6. A random variable X has a binomial distribution with parameters $n = 4$ and $p = \frac{1}{3}$.
   (a) Describe the probability mass function p and sketch its graph.
   (b) Describe the distribution function F and sketch its graph.
   (c) Compute the probabilities $P(1 < X \le 2)$ and $P(1 \le X \le 2)$.
7. Assume that if a thumbtack is tossed on a table, it lands either with point up or in a stable position with point resting on the table. Assume there is a positive probability p that it lands with point up.
   (a) Suppose two identical tacks are tossed simultaneously. Assuming stochastic independence, show that the probability that both land with point up is p^2.
   (b) Continuing part (a), let X denote the random variable which counts the number of tacks which land with point up (the possible values of X are 0, 1, and 2). Compute the probabilities $P(X = 0)$ and $P(X = 1)$.
   (c) Draw the graph of the distribution function F_X when $p = \frac{1}{3}$.
8. Given a random variable X whose possible values are $1, 2, \ldots, n$. Assume that the probability $P(X = k)$ is proportional to k. Determine the constant of proportionality, the probability mass function p_X, and the distribution function F_X.
9. Given a random variable X whose possible values are $0, 1, 2, 3, \ldots$. Assume that $P(X = k)$ is proportional to $c^k/k!$, where c is a fixed real number. Determine the constant of proportionality and the probability mass function p.
10. (a) A fair die is rolled. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. If the number of points on the upturned face is odd a player receives one dollar; otherwise he must pay one dollar. Let X denote the random variable which measures his financial outcome (number of dollars) on each play of the game. (The possible values of X are $+1$ and -1.) Describe the probability mass function p_X and the distribution F_X. Sketch their graphs.
    (b) A fair coin is tossed. The sample space $S = \{H, T\}$. If the outcome is heads a player receives one dollar; if it is tails he must pay one dollar. Let Y denote the random variable which measures his financial outcome (number of dollars) on each play of the game. Show that the mass function p_Y and the distribution F_Y are identical to those in part (a). This example shows that different random variables may have the same probability distribution function. Actually, there are infinitely many random variables having a given probability distribution F. (Why?) Such random variables are said to be *identically distributed.* Each theorem concerning a particular distribution function is applicable to any of an infinite collection of random variables having this distribution.
11. The number of minutes that one has to wait for a train at a certain subway station is known to be a random variable X with the following probability mass function:

$$p(t) = 0 \qquad \text{unless } t = 3k/10 \quad \text{for some } k = 0, 1, 2, \ldots, 10.$$

$$p(t) = \tfrac{1}{12} \qquad \text{if } t = 0, 0.3, 0.6, 0.9, 2.1, 2.4, 2.7, 3.0.$$

$$p(t) = \tfrac{1}{9} \qquad \text{if } t = 1.2, 1.5, 1.8.$$

    Sketch the graph of the corresponding distribution function F. Let A be the event that one has to wait between 0 and 2 minutes (including 0 and 2), and let B be the event that one has to wait between 1 and 3 minutes (including 1 and 3). Compute the following probabilities: $P(A), P(B), P(A \cap B), P(B \mid A), P(A \cup B)$.
12. (a) If $0 < p < 1$ and $q = 1 - p$, show that

$$\binom{n}{k} p^k q^{n-k} = \frac{(np)^k}{k!}\left(1 - \frac{np}{n}\right)^n Q_n,$$

where

$$Q_n = \frac{\prod_{r=2}^{k}\left(1 - \frac{r-1}{n}\right)}{(1-p)^k}.$$

(b) Given $\lambda > 0$, let $p = \lambda/n$ for $n > \lambda$. Show that $Q_n \to 1$ as $n \to \infty$ and that

$$\binom{n}{k} p^k q^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{as} \quad n \to \infty.$$

This result suggests that for large n and small p, the binomial distribution is approximately the same as the Poisson distribution, provided the product np is nearly constant; this constant is the parameter λ of the Poisson distribution.

