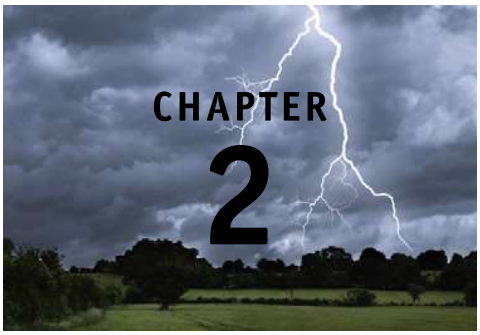
Random Variables



2.1 Random Variables

It frequently occurs that in performing an experiment we are mainly interested in some functions of the outcome as opposed to the outcome itself. For instance, in tossing dice we are often interested in the sum of the two dice and are not really concerned about the actual outcome. That is, we may be interested in knowing that the sum is seven and not be concerned over whether the actual outcome was (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1). These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as *random variables*.

Since the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Example 2.1 Letting X denote the random variable that is defined as the sum of two fair dice; then

$$
\begin{aligned}
P\{X = 2\} &= P\{(1, 1)\} = \tfrac{1}{36},\\
P\{X = 3\} &= P\{(1, 2), (2, 1)\} = \tfrac{2}{36},\\
P\{X = 4\} &= P\{(1, 3), (2, 2), (3, 1)\} = \tfrac{3}{36},\\
P\{X = 5\} &= P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = \tfrac{4}{36},\\
P\{X = 6\} &= P\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} = \tfrac{5}{36},\\
P\{X = 7\} &= P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = \tfrac{6}{36},\\
P\{X = 8\} &= P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = \tfrac{5}{36},\\
P\{X = 9\} &= P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = \tfrac{4}{36},
\end{aligned}
$$

Introduction to Probability Models, Eleventh Edition. http://dx.doi.org/10.1016/B978-0-12-407948-9.00002-5
© 2014 Elsevier Inc. All rights reserved.

$$
\begin{aligned}
P\{X = 10\} &= P\{(4, 6), (5, 5), (6, 4)\} = \tfrac{3}{36}, \\
P\{X = 11\} &= P\{(5, 6), (6, 5)\} = \tfrac{2}{36}, \\
P\{X = 12\} &= P\{(6, 6)\} = \tfrac{1}{36}
\end{aligned} \tag{2.1}
$$

In other words, the random variable X can take on any integral value between two and twelve, and the probability that it takes on each value is given by Equation (2.1). Since X must take on one of the values two through twelve, we must have

$$
1 = P\left\{\bigcup_{i=2}^{12} \{X = n\}\right\} = \sum_{n=2}^{12} P\{X = n\}
$$

which may be checked from Equation (2.1). ■

Example 2.2 For a second example, suppose that our experiment consists of tossing two fair coins. Letting Y denote the number of heads appearing, then Y is a random variable taking on one of the values 0, 1, 2 with respective probabilities

$$
\begin{aligned}
P\{Y = 0\} &= P\{(T, T)\} = \tfrac{1}{4}, \\
P\{Y = 1\} &= P\{(T, H), (H, T)\} = \tfrac{2}{4}, \\
P\{Y = 2\} &= P\{(H, H)\} = \tfrac{1}{4}
\end{aligned}
$$

Of course, $P\{Y = 0\} + P\{Y = 1\} + P\{Y = 2\} = 1$. ■

Example 2.3 Suppose that we toss a coin having a probability p of coming up heads, until the first head appears. Letting N denote the number of flips required, then assuming that the outcome of successive flips are independent, N is a random variable taking on one of the values $1, 2, 3, \ldots,$ with respective probabilities

$$
\begin{aligned}
P\{N = 1\} &= P\{H\} = p, \\
P\{N = 2\} &= P\{(T, H)\} = (1 - p)p, \\
P\{N = 3\} &= P\{(T, T, H)\} = (1 - p)^2 p, \\
&\vdots \\
P\{N = n\} &= P\{(\underbrace{T, T, \ldots, T}_{n-1}, H)\} = (1 - p)^{n-1} p, \qquad n \geq 1
\end{aligned}
$$

As a check, note that

$$
\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} \{N = n\}\right) &= \sum_{n=1}^{\infty} P\{N = n\} \\
&= p \sum_{n=1}^{\infty} (1 - p)^{n-1} \\
&= \frac{p}{1 - (1 - p)} \\
&= 1
\end{aligned}
$$

■

Example 2.4 Suppose that our experiment consists of seeing how long a battery can operate before wearing down. Suppose also that we are not primarily interested in the actual lifetime of the battery but are concerned only about whether or not the battery lasts at least two years. In this case, we may define the random variable I by

$$I = \begin{cases} 1, & \text{if the lifetime of battery is two or more years} \\ 0, & \text{otherwise} \end{cases}$$

If E denotes the event that the battery lasts two or more years, then the random variable I is known as the *indicator* random variable for event E. (Note that I equals 1 or 0 depending on whether or not E occurs.) ■

Example 2.5 Suppose that independent trials, each of which results in any of m possible outcomes with respective probabilities $p_1, \ldots, p_m$, $\sum_{i=1}^{m} p_i = 1$, are continually performed. Let X denote the number of trials needed until each outcome has occurred at least once.

Rather than directly considering $P\{X = n\}$ we will first determine $P\{X > n\}$, the probability that at least one of the outcomes has not yet occurred after n trials. Letting A_i denote the event that outcome i has not yet occurred after the first n trials, $i = 1, \ldots, m$, then

$$\begin{aligned} P\{X > n\} &= P\left(\bigcup_{i=1}^{m} A_i\right) \\ &= \sum_{i=1}^{m} P(A_i) - \sum\sum_{i<j} P(A_i A_j) \\ &\quad + \sum\sum\sum_{i<j<k} P(A_i A_j A_k) - \cdots + (-1)^{m+1} P(A_1 \cdots A_m) \end{aligned}$$

