marked. By taking this as an approximation of r/N, the proportion of animals in the region that are marked, we obtain the estimate rn/i of the number of animals in the region. For instance, if $r = 50$ animals are initially caught, marked, and then released, and a subsequent catch of $n = 100$ animals revealed $X = 25$ of them that were marked, then we would estimate the number of animals in the region to be about 200. ■

There is a relationship between binomial random variables and the hypergeometric distribution that will be useful to us in developing a statistical test concerning two binomial populations.

EXAMPLE 5.3c Let X and Y be independent binomial random variables having respective parameters (n, p) and (m, p). The conditional probability mass function of X given that $X + Y = k$ is as follows.

$$\begin{aligned}
P\{X = i|X + Y = k\} &= \frac{P\{X = i, X + Y = k\}}{P\{X + Y = k\}} \\
&= \frac{P\{X = i, Y = k - i\}}{P\{X + Y = k\}} \\
&= \frac{P\{X = i\}P\{Y = k - i\}}{P\{X + Y = k\}} \\
&= \frac{\binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m-(k-i)}}{\binom{n+m}{k} p^k (1-p)^{n+m-k}} \\
&= \frac{\binom{n}{i}\binom{m}{k-i}}{\binom{n+m}{k}}
\end{aligned}$$

where the next-to-last equality used the fact that $X + Y$ is binomial with parameters $(n + m, p)$. Hence, we see that the conditional distribution of X given the value of $X + Y$ is hypergeometric.

It is worth noting that the preceding is quite intuitive. For suppose that $n + m$ independent trials, each of which has the same probability of being a success, are performed; let X be the number of successes in the first n trials, and let Y be the number of successes in the final m trials. Given a total of k successes in the $n + m$ trials, it is quite intuitive that each subgroup of k trials is equally likely to consist of those trials that resulted in successes. That is, the k success trials are distributed as a random selection of k of the $n + m$ trials, and so the number that are from the first n trials is hypergeometric. ■

5.4 THE UNIFORM RANDOM VARIABLE

A random variable X is said to be uniformly distributed over the interval $[\alpha, \beta]$ if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

A graph of this function is given in Figure 5.4. Note that the foregoing meets the requirements of being a probability density function since

$$\frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} dx = 1$$

The uniform distribution arises in practice when we suppose a certain random variable is equally likely to be near any value in the interval $[\alpha, \beta]$.

The probability that X lies in any subinterval of $[\alpha, \beta]$ is equal to the length of that subinterval divided by the length of the interval $[\alpha, \beta]$. This follows since when $[a, b]$ is a subinterval of $[\alpha, \beta]$ (see Figure 5.5),

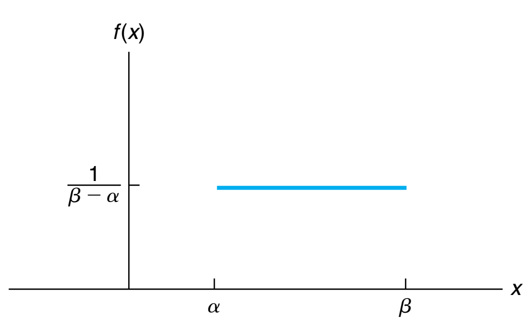


FIGURE 5.4 *Graph of $f(x)$ for a uniform $[\alpha, \beta]$.*

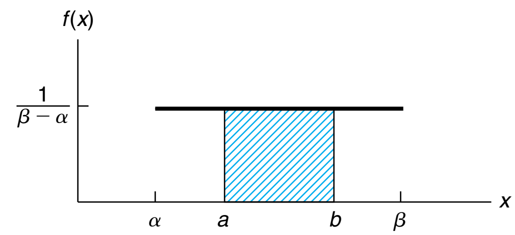


FIGURE 5.5 *Probabilities of a uniform random variable.*

$$P\{a < X < b\} = \frac{1}{\beta - \alpha} \int_a^b dx$$
$$= \frac{b - a}{\beta - \alpha}$$

EXAMPLE 5.4a If X is uniformly distributed over the interval [0, 10], compute the probability that **(a)** $2 < X < 9$, **(b)** $1 < X < 4$, **(c)** $X < 5$, **(d)** $X > 6$.

SOLUTION The respective answers are **(a)** 7/10, **(b)** 3/10, **(c)** 5/10, **(d)** 4/10. ■

EXAMPLE 5.4b Buses arrive at a specified stop at 15-minute intervals starting at 7 A.M. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on. If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits

(a) less than 5 minutes for a bus;
(b) at least 12 minutes for a bus.

SOLUTION Let X denote the time in minutes past 7 A.M. that the passenger arrives at the stop. Since X is a uniform random variable over the interval (0, 30), it follows that the passenger will have to wait less than 5 minutes if he arrives between 7:10 and 7:15 or between 7:25 and 7:30. Hence, the desired probability for **(a)** is

$$P\{10 < X < 15\} + P\{25 < X < 30\} = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}$$

Similarly, he would have to wait at least 12 minutes if he arrives between 7 and 7:03 or between 7:15 and 7:18, and so the probability for **(b)** is

$$P\{0 < X < 3\} + P\{15 < X < 18\} = \frac{3}{30} + \frac{3}{30} = \frac{1}{5} \quad ■$$

The mean of a uniform $[\alpha, \beta]$ random variable is

$$\begin{aligned} E[X] &= \int_\alpha^\beta \frac{x}{\beta - \alpha}\,dx \\ &= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} \\ &= \frac{(\beta - \alpha)(\beta + \alpha)}{2(\beta - \alpha)} \end{aligned}$$

or

$$E[X] = \frac{\alpha + \beta}{2}$$

Or, in other words, the expected value of a uniform $[\alpha, \beta]$ random variable is equal to the midpoint of the interval $[\alpha, \beta]$, which is clearly what one would expect. (Why?)

The variance is computed as follows.

$$\begin{aligned} E[X^2] &= \frac{1}{\beta - \alpha}\int_\alpha^\beta x^2 dx \\ &= \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} \\ &= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} \end{aligned}$$

where the final equation used that

$$\beta^3 - \alpha^3 = (\beta^2 + \alpha\beta + \alpha^2)(\beta - \alpha)$$

Hence,

$$\begin{aligned} \mathrm{Var}(X) &= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \left(\frac{\alpha + \beta}{2}\right)^2 \\ &= \frac{4(\beta^2 + \alpha\beta + \alpha^2) - 3(\alpha^2 + 2\alpha\beta + \beta^2)}{12} \\ &= \frac{\alpha^2 + \beta^2 - 2\alpha\beta}{12} \\ &= \frac{(\beta - \alpha)^2}{12} \end{aligned}$$

EXAMPLE 5.4c The current in a semiconductor diode is often measured by the Shockley equation

$$I = I_0(e^{aV} - 1)$$

where V is the voltage across the diode; I_0 is the reverse current; a is a constant; and I is the resulting diode current. Find $E[I]$ if $a = 5$, $I_0 = 10^{-6}$, and V is uniformly distributed over $(1, 3)$.

SOLUTION

$$\begin{aligned}
E[I] &= E[I_0(e^{aV} - 1)] \\
&= I_0 E[e^{aV} - 1] \\
&= I_0(E[e^{aV}] - 1) \\
&= 10^{-6} \int_1^3 e^{5x} \frac{1}{2} dx - 10^{-6} \\
&= 10^{-7}(e^{15} - e^5) - 10^{-6} \\
&\approx .3269 \quad \blacksquare
\end{aligned}$$

The value of a uniform (0, 1) random variable is called a *random number*. Most computer systems have a built-in subroutine for generating (to a high level of approximation) sequences of independent random numbers — for instance, Table 5.1 presents a set of independent random numbers. Random numbers are quite useful in probability and statistics because their use enables one to empirically estimate various probabilities and expectations.

TABLE 5.1 *A Random Number Table*

.68587	.25848	.85227	.78724	.05302	.70712	.76552	.70326	.80402	.49479
.73253	.41629	.37913	.00236	.60196	.59048	.59946	.75657	.61849	.90181
.84448	.42477	.94829	.86678	.14030	.04072	.45580	.36833	.10783	.33199
.49564	.98590	.92880	.69970	.83898	.21077	.71374	.85967	.20857	.51433
.68304	.46922	.14218	.63014	.50116	.33569	.97793	.84637	.27681	.04354
.76992	.70179	.75568	.21792	.50646	.07744	.38064	.06107	.41481	.93919
.37604	.27772	.75615	.51157	.73821	.29928	.62603	.06259	.21552	.72977
.43898	.06592	.44474	.07517	.44831	.01337	.04538	.15198	.50345	.65288
.86039	.28645	.44931	.59203	.98254	.56697	.55897	.25109	.47585	.59524
.28877	.84966	.97319	.66633	.71350	.28403	.28265	.61379	.13886	.78325
.44973	.12332	.16649	.88908	.31019	.33358	.68401	.10177	.92873	.13065
.42529	.37593	.90208	.50331	.37531	.72208	.42884	.07435	.58647	.84972
.82004	.74696	.10136	.35971	.72014	.08345	.49366	.68501	.14135	.15718
.67090	.08493	.47151	.06464	.14425	.28381	.40455	.87302	.07135	.04507
.62825	.83809	.37425	.17693	.69327	.04144	.00924	.68246	.48573	.24647
.10720	.89919	.90448	.80838	.70997	.98438	.51651	.71379	.10830	.69984
.69854	.89270	.54348	.22658	.94233	.08889	.52655	.83351	.73627	.39018
.71460	.25022	.06988	.64146	.69407	.39125	.10090	.08415	.07094	.14244
.69040	.33461	.79399	.22664	.68810	.56303	.65947	.88951	.40180	.87943
.13452	.36642	.98785	.62929	.88509	.64690	.38981	.99092	.91137	.02411
.94232	.91117	.98610	.71605	.89560	.92921	.51481	.20016	.56769	.60462
.99269	.98876	.47254	.93637	.83954	.60990	.10353	.13206	.33480	.29440
.75323	.86974	.91355	.12780	.01906	.96412	.61320	.47629	.33890	.22099
.75003	.98538	.63622	.94890	.96744	.73870	.72527	.17745	.01151	.47200

For an illustration of the use of random numbers, suppose that a medical center is planning to test a new drug designed to reduce its users' blood cholesterol levels. To test its effectiveness, the medical center has recruited 1,000 volunteers to be subjects in the test. To take into account the possibility that the subjects' blood cholesterol levels may be affected by factors external to the test (such as changing weather conditions), it has been decided to split the volunteers into 2 groups of size 500 — a *treatment* group that will be given the drug and a *control* group that will be given a placebo. Both the volunteers and the administrators of the drug will not be told who is in each group (such a test is called a *double-blind test*). It remains to determine which of the volunteers should be chosen to constitute the treatment group. Clearly, one would want the treatment group and the control group to be as similar as possible in all respects with the exception that members in the first group are to receive the drug while those in the other group receive a placebo; then it will be possible to conclude that any difference in response between the groups is indeed due to the drug. There is general agreement that the best way to accomplish this is to choose the 500 volunteers to be in the treatment group in a completely random fashion. That is, the choice should be made so that each of the $\binom{1000}{500}$ subsets of 500 volunteers is equally likely to constitute the control group. How can this be accomplished?

***EXAMPLE 5.4.d (Choosing a Random Subset)** From a set of n elements — numbered $1, 2, \ldots, n$ — suppose we want to generate a random subset of size k that is to be chosen in such a manner that each of the $\binom{n}{k}$ subsets is equally likely to be the subset chosen. How can we do this?