Now, $P(A_i)$ is the probability that each of the first n trials results in a non-i outcome, and so by independence

$$P(A_i) = (1 - p_i)^n$$

Similarly, $P(A_i A_j)$ is the probability that the first n trials all result in a non-i and non-j outcome, and so

$$P(A_i A_j) = (1 - p_i - p_j)^n$$

As all of the other probabilities are similar, we see that

$$\begin{aligned} P\{X > n\} &= \sum_{i=1}^{m} (1 - p_i)^n - \sum\sum_{i<j} (1 - p_i - p_j)^n \\ &\quad + \sum\sum\sum_{i<j<k} (1 - p_i - p_j - p_k)^n - \cdots \end{aligned}$$

Since $P\{X = n\} = P\{X > n-1\} - P\{X > n\}$, we see, upon using the algebraic identity $(1-a)^{n-1} - (1-a)^n = a(1-a)^{n-1}$, that

$$P\{X = n\} = \sum_{i=1}^{m} p_i(1-p_i)^{n-1} - \sum\sum_{i<j}(p_i + p_j)(1 - p_i - p_j)^{n-1}$$
$$+\sum\sum\sum_{i<j<k}(p_i + p_j + p_k)(1 - p_i - p_j - p_k)^{n-1} - \cdots \qquad \blacksquare$$

In all of the preceding examples, the random variables of interest took on either a finite or a countable number of possible values.* Such random variables are called *discrete*. However, there also exist random variables that take on a continuum of possible values. These are known as *continuous* random variables. One example is the random variable denoting the lifetime of a car, when the car's lifetime is assumed to take on any value in some interval (a, b).

The *cumulative distribution function* (cdf) (or more simply the *distribution function*) $F(\cdot)$ of the random variable X is defined for any real number b, $-\infty < b < \infty$, by

$$F(b) = P\{X \le b\}$$

In words, $F(b)$ denotes the probability that the random variable X takes on a value that is less than or equal to b. Some properties of the cdf F are

(i) $F(b)$ is a nondecreasing function of b,
(ii) $\lim_{b\to\infty} F(b) = F(\infty) = 1$,
(iii) $\lim_{b\to-\infty} F(b) = F(-\infty) = 0$.

Property (i) follows since for $a < b$ the event $\{X \le a\}$ is contained in the event $\{X \le b\}$, and so it must have a smaller probability. Properties (ii) and (iii) follow since X must take on some finite value.

All probability questions about X can be answered in terms of the cdf $F(\cdot)$. For example,

$$P\{a < X \le b\} = F(b) - F(a) \qquad \text{for all } a < b$$

This follows since we may calculate $P\{a < X \le b\}$ by first computing the probability that $X \le b$ (that is, $F(b)$) and then subtracting from this the probability that $X \le a$ (that is, $F(a)$).

If we desire the probability that X is strictly smaller than b, we may calculate this probability by

$$\begin{aligned} P\{X < b\} &= \lim_{h\to 0^+} P\{X \le b - h\} \\ &= \lim_{h\to 0^+} F(b-h) \end{aligned}$$

where $\lim_{h\to 0^+}$ means that we are taking the limit as h decreases to 0. Note that $P\{X < b\}$ does not necessarily equal $F(b)$ since $F(b)$ also includes the probability that X equals b.

* A set is countable if its elements can be put in a one-to-one correspondence with the sequence of positive integers.

2.2 Discrete Random Variables

As was previously mentioned, a random variable that can take on at most a countable number of possible values is said to be *discrete*. For a discrete random variable X, we define the *probability mass function* $p(a)$ of X by

$$p(a) = P\{X = a\}$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a. That is, if X must assume one of the values $x_1, x_2, \ldots$, then

$$\begin{aligned} p(x_i) &> 0, \qquad i = 1, 2, \ldots \\ p(x) &= 0, \qquad \text{all other values of } x \end{aligned}$$

Since X must take on one of the values x_i, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

The cumulative distribution function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{\text{all } x_i \le a} p(x_i)$$

For instance, suppose X has a probability mass function given by

$$p(1) = \tfrac{1}{2}, \qquad p(2) = \tfrac{1}{3}, \qquad p(3) = \tfrac{1}{6}$$

then, the cumulative distribution function F of X is given by

$$F(a) = \begin{cases} 0, & a < 1 \\ \frac{1}{2}, & 1 \le a < 2 \\ \frac{5}{6}, & 2 \le a < 3 \\ 1, & 3 \le a \end{cases}$$

This is graphically presented in Figure 2.1.

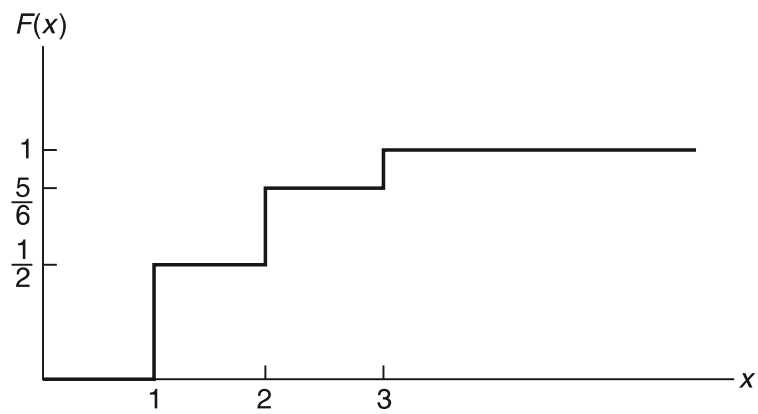


Figure 2.1 Graph of $F(x)$.

Discrete random variables are often classified according to their probability mass functions. We now consider some of these random variables.

2.2.1 The Bernoulli Random Variable

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 equal 1 if the outcome is a success and 0 if it is a failure, then the probability mass function of X is given by

$$\begin{aligned} p(0) &= P\{X = 0\} = 1 - p, \\ p(1) &= P\{X = 1\} = p \end{aligned} \tag{2.2}$$

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 if its probability mass function is given by Equation (2.2) for some $p \in (0, 1)$.