To answer this question, let us work backward and suppose that we have indeed randomly generated such a subset of size k. Now for each $j = 1, \ldots, n$, we set

$$I_j = \begin{cases} 1 & \text{if element } j \text{ is in the subset} \\ 0 & \text{otherwise} \end{cases}$$

and compute the conditional distribution of I_j given $I_1, \ldots, I_{j-1}$. To start, note that the probability that element 1 is in the subset of size k is clearly k/n (which can be seen either by noting that there is probability $1/n$ that element 1 would have been the jth element chosen, $j = 1, \ldots, k$; or by noting that the proportion of outcomes of the random selection that results in element 1 being chosen is $\binom{1}{1}\binom{n-1}{k-1}/\binom{n}{k} = k/n$). Therefore, we have that

$$P\{I_1 = 1\} = k/n \tag{5.4.1}$$

To compute the conditional probability that element 2 is in the subset given I_1, note that if $I_1 = 1$, then aside from element 1 the remaining $k - 1$ members of the subset would have been chosen "at random" from the remaining $n - 1$ elements (in the sense

* Optional.

that each of the subsets of size $k-1$ of the numbers $2, \ldots, n$ is equally likely to be the other elements of the subset). Hence, we have that

$$P\{I_2 = 1 | I_1 = 1\} = \frac{k-1}{n-1} \tag{5.4.2}$$

Similarly, if element 1 is not in the subgroup, then the k members of the subgroup would have been chosen "at random" from the other $n-1$ elements, and thus

$$P\{I_2 = 1 | I_1 = 0\} = \frac{k}{n-1} \tag{5.4.3}$$

From Equations 5.4.2 and 5.4.3, we see that

$$P\{I_2 = 1 | I_1\} = \frac{k - I_1}{n-1}$$

In general, we have that

$$P\{I_j = 1 | I_1, \ldots, I_{j-1}\} = \frac{k - \sum_{i=1}^{j-1} I_i}{n-j+1}, \quad j = 2, \ldots, n \tag{5.4.4}$$

The preceding formula follows since $\sum_{i=1}^{j-1} I_i$ represents the number of the first $j-1$ elements that are included in the subset, and so given $I_1, \ldots, I_{j-1}$ there remain $k - \sum_{i=1}^{j-1} I_i$ elements to be selected from the remaining $n - (j-1)$.

Since $P\{U < a\} = a, 0 \leq a \leq 1$, when U is a uniform (0, 1) random variable, Equations 5.4.1 and 5.4.4 lead to the following method for generating a random subset of size k from a set of n elements: Namely, generate a sequence of (at most n) random numbers $U_1, U_2, \ldots$ and set

$$I_1 = \begin{cases} 1 & \text{if } U_1 < \dfrac{k}{n} \\ 0 & \text{otherwise} \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if } U_2 < \dfrac{k - I_1}{n-1} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$I_j = \begin{cases} 1 & \text{if } U_j < \frac{k - I_1 - \cdots - I_{j-1}}{n-j+1} \\ 0 & \text{otherwise} \end{cases}$$

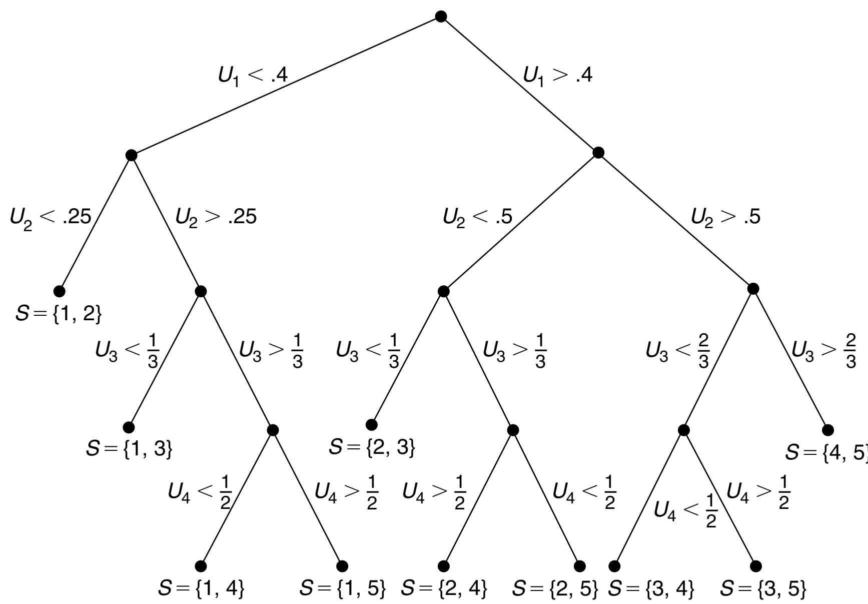


FIGURE 5.6 *Tree diagram.*

This process stops when $I_1 + \cdots + I_j = k$ and the random subset consists of the k elements whose I-value equals 1. That is, $S = \{i : I_i = 1\}$ is the subset.

For instance, if $k = 2$, $n = 5$, then the tree diagram of Figure 5.6 illustrates the foregoing technique. The random subset S is given by the final position on the tree. Note that the probability of ending up in any given final position is equal to 1/10, which can be seen by multiplying the probabilities of moving through the tree to the desired endpoint. For instance, the probability of ending at the point labeled $S = \{2, 4\}$ is $P\{U_1 > .4\}P\{U_2 < .5\}P\{U_3 > \frac{1}{3}\}P\{U_4 > \frac{1}{2}\} = (.6)(.5)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = .1$.

As indicated in the tree diagram (see the rightmost branches that result in $S = \{4, 5\}$), we can stop generating random numbers when the number of remaining places in the subset to be chosen is equal to the remaining number of elements. That is, the general procedure would stop whenever either $\sum_{i=1}^{j} I_i = k$ or $\sum_{i=1}^{j} I_i = k - (n - j)$. In the latter case, $S = \{i \leq j : I_i = 1, j + 1, \ldots, n\}$. ■

EXAMPLE 5.4e The random vector X, Y is said to have a *uniform* distribution over the two-dimensional region R if its joint density function is constant for points in R, and is 0 for points outside of R. That is, if

$$f(x, y) = \begin{cases} c & \text{if } (x, y) \in R \\ 0 & \text{if otherwise} \end{cases}$$

Because

$$\begin{aligned}1 &= \int_R f(x,y)\,dx\,dy \\ &= \int_R c\,dx\,dy \\ &= c \times \text{Area of } R\end{aligned}$$

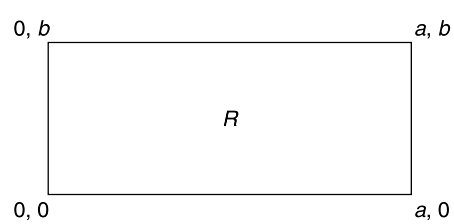
it follows that

$$c = \frac{1}{\text{Area of } R}$$

For any region $A \subset R$,

$$\begin{aligned}P\{(X,Y) \in A\} &= \int\int_{(x,y)\,\in A} f(x,y)\,dx\,dy \\ &= \int\int_{(x,y)\,\in A} c\,dx\,dy \\ &= \frac{\text{Area of } A}{\text{Area of } R}\end{aligned}$$

Suppose now that X, Y is uniformly distributed over the following rectangular region R:



Its joint density function is

$$f(x,y) = \begin{cases} c & \text{if } 0 \le x \le a,\ 0 \le y \le b \\ 0 & \text{otherwise} \end{cases}$$

where $c = \frac{1}{\text{Area of rectangle}} = \frac{1}{ab}$. In this case, X and Y are independent uniform random variables. To show this, note that for $0 \le x \le a$, $0 \le y \le b$

$$P\{X \le x, Y \le y\} = c\int_0^x \int_0^y dy\,dx = \frac{xy}{ab} \tag{5.4.5}$$

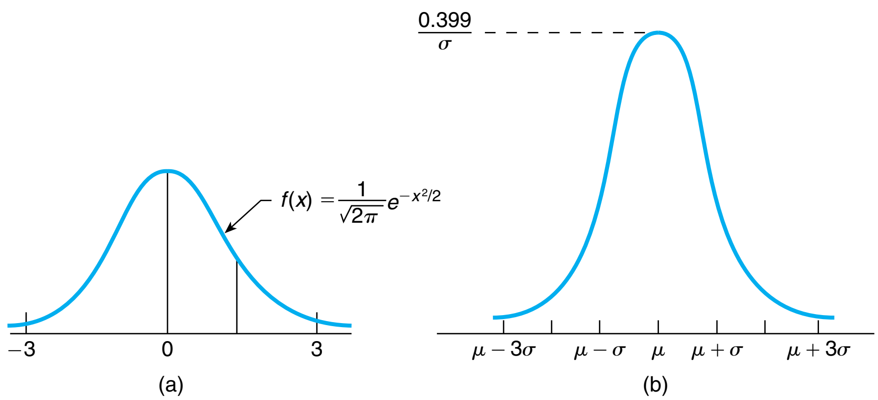


FIGURE 5.7 *The normal density function (a) with $\mu = 0, \sigma = 1$ and (b) with arbitrary μ and σ^2.*

First letting $y = b$, and then letting $x = a$, in the preceding shows that

$$P\{X \leq x\} = \frac{x}{a}, \quad P\{Y \leq y\} = \frac{y}{b} \tag{5.4.6}$$

Thus, from Equations 5.4.5 and 5.4.6 we can conclude that X and Y are independent, with X being uniform on $(0, a)$ and Y being uniform on $(0, b)$. ■

5.5 NORMAL RANDOM VARIABLES

A random variable is said to be normally distributed with parameters μ and σ^2, and we write $X \sim \mathcal{N}(\mu, \sigma^2)$, if its density is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty < x < \infty^*$$

The normal density $f(x)$ is a bell-shaped curve that is symmetric about μ and that attains its maximum value of $\frac{1}{\sqrt{2\pi}\,\sigma} \approx 0.399/\sigma$ at $x = \mu$ (see Figure 5.7).

The normal distribution was introduced by the French mathematician Abraham de Moivre in 1733 and was used by him to approximate probabilities associated with binomial random variables when the binomial parameter n is large. This result was later extended by Laplace and others and is now encompassed in a probability theorem known as the central limit theorem, which gives a theoretical base to the often noted empirical observation that, in practice, many random phenomena obey, at least approximately, a normal probability distribution. Some examples of this behavior are the height of a person, the velocity in any direction of a molecule in gas, and the error made in measuring a physical quantity.

* To verify that this is indeed a density function, see Problem 29.

To compute $E[X]$ note that

$$E[X-\mu] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)e^{-(x-\mu)^2/2\sigma^2}dx$$

Letting $y = (x-\mu)/\sigma$ gives that

$$E[X-\mu] = \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ye^{-y^2/2}dy$$

But

$$\int_{-\infty}^{\infty} ye^{-y^2/2}dy = -e^{-y^2/2}|_{-\infty}^{\infty} = 0$$

showing that $E[X-\mu] = 0$, or equivalently that

$$E[X] = \mu$$

Using this, we now compute $\mathrm{Var}(X)$ as follows:

$$\begin{aligned}
\mathrm{Var}(X) &= E[(X-\mu)^2] \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sigma^2 y^2 e^{-y^2/2}dy \qquad (5.5.1)
\end{aligned}$$

With $u = y$ and $dv = ye^{-y^2/2}$, the integration by parts formula

$$\int u\,dv = uv - \int v\,du$$

yields that

$$\begin{aligned}
\int_{-\infty}^{\infty} y^2 e^{-y^2/2}dy &= -ye^{-y^2/2}|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2}dy \\
&= \int_{-\infty}^{\infty} e^{-y^2/2}dy
\end{aligned}$$

Hence, from (5.5.1)

$$\begin{aligned}
\mathrm{Var}(X) &= \sigma^2\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}dy \\
&= \sigma^2
\end{aligned}$$

where the preceding used that $\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy$ is the density function of a normal random variable with parameters $\mu = 0$ and $\sigma = 1$, so its integral must equal 1.

Thus μ and σ^2 represent, respectively, the mean and variance of the normal distribution.