2.2.2 The Binomial Random Variable

Suppose 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 having parameters (n, p) is given by

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i}, \qquad i = 0, 1, \ldots, n \tag{2.3}$$

where

$$\binom{n}{i} = \frac{n!}{(n-i)!\,i!}$$

equals the number of different groups of i objects that can be chosen from a set of n objects. The validity of Equation (2.3) 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 (2.3) then follows since there are $\binom{n}{i}$ different sequences of the n outcomes leading to i successes and $n - i$ failures. For instance, if $n = 3, i = 2$, then there are $\binom{3}{2} = 3$ ways in which the three trials can result in two successes. Namely, any one of the three outcomes (s, s, f), (s, f, s), (f, s, s), where the outcome (s, s, f) means that the first two trials are successes and the third a failure. Since each of the three outcomes (s, s, f), (s, f, s), (f, s, s) has a probability $p^2(1-p)$ of occurring the desired probability is thus $\binom{3}{2} p^2(1-p)$.

Note that, by the binomial theorem, the probabilities sum to one, that is,

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

Example 2.6 Four fair coins are flipped. If the outcomes are assumed independent, what is the probability that two heads and two tails are obtained?

Solution: Letting X equal the number of heads ("successes") that appear, then X is a binomial random variable with parameters $(n = 4,\ p = \frac{1}{2})$. Hence, by Equation (2.3),

$$P\{X = 2\} = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 = \frac{3}{8} \qquad \blacksquare$$

Example 2.7 It is known that any item produced by a certain machine will be defective with probability 0.1, independently of any other item. What is the probability that in a sample of three items, at most one will be defective?

Solution: If X is the number of defective items in the sample, then X is a binomial random variable with parameters (3, 0.1). Hence, the desired probability is given by

$$P\{X = 0\} + P\{X = 1\} = \binom{3}{0}(0.1)^0(0.9)^3 + \binom{3}{1}(0.1)^1(0.9)^2 = 0.972 \qquad \blacksquare$$

Example 2.8 Suppose that an airplane engine will fail, when in flight, with probability $1 - p$ independently from engine to engine; suppose that the airplane will make a successful flight if at least 50 percent of its engines remain operative. For what values of p is a four-engine plane preferable to a two-engine plane?

Solution: Because each engine is assumed to fail or function independently of what happens with the other engines, it follows that the number of engines remaining operative is a binomial random variable. Hence, the probability that a four-engine plane makes a successful flight is

$$\begin{aligned}&\binom{4}{2} p^2(1-p)^2 + \binom{4}{3} p^3(1-p) + \binom{4}{4} p^4(1-p)^0\\ &\quad = 6p^2(1-p)^2 + 4p^3(1-p) + p^4\end{aligned}$$

whereas the corresponding probability for a two-engine plane is

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

Hence the four-engine plane is safer if

$$6p^2(1-p)^2 + 4p^3(1-p) + p^4 \geq 2p(1-p) + p^2$$

or equivalently if

$$6p(1-p)^2 + 4p^2(1-p) + p^3 \geq 2 - p$$

which simplifies to

$$3p^3 - 8p^2 + 7p - 2 \geq 0 \quad \text{or} \quad (p-1)^2(3p-2) \geq 0$$

which is equivalent to

$$3p - 2 \geq 0 \quad \text{or} \quad p \geq \tfrac{2}{3}$$

Hence, the four-engine plane is safer when the engine success probability is at least as large as $\frac{2}{3}$,whereas the two-engine plane is safer when this probability falls below $\frac{2}{3}$. ■

Example 2.9 Suppose that a particular trait of a person (such as eye color or left handedness) 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 one gene from each parent. If, with respect to a particular trait, two hybrid parents have a total of four children, what is the probability that exactly three of the four children have the outward appearance of the dominant gene?

Solution: If we assume that each child is equally likely to inherit either of two genes from each parent, the probabilities that the child of two hybrid parents will have dd, rr, or rd pairs of genes are, respectively, $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$. Hence, because an offspring will have the outward appearance of the dominant gene if its gene pair is either dd or rd, it follows that the number of such children is binomially distributed with parameters $(4, \frac{3}{4})$. Thus the desired probability is

$$\binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1 = \frac{27}{64}$$ ■

Remark on Terminology If X is a binomial random variable with parameters (n, p), then we say that X has a binomial distribution with parameters (n, p).

2.2.3 The Geometric Random Variable

Suppose that independent trials, each having probability p of being a success, are performed until a success occurs. If we let X be the number of trials required until the first success, then X is said to be a *geometric* random variable with parameter p. Its probability mass function is given by

$$p(n) = P\{X = n\} = (1-p)^{n-1}p, \qquad n = 1, 2, \ldots \tag{2.4}$$

Equation (2.4) follows since in order for X to equal n it is necessary and sufficient that the first $n-1$ trials be failures and the nth trial a success. Equation (2.4) follows since the outcomes of the successive trials are assumed to be independent.

To check that $p(n)$ is a probability mass function, we note that

$$\sum_{n=1}^{\infty} p(n) = p \sum_{n=1}^{\infty} (1-p)^{n-1} = 1$$

2.2.4 The Poisson Random Variable

A random variable X, taking on one of the values $0, 1, 2, \ldots$, is said to be a *Poisson* random variable with parameter λ, if for some $\lambda > 0$,

$$p(i) = P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}, \qquad i = 0, 1, \ldots \tag{2.5}$$

Equation (2.5) defines a probability mass function since

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

The Poisson random variable has a wide range of applications in a diverse number of areas, as will be seen in Chapter 5.

An important property of the Poisson random variable is that it may be used to approximate a binomial random variable when the binomial parameter 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

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

Now, for n large and p small

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

Hence, for n large and p small,

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

Example 2.10 Suppose that the number of typographical errors on a single page of this book has a Poisson distribution with parameter $\lambda = 1$. Calculate the probability that there is at least one error on this page.

Solution:

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1} \approx 0.633 \qquad \blacksquare$$

Example 2.11 If the number of accidents occurring on a highway each day is a Poisson random variable with parameter $\lambda = 3$, what is the probability that no accidents occur today?

Solution:

$$P\{X = 0\} = e^{-3} \approx 0.05 \qquad \blacksquare$$

Example 2.12 Consider an experiment that consists of counting the number of α-particles given off in a one-second interval by one 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 two 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

$$P\{X \leq 2\} = e^{-3.2} + 3.2e^{-3.2} + \frac{(3.2)^2}{2}e^{-3.2} \approx 0.382 \qquad \blacksquare$$

2.3 Continuous Random Variables

In this section, we shall concern ourselves with random variables whose set of possible values is uncountable. Let X be such a random variable. We say that X is a *continuous* random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x)\,dx \tag{2.6}$$

The function $f(x)$ is called the *probability density function* of the random variable X.