A very important property of normal random variables is that if X is normal with mean μ and variance σ^2, then for any constants a and b, $b \neq 0$, the random variable $Y = a + bX$ is also a normal random variable with parameters

$$E[Y] = E[a + bX] = a + bE[X] = a + b\mu$$

and variance

$$\text{Var}(Y) = \text{Var}(a + bX) = b^2\text{Var}(X) = b^2\sigma^2$$

To verify this, let $F_Y(y)$ be the distribution function of Y. Then, for $b > 0$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(a + bX \leq y) \\ &= P\left(X \leq \frac{y-a}{b}\right) \\ &= F_X\left(\frac{y-a}{b}\right) \end{aligned}$$

where F_X is the distribution function of X. Similarly, if $b < 0$, then

$$\begin{aligned} F_Y(y) &= P(a + bX \leq y) \\ &= P\left(X \geq \frac{y-a}{b}\right) \\ &= 1 - F_X\left(\frac{y-a}{b}\right) \end{aligned}$$

Differentiation yields that the density function of Y is

$$f_Y(y) = \begin{cases} \frac{1}{b} f_X\left(\frac{y-a}{b}\right), & \text{if} \quad b > 0 \\ -\frac{1}{b} f_X\left(\frac{y-a}{b}\right), & \text{if} \quad b < 0 \end{cases}$$

which can be written as

$$\begin{aligned} f_Y(y) &= \frac{1}{|b|} f_X\left(\frac{y-a}{b}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma|b|} e^{-\left(\frac{y-a}{b}-\mu\right)^2/2\sigma^2} \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}\sigma|b|} e^{-(y-a-b\mu)^2/2b^2\sigma^2}$$

showing that $Y = a + bX$ is normal with mean $a + b\mu$ and variance $b^2\sigma^2$.

It follows from the foregoing that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

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

is a normal random variable with mean 0 and variance 1. Such a random variable Z is said to have a *standard*, or *unit*, normal distribution. Let $\Phi(\cdot)$ denote its distribution function. That is,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy, \qquad -\infty < x < \infty$$

This result that $Z = (X - \mu)/\sigma$ has a standard normal distribution when X is normal with parameters μ and σ^2 is quite important, for it enables us to write all probability statements about X in terms of probabilities for Z. For instance, to obtain $P\{X < b\}$, we note that X will be less than b if and only if $(X - \mu)/\sigma$ is less than $(b - \mu)/\sigma$, and so

$$\begin{aligned} P\{X < b\} &= P\left\{\frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right\} \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) \end{aligned}$$

Similarly, for any $a < b$,

$$\begin{aligned} P\{a < X < b\} &= P\left\{\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right\} \\ &= P\left\{\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right\} \\ &= P\left\{Z < \frac{b - \mu}{\sigma}\right\} - P\left\{Z < \frac{a - \mu}{\sigma}\right\} \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

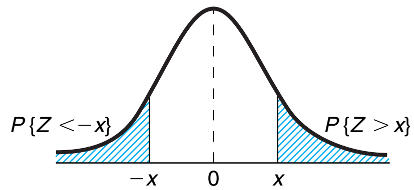


FIGURE 5.8 *Standard normal probabilities.*

It remains for us to compute $\Phi(x)$. This has been accomplished by an approximation and the results are presented in Table A1 of the Appendix, which tabulates $\Phi(x)$ (to a 4-digit level of accuracy) for a wide range of nonnegative values of x. In addition, Program 5.5a of the text disk can be used to obtain $\Phi(x)$.

While Table A1 tabulates $\Phi(x)$ only for nonnegative values of x, we can also obtain $\Phi(-x)$ from the table by making use of the symmetry (about 0) of the standard normal probability density function. That is, for $x > 0$, if Z represents a standard normal random variable, then (see Figure 5.8)

$$\begin{aligned}\Phi(-x) &= P\{Z < -x\}\\ &= P\{Z > x\} \quad \text{by symmetry}\\ &= 1 - \Phi(x)\end{aligned}$$

Thus, for instance,

$$P\{Z < -1\} = \Phi(-1) = 1 - \Phi(1) = 1 - .8413 = .1587$$

EXAMPLE 5.5a If X is a normal random variable with mean $\mu = 3$ and variance $\sigma^2 = 16$, find

(a) $P\{X < 11\}$;
(b) $P\{X > -1\}$;
(c) $P\{2 < X < 7\}$.

SOLUTION

(a)

$$\begin{aligned}P\{X < 11\} &= P\left\{\frac{X-3}{4} < \frac{11-3}{4}\right\}\\ &= \Phi(2)\\ &= .9772\end{aligned}$$

(b)

$$\begin{aligned}P\{X > -1\} &= P\left\{\frac{X-3}{4} > \frac{-1-3}{4}\right\}\\ &= P\{Z > -1\}\end{aligned}$$

$$= P\{Z < 1\}$$
$$= .8413$$

(c)
$$P\{2 < X < 7\} = P\left\{\frac{2-3}{4} < \frac{X-3}{4} < \frac{7-3}{4}\right\}$$
$$= \Phi(1) - \Phi(-1/4)$$
$$= \Phi(1) - (1 - \Phi(1/4))$$
$$= .8413 + .5987 - 1 = .4400 \quad \blacksquare$$

EXAMPLE 5.5b Suppose that a binary message — either "0" or "1" — must be transmitted by wire from location A to location B. However, the data sent over the wire are subject to a channel noise disturbance and so to reduce the possibility of error, the value 2 is sent over the wire when the message is "1" and the value -2 is sent when the message is "0." If x, $x = \pm 2$, is the value sent at location A then R, the value received at location B, is given by $R = x + N$, where N is the channel noise disturbance. When the message is received at location B, the receiver decodes it according to the following rule:

if $R \geq .5$, then "1" is concluded

if $R < .5$, then "0" is concluded

Because the channel noise is often normally distributed, we will determine the error probabilities when N is a standard normal random variable.

There are two types of errors that can occur: One is that the message "1" can be incorrectly concluded to be "0" and the other that "0" is incorrectly concluded to be "1." The first type of error will occur if the message is "1" and $2 + N < .5$, whereas the second will occur if the message is "0" and $-2 + N \geq .5$.

Hence,

$$P\{\text{error}|\text{message is "1"}\} = P\{N < -1.5\}$$
$$= 1 - \Phi(1.5) = .0668$$

and

$$P\{\text{error}|\text{message is "0"}\} = P\{N > 2.5\}$$
$$= 1 - \Phi(2.5) = .0062 \quad \blacksquare$$

EXAMPLE 5.5c The power W dissipated in a resistor is proportional to the square of the voltage V. That is,

$$W = rV^2$$

where r is a constant. If $r = 3$, and V can be assumed (to a very good approximation) to be a normal random variable with mean 6 and standard deviation 1, find

(a) $E[W]$;
(b) $P\{W > 120\}$.

SOLUTION

(a)
$$\begin{aligned} E[W] &= E[3V^2] \\ &= 3E[V^2] \\ &= 3(\text{Var}[V] + E^2[V]) \\ &= 3(1 + 36) = 111 \end{aligned}$$

(b)
$$\begin{aligned} P\{W > 120\} &= P\{3V^2 > 120\} \\ &= P\{V > \sqrt{40}\} \\ &= P\{V - 6 > \sqrt{40} - 6\} \\ &= P\{Z > .3246\} \\ &= 1 - \Phi(.3246) \\ &= .3727 \end{aligned}$$

Let us now compute the moment generating function of a normal random variable. To start, we compute the moment generating function of a standard normal random variable Z.

$$\begin{aligned} E[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2-2tx)/2} dx \\ &= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} dx \\ &= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\ &= e^{-t^2/2} \end{aligned}$$

Now, if Z is a standard normal, then $X = \mu + \sigma Z$ is normal with mean μ and variance σ^2. Using the preceding, its moment generating function is

$$\begin{aligned} E[e^{tX}] &= E[e^{t\mu + t\sigma Z}] \\ &= E[e^{t\mu} e^{t\sigma Z}] \\ &= e^{t\mu} E[e^{t\sigma Z}] \\ &= e^{t\mu} e^{-(\sigma t)^2/2} \\ &= e^{\mu t - \sigma^2 t^2/2} \end{aligned}$$

Another important result is that the sum of independent normal random variables is also a normal random variable. To see this, suppose that X_i, $i = 1, \ldots, n$, are independent, with X_i being normal with mean μ_i and variance σ_i^2. The moment generating function of $\sum_{i=1}^{n} X_i$ is as follows.

$$\begin{aligned}
E\left[e^{t\sum_{i=1}^{n} X_i}\right] &= E\left[e^{tX_1} e^{tX_2} \cdots e^{tX_n}\right] \\
&= \prod_{i=1}^{n} E\left[e^{tX_i}\right] \qquad \text{by independence} \\
&= \prod_{i=1}^{n} e^{\mu_i t + \sigma_i^2 t^2/2} \\
&= e^{\mu t + \sigma^2 t^2/2}
\end{aligned}$$

where

$$\mu = \sum_{i=1}^{n} \mu_i, \qquad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2$$

Therefore, $\sum_{i=1}^{n} X_i$ has the same moment generating function as a normal random variable having mean μ and variance σ^2. Hence, from the one-to-one correspondence between moment generating functions and distributions, we can conclude that $\sum_{i=1}^{n} X_i$ is normal with mean $\sum_{i=1}^{n} \mu_i$ and variance $\sum_{i=1}^{n} \sigma_i^2$.

EXAMPLE 5.5d Data from the National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches.

(a) Find the probability that the total precipitation during the next 2 years will exceed 25 inches.

(b) Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches.
Assume that the precipitation totals for the next 2 years are independent.

SOLUTION Let X_1 and X_2 be the precipitation totals for the next 2 years.

(a) Since $X_1 + X_2$ is normal with mean 24.16 and variance $2(3.1)^2 = 19.22$, it follows that

$$\begin{aligned}
P\{X_1 + X_2 > 25\} &= P\left\{\frac{X_1 + X_2 - 24.16}{\sqrt{19.22}} > \frac{25 - 24.16}{\sqrt{19.22}}\right\} \\
&= P\{Z > .1916\} \\
&\approx .4240
\end{aligned}$$

(b) Since $-X_2$ is a normal random variable with mean -12.08 and variance $(-1)^2(3.1)^2$, it follows that $X_1 - X_2$ is normal with mean 0 and variance 19.22. Hence,

$$\begin{aligned}
P\{X_1 > X_2 + 3\} &= P\{X_1 - X_2 > 3\} \\
&= P\left\{\frac{X_1 - X_2}{\sqrt{19.22}} > \frac{3}{\sqrt{19.22}}\right\} \\
&= P\{Z > .6843\} \\
&\approx .2469
\end{aligned}$$

Thus there is a 42.4 percent chance that the total precipitation in Los Angeles during the next 2 years will exceed 25 inches, and there is a 24.69 percent chance that next year's precipitation will exceed that of the following year by more than 3 inches. ■

For $\alpha \in (0, 1)$, let z_α be such that

$$P\{Z > z_\alpha\} = 1 - \Phi(z_\alpha) = \alpha$$

That is, the probability that a standard normal random variable is greater than z_α is equal to α (see Figure 5.9).

The value of z_α can, for any α, be obtained from Table A1. For instance, since

$$\begin{aligned}
1 - \Phi(1.645) &= .05 \\
1 - \Phi(1.96) &= .025 \\
1 - \Phi(2.33) &= .01
\end{aligned}$$

it follows that

$$z_{.05} = 1.645, \qquad z_{.025} = 1.96, \qquad z_{.01} = 2.33$$

Program 5.5b on the text disk can also be used to obtain the value of z_α.

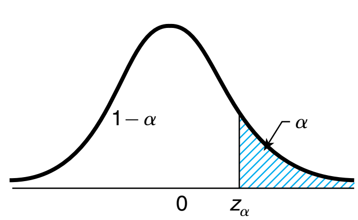


FIGURE 5.9 $P\{Z > z_\alpha\} = \alpha$.

Since

$$P\{Z < z_\alpha\} = 1 - \alpha$$

it follows that $100(1 - \alpha)$ percent of the time a standard normal random variable will be less than z_α. As a result, we call z_α the $100(1 - \alpha)$ *percentile* of the standard normal distribution.