In words, Equation (2.6) states that the probability that X will be in B may be obtained by integrating the probability density function over the set B. Since X must assume some value, $f(x)$ must satisfy

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\,dx$$

All probability statements about X can be answered in terms of $f(x)$. For instance, letting $B = [a, b]$, we obtain from Equation (2.6) that

$$P\{a \leq X \leq b\} = \int_a^b f(x)\,dx \tag{2.7}$$

If we let $a = b$ in the preceding, then

$$P\{X = a\} = \int_a^a f(x)\,dx = 0$$

In words, this equation states that the probability that a continuous random variable will assume any *particular* value is zero.

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x)\,dx$$

Differentiating both sides of the preceding yields

$$\frac{d}{da}F(a) = f(a)$$

That is, the density is the derivative of the cumulative distribution function. A somewhat more intuitive interpretation of the density function may be obtained from Equation (2.7) as follows:

$$P\left\{a - \frac{\varepsilon}{2} \le X \le a + \frac{\varepsilon}{2}\right\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)\,dx \approx \varepsilon f(a)$$

when ε is small. In other words, the probability that X will be contained in an interval of length ε around the point a is approximately $\varepsilon f(a)$. From this, we see that $f(a)$ is a measure of how likely it is that the random variable will be near a.

There are several important continuous random variables that appear frequently in probability theory. The remainder of this section is devoted to a study of certain of these random variables.

2.3.1 The Uniform Random Variable

A random variable is said to be *uniformly distributed* over the interval $(0, 1)$ if its probability density function is given by

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Note that the preceding is a density function since $f(x) \ge 0$ and

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 dx = 1$$

Since $f(x) > 0$ only when $x \in (0, 1)$, it follows that X must assume a value in $(0, 1)$. Also, since $f(x)$ is constant for $x \in (0, 1)$, X is just as likely to be "near" any value in $(0, 1)$ as any other value. To check this, note that, for any $0 < a < b < 1$,

$$P\{a \le X \le b\} = \int_a^b f(x)\,dx = b - a$$

In other words, the probability that X is in any particular subinterval of $(0, 1)$ equals the length of that subinterval.

In general, we say that X is a uniform random variable on the interval (α, β) if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \text{if } \alpha < x < \beta \\ 0, & \text{otherwise} \end{cases} \tag{2.8}$$

Example 2.13 Calculate the cumulative distribution function of a random variable uniformly distributed over (α, β).

Solution: Since $F(a) = \int_{-\infty}^{a} f(x)\,dx$, we obtain from Equation (2.8) that

$$F(a) = \begin{cases} 0, & a \le \alpha \\ \dfrac{a - \alpha}{\beta - \alpha}, & \alpha < a < \beta \\ 1, & a \ge \beta \end{cases}$$

■

Example 2.14 If X is uniformly distributed over $(0, 10)$, calculate the probability that (a) $X < 3$, (b) $X > 7$, (c) $1 < X < 6$.

Solution:

$$P\{X < 3\} = \frac{\int_0^3 dx}{10} = \frac{3}{10},$$

$$P\{X > 7\} = \frac{\int_7^{10} dx}{10} = \frac{3}{10},$$

$$P\{1 < X < 6\} = \frac{\int_1^6 dx}{10} = \frac{1}{2}$$

■

2.3.2 Exponential Random Variables

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

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

is said to be an *exponential random variable* with parameter λ. These random variables will be extensively studied in Chapter 5, so we will content ourselves here with just calculating the cumulative distribution function F:

$$F(a) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}, \qquad a \ge 0$$

Note that $F(\infty) = \int_0^\infty \lambda e^{-\lambda x} dx = 1$, as, of course, it must.

2.3.3 Gamma Random Variables

A continuous random variable whose density is given by

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

Exclude 2.3.3

for some $\lambda > 0, \alpha > 0$ is said to be a *gamma random variable* with parameters α, λ. The quantity $\Gamma(\alpha)$ is called the gamma function and is defined by

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1}\, dx$$

It is easy to show by induction that for integral α, say, $\alpha = n$,

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

2.3.4 Normal Random Variables

We say that X is a *normal random variable* (or simply that X is normally distributed) with parameters μ and σ^2 if the density of X is given by

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

This density function is a bell-shaped curve that is symmetric around μ (see Figure 2.2).

An important fact about normal random variables is that if X is normally distributed with parameters μ and σ^2 then $Y = \alpha X + \beta$ is normally distributed with parameters $\alpha\mu + \beta$ and $\alpha^2\sigma^2$. To prove this, suppose first that $\alpha > 0$ and note that $F_Y(\cdot)$,* the cumulative distribution function of the random variable Y, is given by

$$\begin{aligned} F_Y(a) &= P\{Y \leq a\} \\ &= P\{\alpha X + \beta \leq a\} \end{aligned}$$

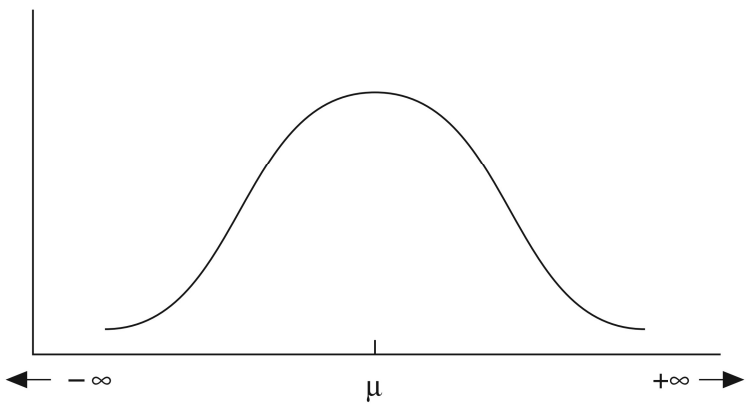


Figure 2.2 Normal density function.