5.6 EXPONENTIAL RANDOM VARIABLES

A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

is said to be an *exponential* random variable (or, more simply, is said to be exponentially distributed) with parameter λ. The cumulative distribution function $F(x)$ of an exponential random variable is given by

$$\begin{aligned} F(x) &= P\{X \leq x\} \\ &= \int_0^x \lambda e^{-\lambda y}\, dy \\ &= 1 - e^{-\lambda x}, \qquad x \geq 0 \end{aligned}$$

The exponential distribution often arises, in practice, as being the distribution of the amount of time until some specific event occurs. For instance, the amount of time (starting from now) until an earthquake occurs, or until a new war breaks out, or until a telephone call you receive turns out to be a wrong number are all random variables that tend in practice to have exponential distributions (see Section 5.6.1 for an explanation).

The moment generating function of the exponential is given by

$$\begin{aligned} \phi(t) &= E[e^{tX}] \\ &= \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx \\ &= \lambda \int_0^\infty e^{-(\lambda - t)x}\, dx \\ &= \frac{\lambda}{\lambda - t}, \qquad t < \lambda \end{aligned}$$

Differentiation yields

$$\phi'(t) = \frac{\lambda}{(\lambda - t)^2}$$
$$\phi''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

and so

$$\begin{aligned}
E[X] &= \phi'(0) = 1/\lambda \\
\text{Var}(X) &= \phi''(0) - (E[X])^2 \\
&= 2/\lambda^2 - 1/\lambda^2 \\
&= 1/\lambda^2
\end{aligned}$$

Thus λ is the reciprocal of the mean, and the variance is equal to the square of the mean.

The key property of an exponential random variable is that it is memoryless, where we say that a nonnegative random variable X is *memoryless* if

$$P\{X > s + t | X > t\} = P\{X > s\} \qquad \text{for all } s, t \geq 0 \tag{5.6.1}$$

To understand why Equation 5.6.1 is called the *memoryless property*, imagine that X represents the length of time that a certain item functions before failing. Now let us consider the probability that an item that is still functioning at age t will continue to function for at least an additional time s. Since this will be the case if the total functional lifetime of the item exceeds $t + s$ given that the item is still functioning at t, we see that

$$\begin{aligned}
&P\{\text{additional functional life of } t\text{-unit-old item exceeds } s\} \\
&\quad = P\{X > t + s | X > t\}
\end{aligned}$$

Thus, we see that Equation 5.6.1 states that the distribution of additional functional life of an item of age t is the same as that of a new item — in other words, when Equation 5.6.1 is satisfied, there is no need to remember the age of a functional item since as long as it is still functional it is "as good as new."

The condition in Equation 5.6.1 is equivalent to

$$\frac{P\{X > s + t, X > t\}}{P\{X > t\}} = P\{X > s\}$$

or

$$P\{X > s + t\} = P\{X > s\}P\{X > t\} \tag{5.6.2}$$

When X is an exponential random variable, then

$$P\{X > x\} = e^{-\lambda x}, \qquad x > 0$$

and so Equation 5.6.2 is satisfied (since $e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t}$). Hence, *exponentially distributed random variables are memoryless* (and in fact it can be shown that they are the only random variables that are memoryless).

EXAMPLE 5.6a Suppose that a number of miles that a car can run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5,000-mile trip, what is the probability that she will be able to complete her trip without having to replace her car battery? What can be said when the distribution is not exponential?

SOLUTION It follows, by the memoryless property of the exponential distribution, that the remaining lifetime (in thousands of miles) of the battery is exponential with parameter $\lambda = 1/10$. Hence the desired probability is

$$\begin{aligned} P\{\text{remaining lifetime} > 5\} &= 1 - F(5) \\ &= e^{-5\lambda} \\ &= e^{-1/2} \approx .604 \end{aligned}$$

However, if the lifetime distribution F is not exponential, then the relevant probability is

$$P\{\text{lifetime} > t + 5|\text{lifetime} > t\} = \frac{1 - F(t+5)}{1 - F(t)}$$

where t is the number of miles that the battery had been in use prior to the start of the trip. Therefore, if the distribution is not exponential, additional information is needed (namely, t) before the desired probability can be calculated. ■

For another illustration of the memoryless property, consider the following example.

EXAMPLE 5.6b A crew of workers has 3 interchangeable machines, of which 2 must be working for the crew to do its job. When in use, each machine will function for an exponentially distributed time having parameter λ before breaking down. The workers decide initially to use machines A and B and keep machine C in reserve to replace whichever of A or B breaks down first. They will then be able to continue working until one of the remaining machines breaks down. When the crew is forced to stop working because only one of the machines has not yet broken down, what is the probability that the still operable machine is machine C?

SOLUTION This can be easily answered, without any need for computations, by invoking the memoryless property of the exponential distribution. The argument is as follows: Consider the moment at which machine C is first put in use. At that time either A or B would have just broken down and the other one — call it machine 0 — will still be functioning. Now even though 0 would have already been functioning for some time, by the memoryless property of the exponential distribution, it follows that its remaining lifetime has the same distribution as that of a machine that is just being put into use. Thus, the remaining lifetimes of machine 0 and machine C have the same distribution and so, by symmetry, the probability that 0 will fail before C is $\frac{1}{2}$. ■

The following proposition presents another property of the exponential distribution.

PROPOSITION 5.6.1 If $X_1, X_2, \ldots, X_n$ are independent exponential random variables having respective parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\min(X_1, X_2, \ldots, X_n)$ is exponential with parameter $\sum_{t=1}^{n} \lambda_i$.

Proof

Since the smallest value of a set of numbers is greater than x if and only if all values are greater than x, we have

$$\begin{aligned} P\{\min(X_1, X_2, \ldots, X_n) > x\} &= P\{X_1 > x, X_2 > x, \ldots, X_n > x\} \\ &= \prod_{i=1}^{n} P\{X_i > x\} \quad \text{by independence} \\ &= \prod_{i=1}^{n} e^{-\lambda_i x} \\ &= e^{-\sum_{i=1}^{n} \lambda_i x} \quad ■ \end{aligned}$$

EXAMPLE 5.6c A series system is one that needs all of its components to function in order for the system itself to be functional. For an n-component series system in which the component lifetimes are independent exponential random variables with respective parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, what is the probability that the system survives for a time t?

SOLUTION Since the system life is equal to the minimal component life, it follows from Proposition 5.6.1 that

$$P\{\text{system life exceeds } t\} = e^{-\sum_i \lambda_i t} \quad ■$$

Another useful property of exponential random variables is that cX is exponential with parameter λ/c when X is exponential with parameter λ, and $c > 0$. This follows since

$$\begin{aligned} P\{cX \leq x\} &= P\{X \leq x/c\} \\ &= 1 - e^{-\lambda x/c} \end{aligned}$$

The parameter λ is called the *rate* of the exponential distribution.

*5.6.1 The Poisson Process

Suppose that "events" are occurring at random time points, and let $N(t)$ denote the number of events that occurs in the time interval $[0, t]$. These events are said to constitute a *Poisson process having rate* λ, $\lambda > 0$, if

(a) $N(0) = 0$
(b) The number of events that occur in disjoint time intervals are independent.
(c) The distribution of the number of events that occur in a given interval depends only on the length of the interval and not on its location.
(d) $\lim_{h \to 0} \frac{P\{N(h) = 1\}}{h} = \lambda$
(e) $\lim_{h \to 0} \frac{P\{N(h) \geq 2\}}{h} = 0$

Thus, Condition (a) states that the process begins at time 0. Condition (b), the *independent increment* assumption, states for instance that the number of events by time t [that is, $N(t)$] is independent of the number of events that occurs between t and $t + s$ [that is, $N(t + s) - N(t)$]. Condition (c), the *stationary increment* assumption, states that probability distribution of $N(t + s) - N(t)$ is the same for all values of t. Conditions (d) and (e) state that in a small interval of length h, the probability of one event occurring is approximately λh, whereas the probability of 2 or more is approximately 0.

We will now show that these assumptions imply that the number of events occurring in any interval of length t is a Poisson random variable with parameter λt. To be precise, let us call the interval $[0, t]$ and denote by $N(t)$ the number of events occurring in that interval. To obtain an expression for $P\{N(t) = k\}$, we start by breaking the interval $[0, t]$ into n nonoverlapping subintervals each of length t/n (Figure 5.10). Now there will be k events in $[0, t]$ if either

(i) $N(t)$ equals k and there is at most one event in each subinterval;
(ii) $N(t)$ equals k and at least one of the subintervals contains 2 or more events.

Since these two possibilities are clearly mutually exclusive, and since Condition (i) is equivalent to the statement that k of the n subintervals contain exactly 1 event and the other $n - k$ contain 0 events, we have that

$$\begin{aligned} P\{N(t) = k\} = P\{&k \text{ of the } n \text{ subintervals contain exactly 1 event} \\ &\text{and the other } n - k \text{ contain 0 events}\} + P\{N(t) = k \\ &\text{and at least 1 subinterval contains 2 or more events}\} \end{aligned} \tag{5.6.3}$$

FIGURE 5.10

* Optional section.

Now it can be shown, using Condition (e), that

$$P\{N(t) = k \text{ and at least 1 subinterval contains 2 or more events}\} \longrightarrow 0 \text{ as } n \to \infty \tag{5.6.4}$$

Also, it follows from Conditions (d) and (e) that

$$P\{\text{exactly 1 event in a subinterval}\} \approx \frac{\lambda t}{n}$$
$$P\{\text{0 events in a subinterval}\} \approx 1 - \frac{\lambda t}{n}$$

Hence, since the number of events that occur in different subintervals are independent [from Condition (b)], it follows that

$$P\{k \text{ of the subintervals contain exactly 1 event and the other } n-k \text{ contain 0 events}\}$$
$$\approx \binom{n}{k}\left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \tag{5.6.5}$$

with the approximation becoming exact as the number of subintervals, n, goes to ∞. However, the probability in Equation 5.6.5 is just the probability that a binomial random variable with parameters n and $p = \lambda t/n$ equals k. Hence, as n becomes larger and larger, this approaches the probability that a Poisson random variable with mean $n\lambda t/n = \lambda t$ equals k. Hence, from Equations 5.6.3, 5.6.4, and 5.6.5, we see upon letting n approach ∞ that

$$P\{N(t) = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

We have shown:

PROPOSITION 5.6.2 For a Poisson process having rate λ

$$P\{N(t) = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad k = 0, 1, \ldots$$

That is, the number of events in any interval of length t has a Poisson distribution with mean λt.

For a Poisson process, let X_1 denote the time of the first event. Further, for $n > 1$, let X_n denote the elapsed time between $(n-1)$st and the nth events. The sequence $\{X_n, n = 1, 2, \ldots\}$ is called the *sequence of interarrival times.* For instance, if $X_1 = 5$ and $X_2 = 10$, then the first event of the Poisson process would have occurred at time 5 and the second at time 15.

We now determine the distribution of the X_n. To do so, we first note that the event $\{X_1 > t\}$ takes place if and only if no events of the Poisson process occur in the interval $[0, t]$, and thus,

$$P\{X_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t}$$

Hence, X_1 has an exponential distribution with mean $1/\lambda$. To obtain the distribution of X_2, note that

$$\begin{aligned} P\{X_2 > t|X_1 = s\} &= P\{0 \text{ events in } (s, s+t]|X_1 = s\} \\ &= P\{0 \text{ events in } (s, s+t]\} \\ &= e^{-\lambda t} \end{aligned}$$

where the last two equations followed from independent and stationary increments. Therefore, from the foregoing we see that X_2 is also an exponential random variable with mean $1/\lambda$, and furthermore, that X_2 is independent of X_1. Repeating the same argument yields:

PROPOSITION 5.6.3 $X_1, X_2, \ldots$ are independent exponential random variables, each with mean $1/\lambda$.