* When there is more than one random variable under consideration, we shall denote the cumulative distribution function of a random variable Z by $F_Z(\cdot)$. Similarly, we shall denote the density of Z by $f_Z(\cdot)$.

$$
\begin{aligned}
&= P\left\{X \le \frac{a-\beta}{\alpha}\right\} \\
&= F_X\left(\frac{a-\beta}{\alpha}\right) \\
&= \int_{-\infty}^{(a-\beta)/\alpha} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}\,dx \\
&= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\,\alpha\sigma} \exp\left\{\frac{-(v-(\alpha\mu+\beta))^2}{2\alpha^2\sigma^2}\right\} dv \qquad (2.9)
\end{aligned}
$$

where the last equality is obtained by the change in variables $v = \alpha x + \beta$. However, since $F_Y(a) = \int_{-\infty}^{a} f_Y(v)\,dv$, it follows from Equation (2.9) that the probability density function $f_Y(\cdot)$ is given by

$$
f_Y(v) = \frac{1}{\sqrt{2\pi}\alpha\sigma} \exp\left\{\frac{-(v-(\alpha\mu+\beta))^2}{2(\alpha\sigma)^2}\right\}, \qquad -\infty < v < \infty
$$

Hence, Y is normally distributed with parameters $\alpha\mu + \beta$ and $(\alpha\sigma)^2$. A similar result is also true when $\alpha < 0$.

One implication of the preceding result is that if X is normally distributed with parameters μ and σ^2 then $Y = (X - \mu)/\sigma$ is normally distributed with parameters 0 and 1. Such a random variable Y is said to have the *standard* or *unit* normal distribution.

2.4 Expectation of a Random Variable

2.4.1 The Discrete Case

If X is a discrete random variable having a probability mass function $p(x)$, then the *expected value* of X is defined by

$$
E[X] = \sum_{x:p(x)>0} x p(x)
$$

In other words, the expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes that value. For example, if the probability mass function of X is given by

$$
p(1) = \tfrac{1}{2} = p(2)
$$

then

$$
E[X] = 1\left(\tfrac{1}{2}\right) + 2\left(\tfrac{1}{2}\right) = \tfrac{3}{2}
$$

is just an ordinary average of the two possible values 1 and 2 that X can assume. On the other hand, if

$$
p(1) = \tfrac{1}{3}, \qquad p(2) = \tfrac{2}{3}
$$

then

$$E[X] = 1\left(\tfrac{1}{3}\right) + 2\left(\tfrac{2}{3}\right) = \tfrac{5}{3}$$

is a weighted average of the two possible values 1 and 2 where the value 2 is given twice as much weight as the value 1 since $p(2) = 2p(1)$.

Example 2.15 Find $E[X]$ where X is the outcome when we roll a fair die.

Solution: Since $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$, we obtain

$$E[X] = 1\left(\tfrac{1}{6}\right) + 2\left(\tfrac{1}{6}\right) + 3\left(\tfrac{1}{6}\right) + 4\left(\tfrac{1}{6}\right) + 5\left(\tfrac{1}{6}\right) + 6\left(\tfrac{1}{6}\right) = \tfrac{7}{2}$$ ■

Example 2.16 (Expectation of a Bernoulli Random Variable) Calculate $E[X]$ when X is a Bernoulli random variable with parameter p.

Solution: Since $p(0) = 1 - p$, $p(1) = p$,we have

$$E[X] = 0(1-p) + 1(p) = p$$

Thus, the expected number of successes in a single trial is just the probability that the trial will be a success. ■

Example 2.17 (Expectation of a Binomial Random Variable) Calculate $E[X]$ when X is binomially distributed with parameters n and p.

Solution:

$$\begin{aligned}
E[X] &= \sum_{i=0}^{n} i p(i) \\
&= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{in!}{(n-i)!\,i!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{n!}{(n-i)!(i-1)!} p^i (1-p)^{n-i} \\
&= np \sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!(i-1)!} p^{i-1} (1-p)^{n-i} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\
&= np[p + (1-p)]^{n-1} \\
&= np
\end{aligned}$$

where the second from the last equality follows by letting $k = i - 1$. Thus, the expected number of successes in n independent trials is n multiplied by the probability that a trial results in a success. ■

Example 2.18 (Expectation of a Geometric Random Variable) Calculate the expectation of a geometric random variable having parameter p.

Solution: By Equation (2.4), we have

$$\begin{aligned} E[X] &= \sum_{n=1}^{\infty} np(1-p)^{n-1} \\ &= p\sum_{n=1}^{\infty} nq^{n-1} \end{aligned}$$

where $q = 1 - p$,

$$\begin{aligned} E[X] &= p\sum_{n=1}^{\infty} \frac{d}{dq}(q^n) \\ &= p\frac{d}{dq}\left(\sum_{n=1}^{\infty} q^n\right) \\ &= p\frac{d}{dq}\left(\frac{q}{1-q}\right) \\ &= \frac{p}{(1-q)^2} \\ &= \frac{1}{p} \end{aligned}$$

In words, the expected number of independent trials we need to perform until we attain our first success equals the reciprocal of the probability that any one trial results in a success. ■

Example 2.19 (Expectation of a Poisson Random Variable) Calculate $E[X]$ if X is a Poisson random variable with parameter λ.

Solution: From Equation (2.5), we have

$$\begin{aligned} E[X] &= \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} \\ &= \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^i}{(i-1)!} \\ &= \lambda e^{-\lambda}\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \end{aligned}$$

$$
\begin{aligned}
&= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}
$$

where we have used the identity $\sum_{k=0}^{\infty} \lambda^k / k! = e^{\lambda}$. ■

2.4.2 The Continuous Case

We may also define the expected value of a continuous random variable. This is done as follows. If X is a continuous random variable having a probability density function $f(x)$, then the expected value of X is defined by

$$
E[X] = \int_{-\infty}^{\infty} x f(x)\, dx
$$

Example 2.20 (Expectation of a Uniform Random Variable) Calculate the expectation of a random variable uniformly distributed over (α, β).

Solution: From Equation (2.8) we have

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

In other words, the expected value of a random variable uniformly distributed over the interval (α, β) is just the midpoint of the interval. ■

Example 2.21 (Expectation of an Exponential Random Variable) Let X be exponentially distributed with parameter λ. Calculate $E[X]$.

Solution:

$$
E[X] = \int_{0}^{\infty} x \lambda e^{-\lambda x}\, dx
$$

Integrating by parts ($dv = \lambda e^{-\lambda x}$, $u = x$) yields

$$
\begin{aligned}
E[X] &= -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx \\
&= 0 - \frac{e^{-\lambda x}}{\lambda} \Bigg|_0^{\infty} \\
&= \frac{1}{\lambda}
\end{aligned}
$$

■

Example 2.22 (Expectation of a Normal Random Variable) Calculate $E[X]$ when X is normally distributed with parameters μ and σ^2.

Solution:

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

Writing x as $(x - \mu) + \mu$ yields

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

Letting $y = x - \mu$ leads to

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

where $f(x)$ is the normal density. By symmetry, the first integral must be 0, and so

$$E[X] = \mu \int_{-\infty}^{\infty} f(x)\, dx = \mu$$ ■

2.4.3 Expectation of a Function of a Random Variable

Suppose now that we are given a random variable X and its probability distribution (that is, its probability mass function in the discrete case or its probability density function in the continuous case). Suppose also that we are interested in calculating not the expected value of X, but the expected value of some function of X, say, $g(X)$. How do we go about doing this? One way is as follows. Since $g(X)$ is itself a random variable, it must have a probability distribution, which should be computable from a knowledge of the distribution of X. Once we have obtained the distribution of $g(X)$, we can then compute $E[g(X)]$ by the definition of the expectation.

Example 2.23 Suppose X has the following probability mass function:

$$p(0) = 0.2, \qquad p(1) = 0.5, \qquad p(2) = 0.3$$

Calculate $E[X^2]$.

Solution: Letting $Y = X^2$, we have that Y is a random variable that can take on one of the values 0^2, 1^2, 2^2 with respective probabilities

$$\begin{aligned} p_Y(0) &= P\{Y = 0^2\} = 0.2, \\ p_Y(1) &= P\{Y = 1^2\} = 0.5, \\ p_Y(4) &= P\{Y = 2^2\} = 0.3 \end{aligned}$$

Hence,

$$E[X^2] = E[Y] = 0(0.2) + 1(0.5) + 4(0.3) = 1.7$$

Note that

$$1.7 = E[X^2] \neq (E[X])^2 = 1.21 \qquad \blacksquare$$

Example 2.24 Let X be uniformly distributed over $(0, 1)$. Calculate $E[X^3]$.

Solution: Letting $Y = X^3$, we calculate the distribution of Y as follows. For $0 \leq a \leq 1$,

$$\begin{aligned} F_Y(a) &= P\{Y \leq a\} \\ &= P\{X^3 \leq a\} \\ &= P\{X \leq a^{1/3}\} \\ &= a^{1/3} \end{aligned}$$

where the last equality follows since X is uniformly distributed over $(0, 1)$. By differentiating $F_Y(a)$, we obtain the density of Y, namely,

$$f_Y(a) = \tfrac{1}{3}a^{-2/3}, \qquad 0 \leq a \leq 1$$

Hence,

$$\begin{aligned} E[X^3] = E[Y] &= \int_{-\infty}^{\infty} a f_Y(a)\, da \\ &= \int_0^1 a \tfrac{1}{3} a^{-2/3}\, da \\ &= \tfrac{1}{3} \int_0^1 a^{1/3}\, da \\ &= \tfrac{1}{3}\tfrac{3}{4} a^{4/3} \Big|_0^1 \\ &= \tfrac{1}{4} \end{aligned} \qquad \blacksquare$$

While the foregoing procedure will, in theory, always enable us to compute the expectation of any function of X from a knowledge of the distribution of X, there is, fortunately, an easier way to do this. The following proposition shows how we can calculate the expectation of $g(X)$ without first determining its distribution.

Proposition 2.1 (a) If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g,

$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$$

(b) If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx \qquad \blacksquare$$

Example 2.25 Applying the proposition to Example 2.23 yields

$$E[X^2] = 0^2(0.2) + (1^2)(0.5) + (2^2)(0.3) = 1.7$$

which, of course, checks with the result derived in Example 2.23. ■

Example 2.26 Applying the proposition to Example 2.24 yields

$$\begin{aligned} E[X^3] &= \int_0^1 x^3\,dx \qquad (\text{since } f(x) = 1, 0 < x < 1) \\ &= \tfrac{1}{4} \end{aligned}$$

■

A simple corollary of Proposition 2.1 is the following.

Corollary 2.2 If a and b are constants, then

$$E[aX + b] = aE[X] + b$$

Proof. In the discrete case,

$$\begin{aligned} E[aX + b] &= \sum_{x:p(x)>0} (ax + b)p(x) \\ &= a \sum_{x:p(x)>0} xp(x) + b \sum_{x:p(x)>0} p(x) \\ &= aE[X] + b \end{aligned}$$

In the continuous case,

$$\begin{aligned} E[aX + b] &= \int_{-\infty}^{\infty} (ax + b)f(x)\,dx \\ &= a\int_{-\infty}^{\infty} xf(x)\,dx + b\int_{-\infty}^{\infty} f(x)\,dx \\ &= aE[X] + b \end{aligned}$$

■

The expected value of a random variable X, $E[X]$, is also referred to as the *mean* or the first *moment* of X. The quantity $E[X^n]$, $n \geq 1$, is called the nth moment of X. By Proposition 2.1, we note that

$$E[X^n] = \begin{cases} \displaystyle\sum_{x:p(x)>0} x^n p(x), & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} x^n f(x)\,dx, & \text{if } X \text{ is continuous} \end{cases}$$

Another quantity of interest is the variance of a random variable X, denoted by $\mathrm{Var}(X)$, which is defined by

$$\mathrm{Var}(X) = E\Big[(X - E[X])^2\Big]$$

Thus, the variance of X measures the expected square of the deviation of X from its expected value.

Example 2.27 (Variance of the Normal Random Variable) Let X be normally distributed with parameters μ and σ^2. Find $\mathrm{Var}(X)$.