*5.6.2 The Pareto Distribution

If X is an exponential random variable with rate λ, then

$$Y = \alpha\, e^X$$

is said to be a *Pareto* random variable with parameters α and λ. The parameter $\lambda > 0$ is called the index parameter, and α is called the minimum parameter (because $P(Y \geq \alpha) = 1$). The distribution function of Y is derived as follows: For $y \geq \alpha$,

$$\begin{aligned} P\{Y > y\} &= P\{\alpha\, e^X > y\} \\ &= P\{e^X > y/\alpha\} \\ &= P\{X > \log(y/\alpha)\} \\ &= e^{-\lambda \log(y/\alpha)} \\ &= e^{-\log((y/\alpha)^\lambda)} \\ &= (\alpha/y)^\lambda \end{aligned}$$

Hence, the distribution function of Y is

$$F_Y(y) = 1 - P(Y > y) = 1 - \alpha^\lambda y^{-\lambda}, \quad y \geq \alpha$$

Differentiating the distribution function yields the density function of Y:

$$f_Y(y) = \lambda\alpha^\lambda y^{-(\lambda+1)}, \quad y \geq \alpha$$

* Optional section.

It can be shown (see Problem 5-49) that $E[Y] = \infty$ when $\lambda \le 1$. When $\lambda > 1$, the mean is obtained as follows.

$$\begin{aligned}
E[Y] &= \int_{\alpha}^{\infty} y\,\lambda\alpha^{\lambda}\,y^{-(\lambda+1)}\,dy \\
&= \lambda\alpha^{\lambda}\int_{\alpha}^{\infty} y^{-\lambda}\,dy \\
&= \alpha^{\lambda}\frac{\lambda}{1-\lambda}y^{1-\lambda}\Big|_{\alpha}^{\infty} \\
&= \alpha^{\lambda}\frac{\lambda}{\lambda-1}\alpha^{1-\lambda} \\
&= \alpha\frac{\lambda}{\lambda-1}
\end{aligned}$$

An important feature of Pareto distributions is that for $y_0 > \alpha$ the conditional distribution of a Pareto random variable Y with parameters α and λ, given that it exceeds y_0, is the Pareto distribution with parameters y_0 and λ. To verify this, note for $y > y_0$ that

$$P\{Y > y | Y > y_0\} = \frac{P\{Y > y,\ Y > y_0\}}{P\{Y > y_0\}} = \frac{P\{Y > y\}}{P\{Y > y_0\}} = \frac{\alpha^{\lambda}\,y^{-\lambda}}{\alpha^{\lambda}\,y_0^{-\lambda}} = y_0^{\lambda}\,y^{-\lambda}$$

Thus, the conditional distribution is indeed Pareto with parameters y_0 and λ.

One of the earliest uses of the Pareto was as the distribution of the annual income of the members of a population. In fact, it has been widely supposed that incomes in many populations can be modeled as coming from a Pareto distribution with index parameter $\lambda = \log(5)/\log(4) \approx 1.161$. Under this supposition, it turns out that the total income of the top 20 percent of earners is 80 percent of the population's total income earnings, and that the top 20 percent of these high earners earn 80 percent of the total of all high earners income, and that the top 20 percent of these very high earners earn 80 percent of the total of all very high earners income, and so on.

To verify the preceding claim, let $y_{.8}$ be the 80 percentile of the Pareto distribution. Because $F_Y(y) = 1 - (\alpha/y)^{\lambda}$, we see that $.8 = F(y_{.8}) = 1 - (\alpha/y_{.8})^{\lambda}$, showing that

$$(\alpha/y_{.8})^{\lambda} = .2 \quad \text{or} \quad (y_{.8}/\alpha)^{\lambda} = 5$$

and thus

$$y_{.8} = \alpha\, 5^{1/\lambda}$$

Now suppose, from this point on, that $\lambda = \log(5)/\log(4)$, and note that $\log(4) = (1/\lambda)\log(5) = \log(5^{1/\lambda})$, showing that $4 = 5^{1/\lambda}$, or equivalently that $1/\lambda = \log_5(4)$. Hence,

$$y_{.8} = \alpha\, 5^{\log_5(4)} = 4\alpha$$

The average income of a randomly chosen individual in the top 20 percent is $E[Y|Y > y_{.8}]$, which is easily obtained by using that the conditional distribution of Y given that it exceeds $y_{.8}$ is Pareto with parameters $y_{.8}$ and λ. Using the previously derived formula for $E[Y]$, this yields that

$$E[Y|Y > y_{.8}] = y_{.8}\frac{\lambda}{\lambda - 1} = 4\alpha\,\frac{\lambda}{\lambda - 1}$$

To obtain $E[Y|Y < y_{.8}]$, the average income of a randomly chosen individual in the bottom 80 percent, we use the identity

$$E[Y] = E[Y|Y < y_{.8}](.8) + E[Y|Y > y_{.8}](.2)$$

Using the previously derived expressions for $E[Y]$ and $E[Y|Y > y_{.8}]$, the preceding equation yields that

$$\alpha\frac{\lambda}{\lambda - 1} = \frac{4}{5}E[Y|Y < y_{.8}] + \frac{4}{5}\alpha\,\frac{\lambda}{\lambda - 1}$$

showing that

$$E[Y|Y < y_{.8}] = \frac{\alpha}{4}\frac{\lambda}{\lambda - 1}$$

Thus,

$$E[Y|Y < y_{.8}] = \frac{1}{16}\,E[Y|Y > y_{.8}]$$

Hence, the average earnings of someone in the top 20 percent of income earned is 16 times that of someone in the lower 80 percent, thus showing that, although there are 4 times as many people in the lower earning group, the total income of the lower income group is only 20 percent of the total earnings of the population. (On average, for every 5 people in the population, 4 are in the lower 80 percent and 1 is in the upper 20 percent; the 4 in the lower earnings group earn on average a total of $4\,\frac{a}{4}\,\frac{\lambda}{\lambda-1} = a\,\frac{\lambda}{\lambda-1}$, whereas the one in the higher income group earns on average $4a\,\frac{\lambda}{\lambda-1}$. Thus, 4 out of every 5 dollars of the population's total income is earned by someone in the highest 20 percent.)

Because the conditional distribution of a high income earner (that is, one who earns more than $y_{.8}$) is Pareto with parameters $y_{.8}$ and λ, it also follows from the preceding that 80 percent of the total of the earnings of this group are earned by the top 20 percent of these high earners, and so on.

The Pareto distribution has been applied in a variety of areas. For instance, it has been used as the distribution of

(a) the file size of internet traffic (under the TCP protocol);

(b) the time to compete a job assigned to a supercomputer;

(c) the size of a meteorite;

(d) the yearly maximum one day rainfalls in different regions.

*5.7 THE GAMMA DISTRIBUTION

A random variable is said to have a gamma distribution with parameters (α, λ), $\lambda > 0$, $\alpha > 0$, if its density function is given by

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty \lambda e^{-\lambda x}(\lambda x)^{\alpha-1}\,dx \\ &= \int_0^\infty e^{-y}y^{\alpha-1}\,dy \quad \text{(by letting } y = \lambda x)\end{aligned}$$

The integration by parts formula $\int u\,dv = uv - \int v\,du$ yields, with $u = y^{\alpha-1}$, $dv = e^{-y}dy$, $v = -e^{-y}$, that for $\alpha > 1$,

$$\begin{aligned}\int_0^\infty e^{-y}y^{\alpha-1}\,dy &= -e^{-y}y^{\alpha-1}\Big|_{y=0}^{y=\infty} + \int_0^\infty e^{-y}(\alpha-1)y^{\alpha-2}\,dy \\ &= (\alpha-1)\int_0^\infty e^{-y}y^{\alpha-2}\,dy\end{aligned}$$

or

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \tag{5.7.1}$$

When α is an integer — say, $\alpha = n$ — we can iterate the foregoing to obtain that

$$\begin{aligned}\Gamma(n) &= (n-1)\Gamma(n-1) \\ &= (n-1)(n-2)\Gamma(n-2) && \text{by letting } \alpha = n-1 \text{ in Eq. 5.7.1} \\ &= (n-1)(n-2)(n-3)\Gamma(n-3) && \text{by letting } \alpha = n-2 \text{ in Eq. 5.7.1} \\ &\ \ \vdots \\ &= (n-1)!\Gamma(1)\end{aligned}$$

Because

$$\Gamma(1) = \int_0^\infty e^{-y}\,dy = 1$$

we see that

$$\Gamma(n) = (n-1)!$$

The function $\Gamma(\alpha)$ is called the *gamma* function.

---

* Optional section.

It should be noted that when $\alpha = 1$, the gamma distribution reduces to the exponential with mean $1/\lambda$.

The moment generating function of a gamma random variable X with parameters (α, λ) is obtained as follows:

$$\begin{aligned}
\phi(t) &= E[e^{tX}] \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{tx} e^{-\lambda x} x^{\alpha-1}\, dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-(\lambda-t)x} x^{\alpha-1}\, dx \\
&= \left(\frac{\lambda}{\lambda-t}\right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-y} y^{\alpha-1}\, dy \quad [\text{by } y = (\lambda-t)x] \\
&= \left(\frac{\lambda}{\lambda-t}\right)^\alpha
\end{aligned} \tag{5.7.2}$$

where the final equality used that $e^{-y}y^{\alpha-1}/\Gamma(\alpha)$ is a density function, and thus integrates to 1.

Differentiation of Equation 5.7.2 yields

$$\phi'(t) = \frac{\alpha\lambda^\alpha}{(\lambda-t)^{\alpha+1}}$$

$$\phi''(t) = \frac{\alpha(\alpha+1)\lambda^\alpha}{(\lambda-t)^{\alpha+2}}$$

Hence,

$$E[X] = \phi'(0) = \frac{\alpha}{\lambda} \tag{5.7.3}$$

$$\begin{aligned}
\text{Var}(X) &= E[X^2] - (E[X])^2 \\
&= \phi''(0) - \left(\frac{\alpha}{\lambda}\right)^2 \\
&= \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}
\end{aligned} \tag{5.7.4}$$

An important property of the gamma is that if X_1 and X_2 are independent gamma random variables having respective parameters (α_1, λ) and (α_2, λ), then $X_1 + X_2$ is a gamma random variable with parameters $(\alpha_1 + \alpha_2, \lambda)$. This result easily follows since

$$\begin{aligned}
\phi_{X_1+X_2}(t) &= E[e^{t(X_1+X_2)}] \qquad (5.7.5) \\
&= \phi_{X_1}(t)\phi_{X_2}(t) \\
&= \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1} \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_2} \qquad \text{from Equation 5.7.2} \\
&= \left(\frac{\lambda}{\lambda-t}\right)^{\alpha_1+\alpha_2}
\end{aligned}$$

which is seen to be the moment generating function of a gamma $(\alpha_1 + \alpha_2, \lambda)$ random variable. Since a moment generating function uniquely characterizes a distribution, the result entails.

The foregoing result easily generalizes to yield the following proposition.

PROPOSITION 5.7.1 If $X_i, i = 1, \ldots, n$ are independent gamma random variables with respective parameters (α_i, λ), then $\sum_{i=1}^{n} X_i$ is gamma with parameters $\sum_{i=1}^{n} \alpha_i, \lambda$.

Since the gamma distribution with parameters $(1, \lambda)$ reduces to the exponential with the rate λ, we have thus shown the following useful result.

Corollary 5.7.2

If $X_1, \ldots, X_n$ are independent exponential random variables, each having rate λ, then $\sum_{i=1}^{n} X_i$ is a gamma random variable with parameters (n, λ).

EXAMPLE 5.7a The lifetime of a battery is exponentially distributed with rate λ. If a stereo cassette requires one battery to operate, then the total playing time one can obtain from a total of n batteries is a gamma random variable with parameters (n, λ). ■

Figure 5.11 presents a graph of the gamma $(\alpha, 1)$ density for a variety of values of α. It should be noted that as α becomes large, the density starts to resemble the normal density. This is theoretically explained by the central limit theorem, which will be presented in the next chapter.

5.8 DISTRIBUTIONS ARISING FROM THE NORMAL

5.8.1 The Chi-Square Distribution

Definition

If $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables, then X, defined by

$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \tag{5.8.1}$$

is said to have a *chi-square distribution with n degrees of freedom*. We will use the notation

$$X \sim \chi_n^2$$

to signify that X has a chi-square distribution with n degrees of freedom.

The chi-square distribution has the additive property that if X_1 and X_2 are independent chi-square random variables with n_1 and n_2 degrees of freedom, respectively, then $X_1 + X_2$ is chi-square with $n_1 + n_2$ degrees of freedom. This can be formally shown either by the use of moment generating functions or, most easily, by noting that $X_1 + X_2$ is the sum of squares of $n_1 + n_2$ independent standard normals and thus has a chi-square distribution with $n_1 + n_2$ degrees of freedom.

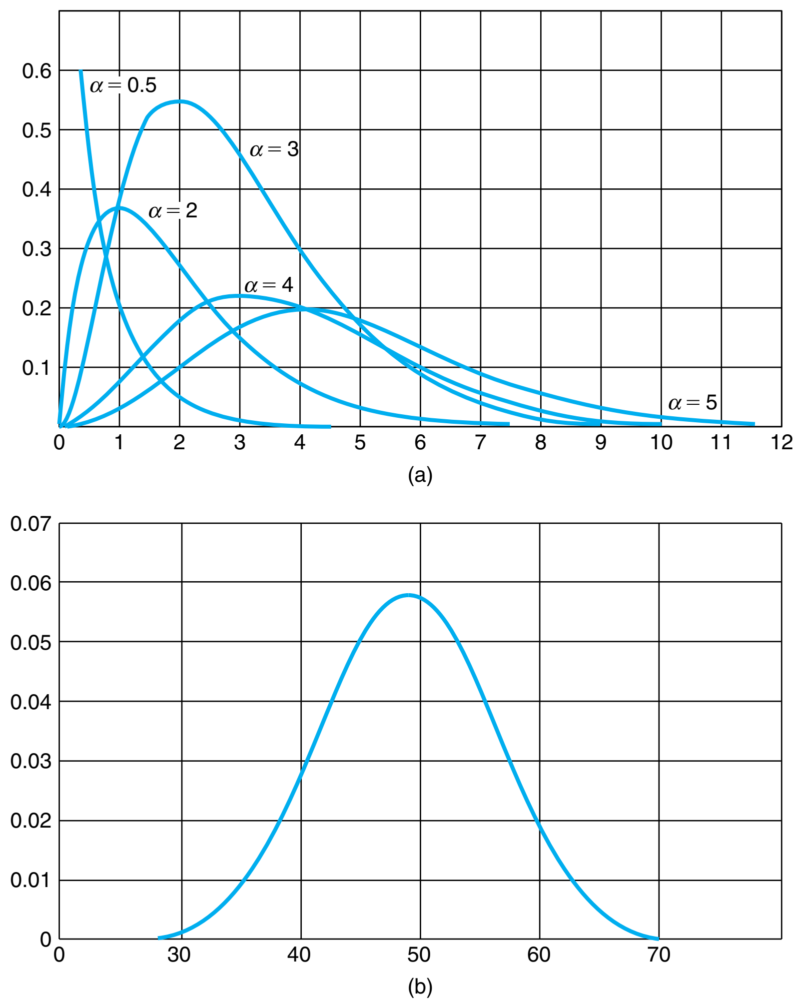


FIGURE 5.11 *Graphs of the gamma $(\alpha, 1)$ density for (a) $\alpha = .5, 2, 3, 4, 5$ and (b) $\alpha = 50$.*

If X is a chi-square random variable with n degrees of freedom, then for any $\alpha \in (0, 1)$, the quantity $\chi^2_{\alpha,n}$ is defined to be such that

$$P\{X \geq \chi^2_{\alpha,n}\} = \alpha$$

This is illustrated in Figure 5.12.

In Table A2 of the Appendix, we list $\chi^2_{\alpha,n}$ for a variety of values of α and n (including all those needed to solve problems and examples in this text). In addition, Programs 5.8.1a and 5.8.1b on the text disk can be used to obtain chi-square probabilities and the values of $\chi^2_{\alpha,n}$.

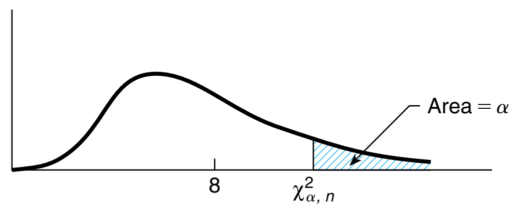


FIGURE 5.12 *The chi-square density function with 8 degrees of freedom.*

EXAMPLE 5.8a Determine $P\{\chi^2_{26} \leq 30\}$ when χ^2_{26} is a chi-square random variable with 26 degrees of freedom.

SOLUTION Using Program 5.8.1a gives the result

$$P\{\chi^2_{26} \leq 30\} = .7325 \quad \blacksquare$$

EXAMPLE 5.8b Find $\chi^2_{.05,15}$.

SOLUTION Use Program 5.8.1b to obtain:

$$\chi^2_{.05,15} = 24.996 \quad \blacksquare$$

EXAMPLE 5.8c Suppose that we are attempting to locate a target in three-dimensional space, and that the three coordinate errors (in meters) of the point chosen are independent normal random variables with mean 0 and standard deviation 2. Find the probability that the distance between the point chosen and the target exceeds 3 meters.

SOLUTION If D is the distance, then

$$D^2 = X_1^2 + X_2^2 + X_3^2$$

where X_i is the error in the ith coordinate. Since $Z_i = X_i/2, i = 1, 2, 3$, are all standard normal random variables, it follows that

$$\begin{aligned} P\{D^2 > 9\} &= P\{Z_1^2 + Z_2^2 + Z_3^2 > 9/4\} \\ &= P\{\chi^2_3 > 9/4\} \\ &= .5222 \end{aligned}$$

where the final equality was obtained from Program 5.8.1a. $\blacksquare$

* Optional section.

*5.8.1.1 The Relation Between Chi-Square and Gamma Random Variables

Let us compute the moment generating function of a chi-square random variable with n degrees of freedom. To begin, we have, when $n = 1$, that

$$
\begin{aligned}
E[e^{tX}] &= E[e^{tZ^2}] \text{ where } Z \sim \mathcal{N}(0,1) \qquad (5.8.2)\\
&= \int_{-\infty}^{\infty} e^{tx^2} f_Z(x)\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-x^2/2}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2(1-2t)/2}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2\bar{\sigma}^2}\, dx \quad \text{where } \bar{\sigma}^2 = (1-2t)^{-1} \\
&= (1-2t)^{-1/2} \frac{1}{\sqrt{2\pi}\,\bar{\sigma}} \int_{-\infty}^{\infty} e^{-x^2/2\bar{\sigma}^2}\, dx \\
&= (1-2t)^{-1/2}
\end{aligned}
$$

where the last equality follows since the integral of the normal $(0, \bar{\sigma}^2)$ density equals 1. Hence, in the general case of n degrees of freedom

$$
\begin{aligned}
E[e^{tX}] &= E\left[e^{t\sum_{i=1}^{n} Z_i^2}\right] \\
&= E\left[\prod_{i=1}^{n} e^{tZ_i^2}\right] \\
&= \prod_{i=1}^{n} E[e^{tZ_i^2}] \quad \text{by independence of the } Z_i \\
&= \left(\frac{1/2}{1/2 - t}\right)^{n/2} \\
&= (1-2t)^{-n/2} \quad \text{from Equation 5.8.2}
\end{aligned}
$$

which we recognize as being the moment generating function of a gamma random variable with parameters $(n/2, 1/2)$. Hence, by the uniqueness of moment generating functions, it follows that these two distributions — chi-square with n degrees of freedom and gamma

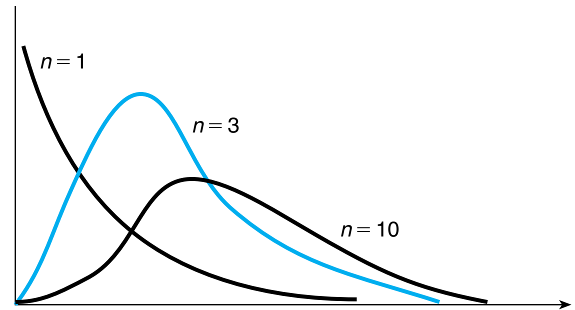


FIGURE 5.13 *The chi-square density function with n degrees of freedom.*

with parameters $n/2$ and $1/2$ — are identical, and thus we can conclude that the density of X is given by

$$f(x) = \frac{\frac{1}{2}e^{-x/2}\left(\frac{x}{2}\right)^{(n/2)-1}}{\Gamma\left(\frac{n}{2}\right)}, \qquad x > 0$$

The chi-square density functions having 1, 3, and 10 degrees of freedom, respectively, are plotted in Figure 5.13.

Let us reconsider Example 5.8c, this time supposing that the target is located in the two-dimensional plane.

EXAMPLE 5.8d When we attempt to locate a target in two-dimensional space, suppose that the coordinate errors are independent normal random variables with mean 0 and standard deviation 2. Find the probability that the distance between the point chosen and the target exceeds 3.

SOLUTION If D is the distance and X_i, $i = 1, 2$, are the coordinate errors, then

$$D^2 = X_1^2 + X_2^2$$

Since $Z_i = X_i/2$, $i = 1, 2$, are standard normal random variables, we obtain

$$P\{D^2 > 9\} = P\{Z_1^2 + Z_2^2 > 9/4\} = P\{\chi_2^2 > 9/4\} = e^{-9/8} \approx .3247$$

where the preceding calculation used the fact that the chi-square distribution with 2 degrees of freedom is the same as the exponential distribution with parameter 1/2. ■

Since the chi-square distribution with n degrees of freedom is identical to the gamma distribution with parameters $\alpha = n/2$ and $\lambda = 1/2$, it follows from Equations 5.7.3 and 5.7.4 that the mean and variance of a random variable X having this distribution is

$$E[X] = n, \qquad \text{Var}(X) = 2n$$

5.8.2 The t-Distribution

If Z and χ_n^2 are independent random variables, with Z having a standard normal distribution and χ_n^2 having a chi-square distribution with n degrees of freedom, then the random variable T_n defined by

$$T_n = \frac{Z}{\sqrt{\chi_n^2/n}}$$

is said to have a *t-distribution with n degrees of freedom*. A graph of the density function of T_n is given in Figure 5.14 for $n = 1, 5$, and 10.

Like the standard normal density, the t-density is symmetric about zero. In addition, as n becomes larger, it becomes more and more like a standard normal density. To understand why, recall that χ_n^2 can be expressed as the sum of the squares of n standard normals, and so

$$\frac{\chi_n^2}{n} = \frac{Z_1^2 + \cdots + Z_n^2}{n}$$

where $Z_1, \ldots, Z_n$ are independent standard normal random variables. It now follows from the weak law of large numbers that, for large n, χ_n^2/n will, with probability close to 1, be approximately equal to $E[Z_i^2] = 1$. Hence, for n large, $T_n = Z/\sqrt{\chi_n^2/n}$ will have approximately the same distribution as Z.

Figure 5.15 shows a graph of the t-density function with 5 degrees of freedom compared with the standard normal density. Notice that the t-density has thicker "tails," indicating greater variability, than does the normal density.