Solution: Recalling (see Example 2.22) that $E[X] = \mu$, we have that

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

Substituting $y = (x-\mu)/\sigma$ yields

$$\mathrm{Var}(X) = \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-y^2/2}\,dy$$

Integrating by parts ($u = y$, $dv = ye^{-y^2/2}dy$) gives

$$\begin{aligned}\mathrm{Var}(X) &= \frac{\sigma^2}{\sqrt{2\pi}}\left(-ye^{-y^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-y^2/2}\,dy\right) \\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}\,dy \\ &= \sigma^2\end{aligned}$$

Another derivation of $\mathrm{Var}(X)$ will be given in Example 2.42. ■

Suppose that X is continuous with density f, and let $E[X] = \mu$. Then,

$$\begin{aligned}\mathrm{Var}(X) &= E[(X-\mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= \int_{-\infty}^{\infty}(x^2 - 2\mu x + \mu^2) f(x)\,dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)\,dx - 2\mu\int_{-\infty}^{\infty} x f(x)\,dx + \mu^2\int_{-\infty}^{\infty} f(x)\,dx \\ &= E[X^2] - 2\mu\mu + \mu^2 \\ &= E[X^2] - \mu^2\end{aligned}$$

A similar proof holds in the discrete case, and so we obtain the useful identity

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

Example 2.28 Calculate $\mathrm{Var}(X)$ when X represents the outcome when a fair die is rolled.

Solution: As previously noted in Example 2.15, $E[X] = \frac{7}{2}$. Also,

$$E[X^2] = 1\left(\tfrac{1}{6}\right) + 2^2\left(\tfrac{1}{6}\right) + 3^2\left(\tfrac{1}{6}\right) + 4^2\left(\tfrac{1}{6}\right) + 5^2\left(\tfrac{1}{6}\right) + 6^2\left(\tfrac{1}{6}\right) = \left(\tfrac{1}{6}\right)(91)$$

Hence,

$$\mathrm{Var}(X) = \tfrac{91}{6} - \left(\tfrac{7}{2}\right)^2 = \tfrac{35}{12}$$ ■

2.8 Limit Theorems

We start this section by proving a result known as Markov's inequality.

Proposition 2.6 (Markov's Inequality) If X is a random variable that takes only nonnegative values, then for any value $a > 0$

$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

Proof. We give a proof for the case where X is continuous with density f.

$$\begin{aligned}
E[X] &= \int_0^\infty x f(x)\, dx \\
&= \int_0^a x f(x)\, dx + \int_a^\infty x f(x)\, dx \\
&\geq \int_a^\infty x f(x)\, dx \\
&\geq \int_a^\infty a f(x)\, dx \\
&= a \int_a^\infty f(x)\, dx \\
&= a P\{X \geq a\}
\end{aligned}$$

and the result is proven.

As a corollary, we obtain the following.

Proposition 2.7 (Chebyshev's Inequality) If X is a random variable with mean μ and variance σ^2, then, for any value $k > 0$,

$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Proof. Since $(X - \mu)^2$ is a nonnegative random variable, we can apply Markov's inequality (with $a = k^2$) to obtain

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E[(X - \mu)^2]}{k^2}$$

But since $(X - \mu)^2 \geq k^2$ if and only if $|X - \mu| \geq k$, the preceding is equivalent to

$$P\{|X - \mu| \geq k\} \leq \frac{E[(X - \mu)^2]}{k^2} = \frac{\sigma^2}{k^2}$$

and the proof is complete.

The importance of Markov's and Chebyshev's inequalities is that they enable us to derive bounds on probabilities when only the mean, or both the mean and the variance, of the probability distribution are known. Of course, if the actual distribution were known, then the desired probabilities could be exactly computed, and we would not need to resort to bounds.

Example 2.49 Suppose we know that the number of items produced in a factory during a week is a random variable with mean 500.

(a) What can be said about the probability that this week's production will be at least 1000?
(b) If the variance of a week's production is known to equal 100, then what can be said about the probability that this week's production will be between 400 and 600?

Solution: Let X be the number of items that will be produced in a week.

(a) By Markov's inequality,

$$P\{X \geq 1000\} \leq \frac{E[X]}{1000} = \frac{500}{1000} = \frac{1}{2}$$

(b) By Chebyshev's inequality,

$$P\{|X - 500| \geq 100\} \leq \frac{\sigma^2}{(100)^2} = \frac{1}{100}$$

Hence,

$$P\{|X - 500| < 100\} \geq 1 - \frac{1}{100} = \frac{99}{100}$$

and so the probability that this week's production will be between 400 and 600 is at least 0.99. ■

The following theorem, known as the *strong law of large numbers*, is probably the most well-known result in probability theory. It states that the average of a sequence of independent random variables having the same distribution will, with probability 1, converge to the mean of that distribution.

Theorem 2.1 (Strong Law of Large Numbers) Let $X_1, X_2, \ldots$ be a sequence of independent random variables having a common distribution, and let $E[X_i] = \mu$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \qquad \text{as } n \to \infty$$

As an example of the preceding, suppose that a sequence of independent trials is performed. Let E be a fixed event and denote by $P(E)$ the probability that E occurs on any particular trial. Letting

$$X_i = \begin{cases} 1, & \text{if } E \text{ occurs on the } i\text{th trial} \\ 0, & \text{if } E \text{ does not occur on the } i\text{th trial} \end{cases}$$

we have by the strong law of large numbers that, with probability 1,

$$\frac{X_1 + \cdots + X_n}{n} \to E[X] = P(E) \tag{2.25}$$

Since $X_1 + \cdots + X_n$ represents the number of times that the event E occurs in the first n trials, we may interpret Equation (2.25) as stating that, with probability 1, the limiting proportion of time that the event E occurs is just $P(E)$.

Running neck and neck with the strong law of large numbers for the honor of being probability theory's number one result is the *central limit theorem*. Besides its theoretical interest and importance, this theorem provides a simple method for computing approximate probabilities for sums of independent random variables. It also explains the remarkable fact that the empirical frequencies of so many natural "populations" exhibit a bell-shaped (that is, normal) curve.

Theorem 2.2 (Central Limit Theorem) Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables, each with mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal as $n \to \infty$. That is,

$$P\left\{\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le a\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}\,dx$$

as $n \to \infty$.