The mean and variance of T_n can be shown to equal

$$E[T_n] = 0, \qquad n > 1$$

$$\text{Var}(T_n) = \frac{n}{n-2}, \qquad n > 2$$

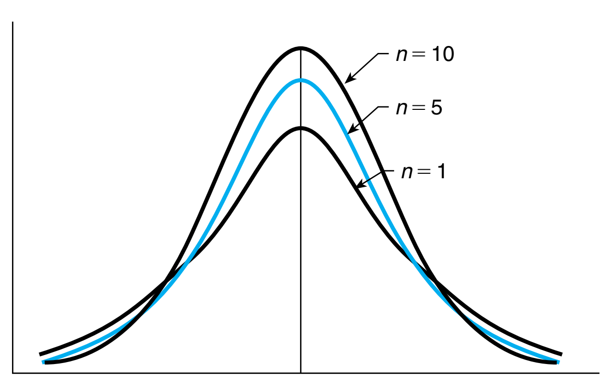


FIGURE 5.14 *Density function of T_n.*

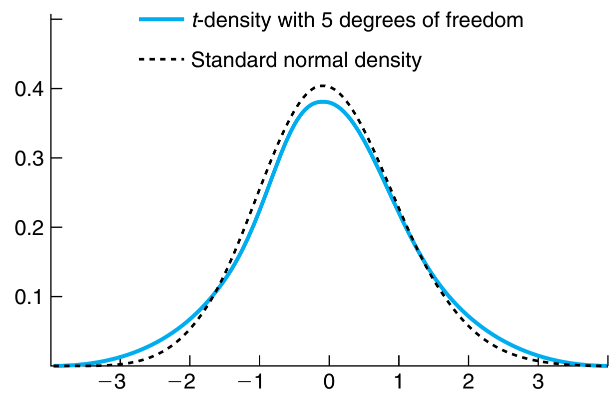


FIGURE 5.15 *Comparing standard normal density with the density of T_5.*

Thus the variance of T_n decreases to 1 — the variance of a standard normal random variable — as n increases to ∞. For α, $0 < \alpha < 1$, let $t_{\alpha,n}$ be such that

$$P\{T_n \geq t_{\alpha,n}\} = \alpha$$

It follows from the symmetry about zero of the t-density function that $-T_n$ has the same distribution as T_n, and so

$$\begin{aligned}\alpha &= P\{-T_n \geq t_{\alpha,n}\} \\ &= P\{T_n \leq -t_{\alpha,n}\} \\ &= 1 - P\{T_n > -t_{\alpha,n}\}\end{aligned}$$

Therefore,

$$P\{T_n \geq -t_{\alpha,n}\} = 1 - \alpha$$

leading to the conclusion that

$$-t_{\alpha,n} = t_{1-\alpha,n}$$

which is illustrated in Figure 5.16.

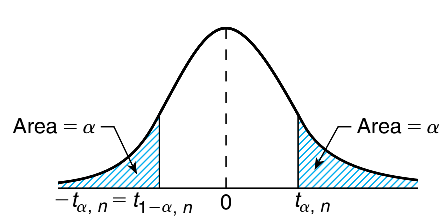


FIGURE 5.16 $t_{1-\alpha,n} = -t_{\alpha,n}$.

The values of $t_{\alpha,n}$ for a variety of values of n and α have been tabulated in Table A3 in the Appendix. In addition, Programs 5.8.2a and 5.8.2b on the text disk compute the t-distribution function and the values $t_{\alpha,n}$, respectively.

EXAMPLE 5.8e Find **(a)** $P\{T_{12} \leq 1.4\}$ and **(b)** $t_{.025,9}$.

SOLUTION Run Programs 5.8.2a and 5.8.2b to obtain the results.

(a) .9066 **(b)** 2.2625 ■

5.8.3 The F-Distribution

If χ_n^2 and χ_m^2 are independent chi-square random variables with n and m degrees of freedom, respectively, then the random variable $F_{n,m}$ defined by

$$F_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m}$$

is said to have an *F-distribution with n and m degrees of freedom.*

For any $\alpha \in (0, 1)$, let $F_{\alpha,n,m}$ be such that

$$P\{F_{n,m} > F_{\alpha,n,m}\} = \alpha$$

This is illustrated in Figure 5.17.

The quantities $F_{\alpha,n,m}$ are tabulated in Table A4 of the Appendix for different values of n, m, and $\alpha \leq \frac{1}{2}$. If $F_{\alpha,n,m}$ is desired when $\alpha > \frac{1}{2}$, it can be obtained by using the following equalities:

$$\begin{aligned}
\alpha &= P\left\{\frac{\chi_n^2/n}{\chi_m^2/m} > F_{\alpha,n,m}\right\} \\
&= P\left\{\frac{\chi_m^2/m}{\chi_n^2/n} < \frac{1}{F_{\alpha,n,m}}\right\} \\
&= 1 - P\left\{\frac{\chi_m^2/m}{\chi_n^2/n} \geq \frac{1}{F_{\alpha,n,m}}\right\}
\end{aligned}$$

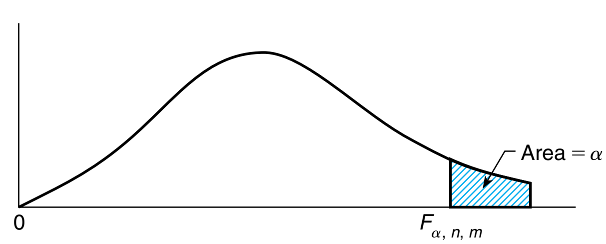


FIGURE 5.17 *Density function of $F_{n,m}$.*

or, equivalently,

$$P\left\{\frac{\chi_m^2/m}{\chi_n^2/n} \geq \frac{1}{F_{\alpha,n,m}}\right\} = 1 - \alpha \tag{5.8.3}$$

But because $(\chi_m^2/m)/(\chi_n^2/n)$ has an F-distribution with degrees of freedom m and n, it follows that

$$1 - \alpha = P\left\{\frac{\chi_m^2/m}{\chi_n^2/n} \geq F_{1-\alpha,m,n}\right\}$$

implying, from Equation 5.8.3, that

$$\frac{1}{F_{\alpha,n,m}} = F_{1-\alpha,m,n}$$

Thus, for instance, $F_{.9,5,7} = 1/F_{.1,7,5} = 1/3.37 = .2967$ where the value of $F_{.1,7,5}$ was obtained from Table A4 of the Appendix.

Program 5.8.3 computes the distribution function of $F_{n,m}$.

EXAMPLE 5.8f Determine $P\{F_{6,14} \leq 1.5\}$.

SOLUTION Run Program 5.8.3 to obtain the solution .7518. ■

*5.9 THE LOGISTICS DISTRIBUTION

A random variable X is said to have a *logistics* distribution with parameters μ and $v > 0$ if its distribution function is

$$F(x) = \frac{e^{(x-\mu)/v}}{1 + e^{(x-\mu)/v}}, \qquad -\infty < x < \infty$$

Differentiating $F(x) = 1 - 1/(1 + e^{(x-\mu)/v})$ yields the density function

$$f(x) = \frac{e^{(x-\mu)/v}}{v(1 + e^{(x-\mu)/v})^2}, \qquad -\infty < x < \infty$$

To obtain the mean of a logistics random variable,

$$E[X] = \int_{-\infty}^{\infty} x \frac{e^{(x-\mu)/v}}{v(1 + e^{(x-\mu)/v})^2}\, dx$$

* Optional section.

make the substitution $y = (x - \mu)/v$. This yields

$$\begin{aligned} E[X] &= v \int_{-\infty}^{\infty} \frac{ye^y}{(1+e^y)^2}\, dy + \mu \int_{-\infty}^{\infty} \frac{e^y}{(1+e^y)^2}\, dy \\ &= v \int_{-\infty}^{\infty} \frac{ye^y}{(1+e^y)^2}\, dy + \mu \end{aligned} \tag{5.9.1}$$

where the preceding equality used that $e^y/((1+e^y)^2)$ is the density function of a logistic random variable with parameters $\mu = 0$, $v = 1$ (such a random variable is called a *standard logistic*) and thus integrates to 1. Now,

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{ye^y}{(1+e^y)^2}\, dy &= \int_{-\infty}^{0} \frac{ye^y}{(1+e^y)^2}\, dy + \int_{0}^{\infty} \frac{ye^y}{(1+e^y)^2}\, dy \\ &= -\int_{0}^{\infty} \frac{xe^{-x}}{(1+e^{-x})^2}\, dx + \int_{0}^{\infty} \frac{ye^y}{(1+e^y)^2}\, dy \\ &= -\int_{0}^{\infty} \frac{xe^{x}}{(e^{x}+1)^2}\, dx + \int_{0}^{\infty} \frac{ye^y}{(1+e^y)^2}\, dy \\ &= 0 \end{aligned} \tag{5.9.2}$$

where the second equality is obtained by making the substitution $x = -y$, and the third by multiplying the numerator and denominator by e^{2x}. From Equations 5.9.1 and 5.9.2 we obtain

$$E[X] = \mu$$

Thus μ is the mean of the logistic; v is called the dispersion parameter.

Problems

1. A satellite system consists of 4 components and can function adequately if at least 2 of the 4 components are in working condition. If each component is, independently, in working condition with probability .6, what is the probability that the system functions adequately?
2. A communications channel transmits the digits 0 and 1. However, due to static, the digit transmitted is incorrectly received with probability .2. Suppose that we want to transmit an important message consisting of one binary digit. To reduce the chance of error, we transmit 00000 instead of 0 and 11111 instead of 1. If the receiver of the message uses "majority" decoding, what is the probability

that the message will be incorrectly decoded? What independence assumptions are you making? (By majority decoding we mean that the message is decoded as "0" if there are at least three zeros in the message received and as "1" otherwise.)

3. If each voter is for Proposition A with probability .7, what is the probability that exactly 7 of 10 voters are for this proposition?

4. Suppose that a particular trait (such as eye color or left-handedness) of a person is classified on the basis of one pair of genes, and suppose that d represents a dominant gene and r a recessive gene. Thus, a person with dd genes is pure dominance, one with rr is pure recessive, and one with rd is hybrid. The pure dominance and the hybrid are alike in appearance. Children receive 1 gene from each parent. If, with respect to a particular trait, 2 hybrid parents have a total of 4 children, what is the probability that 3 of the 4 children have the outward appearance of the dominant gene?

5. At least one-half of an airplane's engines are required to function in order for it to operate. If each engine independently functions with probability p, for what values of p is a 4-engine plane more likely to operate than a 2-engine plane?

6. Let X be a binomial random variable with

$$E[X] = 7 \quad \text{and} \quad \text{Var}(X) = 2.1$$

Find

(a) $P\{X = 4\}$;
(b) $P\{X > 12\}$.

7. If X and Y are binomial random variables with respective parameters (n, p) and $(n, 1 - p)$, verify and explain the following identities:

(a) $P\{X \leq i\} = P\{Y \geq n - i\}$;
(a) $P\{X = k\} = P\{Y = n - k\}$.

8. If X is a binomial random variable with parameters n and p, where $0 < p < 1$, show that

(a) $P\{X = k + 1\} = \dfrac{p}{1-p}\dfrac{n-k}{k+1}P\{X = k\}, k = 0, 1, \ldots, n - 1.$
(b) As k goes from 0 to n, $P\{X = k\}$ first increases and then decreases, reaching its largest value when k is the largest integer less than or equal to $(n + 1)p$.

9. Derive the moment generating function of a binomial random variable and then use your result to verify the formulas for the mean and variance given in the text.

10. Compare the Poisson approximation with the correct binomial probability for the following cases:

(a) $P\{X = 2\}$ when $n = 10, p = .1$;
(b) $P\{X = 0\}$ when $n = 10, p = .1$;
(c) $P\{X = 4\}$ when $n = 9, p = .2$.

11. If you buy a lottery ticket in 50 lotteries, in each of which your chance of winning a prize is $\frac{1}{100}$, what is the (approximate) probability that you will win a prize **(a)** at least once, **(b)** exactly once, and **(c)** at least twice?

12. The number of times that an individual contracts a cold in a given year is a Poisson random variable with parameter $\lambda = 3$. Suppose a new wonder drug (based on large quantities of vitamin C) has just been marketed that reduces the Poisson parameter to $\lambda = 2$ for 75 percent of the population. For the other 25 percent of the population, the drug has no appreciable effect on colds. If an individual tries the drug for a year and has 0 colds in that time, how likely is it that the drug is beneficial for him or her?

13. In the 1980s, an average of 121.95 workers died on the job each week. Give estimates of the following quantities:

(a) the proportion of weeks having 130 deaths or more;
(b) the proportion of weeks having 100 deaths or less.

Explain your reasoning.

14. Approximately 80,000 marriages took place in the state of New York last year. Estimate the probability that for at least one of these couples

(a) both partners were born on April 30;
(b) both partners celebrated their birthday on the same day of the year.

State your assumptions.

15. The game of frustration solitaire is played by turning the cards of a randomly shuffled deck of 52 playing cards over one at a time. Before you turn over the first card, say ace; before you turn over the second card, say two, before you turn over the third card, say three. Continue in this manner (saying ace again before turning over the fourteenth card, and so on). You lose if you ever turn over a card that matches what you have just said. Use the Poisson paradigm to approximate the probability of winning. (The actual probability is .01623.)

16. The probability of error in the transmission of a binary digit over a communication channel is $1/10^3$. Write an expression for the exact probability of more than 3 errors when transmitting a block of 10^3 bits. What is its approximate value? Assume independence.

17. If X is a Poisson random variable with mean λ, show that $P\{X = i\}$ first increases and then decreases as i increases, reaching its maximum value when i is the largest integer less than or equal to λ.

18. A contractor purchases a shipment of 100 transistors. It is his policy to test 10 of these transistors and to keep the shipment only if at least 9 of the 10 are in working condition. If the shipment contains 20 defective transistors, what is the probability it will be kept?

19. Let X denote a hypergeometric random variable with parameters n, m, and k. That is,

$$P\{X = i\} = \frac{\binom{n}{i}\binom{m}{k-i}}{\binom{n+m}{k}}, \qquad i = 0, 1, \ldots, \min(k, n)$$

(a) Derive a formula for $P\{X = i\}$ in terms of $P\{X = i - 1\}$.
(b) Use part (a) to compute $P\{X = i\}$ for $i = 0, 1, 2, 3, 4, 5$ when $n = m = 10$, $k = 5$, by starting with $P\{X = 0\}$.
(c) Based on the recursion in part (a), write a program to compute the hypergeometric distribution function.
(d) Use your program from part (c) to compute $P\{X \le 10\}$ when $n = m = 30$, $k = 15$.

20. Independent trials, each of which is a success with probability p, are successively performed. Let X denote the first trial resulting in a success. That is, X will equal k if the first $k-1$ trials are all failures and the kth a success. X is called a *geometric* random variable. Compute

(a) $P\{X = k\}, k = 1, 2, \ldots;$
(b) $E[X]$.

Let Y denote the number of trials needed to obtain r successes. Y is called a *negative binomial random variable*. Compute

(c) $P\{Y = k\}, k = r, r + 1, \ldots.$

(*Hint*: In order for Y to equal k, how many successes must result in the first $k-1$ trials and what must be the outcome of trial k?)

(d) Show that

$$E[Y] = r/p$$

(*Hint*: Write $Y = Y_1 + \cdots + Y_r$ where Y_i is the number of trials needed to go from a total of $i - 1$ to a total of i successes.)

21. If U is uniformly distributed on (0, 1), show that $a + (b - a)U$ is uniform on (a, b).

22. You arrive at a bus stop at 10 o'clock, knowing that the bus will arrive at some time uniformly distributed between 10 and 10:30. What is the probability that you will have to wait longer than 10 minutes? If at 10:15 the bus has not yet arrived, what is the probability that you will have to wait at least an additional 10 minutes?

23. If X is a normal random variable with parameters $\mu = 10, \sigma^2 = 36$, compute

(a) $P\{X > 5\}$;
(b) $P\{4 < X < 16\}$;
(c) $P\{X < 8\}$;
(d) $P\{X < 20\}$;
(e) $P\{X > 16\}$.

24. The Scholastic Aptitude Test mathematics test scores across the population of high school seniors follow a normal distribution with mean 500 and standard deviation 100. If five seniors are randomly chosen, find the probability that **(a)** all scored below 600 and **(b)** exactly three of them scored above 640.

25. The annual rainfall (in inches) in a certain region is normally distributed with $\mu = 40, \sigma = 4$. What is the probability that in 2 of the next 4 years the rainfall will exceed 50 inches? Assume that the rainfalls in different years are independent.

26. The weekly demand for a product approximately has a normal distribution with mean 1,000 and standard deviation 200. The current on hand inventory is 2,200 and no deliveries will be occurring in the next two weeks. Assuming that the demands in different weeks are independent,

(a) what is the probability that the demand in each of the next 2 weeks is less than 1,100?
(b) what is the probability that the total of the demands in the next 2 weeks exceeds 2,200?

27. A certain type of lightbulb has an output that is normally distributed with mean 2,000 end foot candles and standard deviation 85 end foot candles. Determine a lower specification limit L so that only 5 percent of the lightbulbs produced will be below this limit. (That is, determine L so that $P\{X \geq L\} = .95$, where X is the output of a bulb.)

28. A manufacturer produces bolts that are specified to be between 1.19 and 1.21 inches in diameter. If its production process results in a bolt's diameter being normally distributed with mean 1.20 inches and standard deviation .005, what percentage of bolts will not meet specifications?

29. Let $I = \int_{-\infty}^{\infty} e^{-x^2/2}\,dx$.

(a) Show that for any μ and σ

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx = 1$$

is equivalent to $I = \sqrt{2\pi}$.

(b) Show that $I = \sqrt{2\pi}$ by writing

$$I^2 = \int_{-\infty}^{\infty} e^{-x^2/2}\,dx \int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy$$

and then evaluating the double integral by means of a change of variables to polar coordinates. (That is, let $x = r\cos\theta$, $y = r\sin\theta$, $dx\,dy = r\,dr\,d\theta$.)

30. A random variable X is said to have a lognormal distribution if $\log X$ is normally distributed. If X is lognormal with $E[\log X] = \mu$ and $\text{Var}(\log X) = \sigma^2$, determine the distribution function of X. That is, what is $P\{X \le x\}$?

31. The salaries of pediatric physicians are approximately normally distributed. If 25 percent of these physicians earn below 180, 000 and 25 percent earn above 320, 000, what fraction earn

(a) below 250, 000;

(b) between 260, 00 and 300, 000?

32. The sample mean and sample standard deviation on your economics examination were 60 and 20, respectively; the sample mean and sample standard deviation on your statistics examination were 55 and 10, respectively. You scored 70 on the economics exam and 62 on the statistics exam. Assuming that the two histograms of test scores are approximately normal histograms,

(a) on which exam was your percentile score highest?

(b) approximate the percentage of the scores on the economics exam that were below your score.

(c) approximate the percentage of the scores on the statistics exam that were below your score.

33. Value at risk (VAR) has become a key concept in financial calculations. The VAR of an investment is defined as that value v such that there is only a 1 percent chance that the loss from the investment will exceed v.

(a) If the gain from an investment is a normal random variable with mean 10 and variance 49, determine the value at risk. (If X is the gain, then $-X$ is the loss.)

(b) Among a set of investments whose gains are all normally distributed show that the one having the smallest VAR is the one having the largest value of

$\mu - 2.33\sigma$, where μ and σ^2 are the mean and variance of the gain from the investment.

34. The annual rainfall in Cincinnati is normally distributed with mean 40.14 inches and standard deviation 8.7 inches.

(a) What is the probability this year's rainfall will exceed 42 inches?
(b) What is the probability that the sum of the next 2 years' rainfall will exceed 84 inches?
(c) What is the probability that the sum of the next 3 years' rainfall will exceed 126 inches?
(d) For parts (b) and (c), what independence assumptions are you making?

35. The height of adult women in the United States is normally distributed with mean 64.5 inches and standard deviation 2.4 inches. Find the probability that a randomly chosen woman is

(a) less than 63 inches tall;
(b) less than 70 inches tall;
(c) between 63 and 70 inches tall.
(d) Alice is 72 inches tall. What percentage of women is shorter than Alice?
(e) Find the probability that the average of the heights of two randomly chosen women exceeds 66 inches.
(f) Repeat part (e) for four randomly chosen women.

36. An IQ test produces scores that are normally distributed with mean value 100 and standard deviation 14.2. The top 1 percent of all scores are in what range?

37. The time (in hours) required to repair a machine is an exponentially distributed random variable with parameter $\lambda = 1$.

(a) What is the probability that a repair time exceeds 2 hours?
(b) What is the conditional probability that a repair takes at least 3 hours, given that its duration exceeds 2 hours?

38. The number of years a radio functions is exponentially distributed with parameter $\lambda = \frac{1}{8}$. If Jones buys a used radio, what is the probability that it will be working after an additional 10 years?

39. Jones figures that the total number of thousands of miles that a used auto can be driven before it would need to be junked is an exponential random variable with parameter $\frac{1}{20}$. Smith has a used car that he claims has been driven only 10,000 miles. If Jones purchases the car, what is the probability that she would get at least 20,000 additional miles out of it? Repeat under the assumption that the lifetime mileage of the car is not exponentially distributed but rather is (in thousands of miles) uniformly distributed over (0, 40).

*40. Let $X_1, X_2, \ldots, X_n$ denote the first n interarrival times of a Poisson process and set $S_n = \sum_{i=1}^{n} X_i$.

(a) What is the interpretation of S_n?
(b) Argue that the two events $\{S_n \le t\}$ and $\{N(t) \ge n\}$ are identical.
(c) Use part (b) to show that

$$P\{S_n \le t\} = 1 - \sum_{j=0}^{n-1} e^{-\lambda t} (\lambda t)^j / j!$$

(d) By differentiating the distribution function of S_n given in part (c), conclude that S_n is a gamma random variable with parameters n and λ. (This result also follows from Corollary 5.7.2.)

*41. Earthquakes occur in a given region in accordance with a Poisson process with rate 5 per year.

(a) What is the probability there will be at least two earthquakes in the first half of 2015?
(b) Assuming that the event in part (a) occurs, what is the probability that there will be no earthquakes during the first 9 months of 2016?
(c) Assuming that the event in part (a) occurs, what is the probability that there will be at least four earthquakes over the first 9 months of the year 2015?

*42. When shooting at a target in a two-dimensional plane, suppose that the horizontal miss distance is normally distributed with mean 0 and variance 4 and is independent of the vertical miss distance, which is also normally distributed with mean 0 and variance 4. Let D denote the distance between the point at which the shot lands and the target. Find $E[D]$.

43. If X is a chi-square random variable with 6 degrees of freedom, find

(a) $P\{X \le 6\}$;
(b) $P\{3 \le X \le 9\}$.

44. If X and Y are independent chi-square random variables with 3 and 6 degrees of freedom, respectively, determine the probability that $X + Y$ will exceed 10.

45. Show that $\Gamma(1/2) = \sqrt{\pi}$ (*Hint*: Evaluate $\int_0^\infty e^{-x} x^{-1/2}\,dx$ by letting $x = y^2/2$, $dx = y\,dy$.)

46. If T has a t-distribution with 8 degrees of freedom, find **(a)** $P\{T \ge 1\}$, **(b)** $P\{T \le 2\}$, and **(c)** $P\{-1 < T < 1\}$.

47. If T_n has a t-distribution with n degrees of freedom, show that T_n^2 has an F-distribution with 1 and n degrees of freedom.

---

* From optional sections.

48. Let Φ be the standard normal distribution function. If, for constants a and $b > 0$

$$P\{X \le x\} = \Phi\left(\frac{x-a}{b}\right)$$

characterize the distribution of X.

***49.** Suppose that Y has a Pareto distribution with minimal parameter α and index parameter λ.

(a) Find $E[Y]$ when $\lambda > 1$, and show that $E[Y] = \infty$ when $\lambda \le 1$.
(b) Find $\mathrm{Var}(Y)$ when $\lambda > 2$.

***50.** Suppose that $Y = \alpha e^X$, where X is exponential with rate λ. Use the lack of memory property of the exponential to argue that the conditional distribution of Y given that $Y > y_0 > \alpha$ is Pareto with parameters y_0 and λ.

* From optional sections.