Note that like the other results of this section, this theorem holds for *any* distribution of the X_is; herein lies its power.

If X is binomially distributed with parameters n and p, then X has the same distribution as the sum of n independent Bernoulli random variables, each with parameter p. (Recall that the Bernoulli random variable is just a binomial random variable whose parameter n equals 1.) Hence, the distribution of

$$\frac{X - E[X]}{\sqrt{\mathrm{Var}(X)}} = \frac{X - np}{\sqrt{np(1-p)}}$$

approaches the standard normal distribution as n approaches ∞. The normal approximation will, in general, be quite good for values of n satisfying $np(1-p) \ge 10$.

Example 2.50 (Normal Approximation to the Binomial) Let X be the number of times that a fair coin, flipped 40 times, lands heads. Find the probability that $X = 20$. Use the normal approximation and then compare it to the exact solution.

Solution: Since the binomial is a discrete random variable, and the normal a continuous random variable, it leads to a better approximation to write the desired probability as

$$\begin{aligned}
P\{X = 20\} &= P\{19.5 < X < 20.5\} \\
&= P\left\{\frac{19.5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} < \frac{20.5 - 20}{\sqrt{10}}\right\} \\
&= P\left\{-0.16 < \frac{X - 20}{\sqrt{10}} < 0.16\right\} \\
&\approx \Phi(0.16) - \Phi(-0.16)
\end{aligned}$$

where $\Phi(x)$, the probability that the standard normal is less than x is given by

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

By the symmetry of the standard normal distribution

$$\Phi(-0.16) = P\{N(0, 1) > 0.16\} = 1 - \Phi(0.16)$$

where $N(0, 1)$ is a standard normal random variable. Hence, the desired probability is approximated by

$$P\{X = 20\} \approx 2\Phi(0.16) - 1$$

Using Table 2.3, we obtain

$$P\{X = 20\} \approx 0.1272$$

The exact result is

$$P\{X = 20\} = \binom{40}{20} \left(\frac{1}{2}\right)^{40}$$

which can be shown to equal 0.1268. ■

Example 2.51 Let X_i, $i = 1, 2, \ldots, 10$ be independent random variables, each being uniformly distributed over (0, 1). Estimate $P\{\sum_1^{10} X_i > 7\}$.

Solution: Since $E[X_i] = \frac{1}{2}$, $\mathrm{Var}(X_i) = \frac{1}{12}$ we have by the central limit theorem that

$$\begin{aligned} P\left\{\sum_1^{10} X_i > 7\right\} &= P\left\{\frac{\sum_1^{10} X_i - 5}{\sqrt{10\left(\frac{1}{12}\right)}} > \frac{7-5}{\sqrt{10\left(\frac{1}{12}\right)}}\right\} \\ &\approx 1 - \Phi(2.2) \\ &= 0.0139 \end{aligned}$$

■

Example 2.52 The lifetime of a special type of battery is a random variable with mean 40 hours and standard deviation 20 hours. A battery is used until it fails, at which point it is replaced by a new one. Assuming a stockpile of 25 such batteries, the lifetimes of which are independent, approximate the probability that over 1100 hours of use can be obtained.

Solution: If we let X_i denote the lifetime of the ith battery to be put in use, then we desire $p = P\{X_1 + \cdots + X_{25} > 1100\}$, which is approximated as follows:

$$\begin{aligned} p &= P\left\{\frac{X_1 + \cdots + X_{25} - 1000}{20\sqrt{25}} > \frac{1100 - 1000}{20\sqrt{25}}\right\} \\ &\approx P\{N(0, 1) > 1\} \\ &= 1 - \Phi(1) \\ &\approx 0.1587 \end{aligned}$$

■

Table 2.3 Area $\Phi(x)$ under the Standard Normal Curve to the Left of x

x	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5597	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8557	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

We now present a heuristic proof of the central limit theorem. Suppose first that the X_i have mean 0 and variance 1, and let $E[e^{tX}]$ denote their common moment generating function. Then, the moment generating function of $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ is

$$E\left[\exp\left\{t\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right\}\right] = E[e^{tX_1/\sqrt{n}}e^{tX_2/\sqrt{n}}\cdots e^{tX_n/\sqrt{n}}]$$
$$= (E[e^{tX/\sqrt{n}}])^n \qquad \text{by independence}$$

Now, for n large, we obtain from the Taylor series expansion of e^y that

$$e^{tX/\sqrt{n}} \approx 1 + \frac{tX}{\sqrt{n}} + \frac{t^2X^2}{2n}$$

Taking expectations shows that when n is large

$$E[e^{tX/\sqrt{n}}] \approx 1 + \frac{tE[X]}{\sqrt{n}} + \frac{t^2E[X^2]}{2n}$$
$$= 1 + \frac{t^2}{2n} \qquad \text{because } E[X] = 0,\ E[X^2] = 1$$

Therefore, we obtain that when n is large

$$E\left[\exp\left\{t\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right\}\right] \approx \left(1+\frac{t^2}{2n}\right)^n$$

When n goes to ∞ the approximation can be shown to become exact and we have

$$\lim_{n\to\infty} E\left[\exp\left\{t\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right\}\right] = e^{t^2/2}$$

Thus, the moment generating function of $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the moment generating function of a (standard) normal random variable with mean 0 and variance 1. Using this, it can be proven that the distribution function of the random variable $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the standard normal distribution function Φ.

When the X_i have mean μ and variance σ^2, the random variables $\frac{X_i-\mu}{\sigma}$ have mean 0 and variance 1. Thus, the preceding shows that

$$P\left\{\frac{X_1-\mu+X_2-\mu+\cdots+X_n-\mu}{\sigma\sqrt{n}} \le a\right\} \to \Phi(a)$$

which proves the central limit theorem.

2.9 Stochastic Processes

A *stochastic process* $\{X(t), t \in T\}$ is a collection of random variables. That is, for each $t \in T$, $X(t)$ is a random variable. The index t is often interpreted as time and, as a result, we refer to $X(t)$ as the *state* of the process at time t. For example, $X(t)$ might equal the total number of customers that have entered a supermarket by time t; or the number of customers in the supermarket at time t; or the total amount of sales that have been recorded in the market by time t; etc.