14.9 Continuous distributions. Density functions

Let X be a one-dimensional random variable and let F be its distribution function, so that $F(t) = P(X \leq t)$ for every real t. If the probability $P(X = t)$ is zero for every t then, because of Theorem 14.5, F is continuous everywhere on the real axis. In this case F is called a *continuous distribution* and X is called a *continuous random variable.* If the derivative F' exists and is continuous on an interval $[a, t]$ we can use the second fundamental theorem of calculus to write

$$F(t) - F(a) = \int_a^t f(u)\, du, \tag{14.13}$$

where f is the derivative of F. The difference $F(t) - F(a)$ is, of course, the probability $P(a < X \leq t)$, and Equation (14.13) expresses this probability as an integral.

Sometimes the distribution function F can be expressed as an integral of the form (14.13), in which the integrand f is integrable but not necessarily continuous. Whenever an equation such as (14.13) holds for all intervals $[a, t]$, the integrand f is called *a probability density function* of the random variable X (or of the distribution F) provided thatfis nonnegative. In other words, we have the following definition:

DEFINITION OF A PROBABILITY DENSITY FUNCTION. *Let X be a one-dimensional random variable with a continuous distribution function F. A nonnegative function f is called a probability density of X (or of F) if f is integrable on every interval $[a, t]$ and if*

$$F(t) - F(a) = \int_a^t f(u)\, du. \tag{14.14}$$

If we let $a \to -\infty$ in (14.14) then $F(a) \to 0$ and we obtain the important formula

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(u)\, du, \tag{14.15}$$

valid for all real t. If we now let $t \to +\infty$ and remember that $F(t) \to 1$ we find that

$$\int_{-\infty}^{+\infty} f(u)\, du = 1. \tag{14.16}$$

For discrete random variables the sum of all the probabilities $P(X = t)$ is equal to 1. Formula (14.16) is the continuous analog of this statement There is also a strong analogy between formulas (14.11) and (14.15). The density function f plays the same role for continuous distributions that the probability mass function p plays for discrete distributions — integration takes the place of summation in the computation of probabilities. There is one important difference, however. In the discrete case $p(t)$ is the probability that $X = t$, but in the continuous case $f(t)$ is *not* the probability that $X = t$. In fact, this probability is zero because F is continuous for every t. Of course, this also means that for a continuous distribution we have

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b).$$

If F has a density f each of these probabilities is equal to the integral $\int_a^b f(u)\,du$.

Note: A given distribution can have more than one density since the value of the integrand in (14.14) can be changed at a finite number of points without altering the integral. But if f is *continuous* at t then $f(t) = F'(t)$; in this case the value of the density function at t is uniquely determined by F.

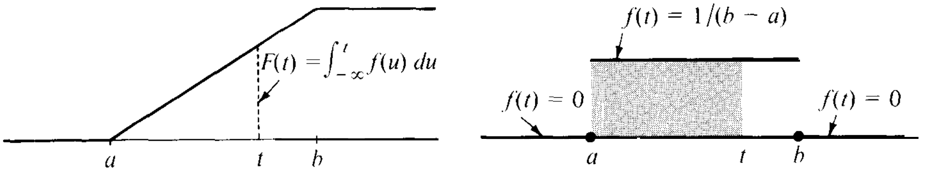
Since f is nonnegative, the right-hand member of Equation (14.14) can be interpreted geometrically as the area of that portion of the ordinate set of f lying to the left of the line $x = t$. The area of the entire ordinate set is equal to 1. The area of the portion of the ordinate set above a given interval (whether it is open, closed, or half-open) is the probability that the random variable X takes on a value in that interval. Figure 14.6 shows an example of a continuous distribution function F and its density function f. The ordinate $F(t)$ in Figure 14.6(a) is equal to the area of the shaded region in Figure 14.6(b).

The next few sections describe some important examples of continuous distributions.

14.10 Uniform distribution over an interval

A one-dimensional random variable X is said to have a uniform distribution function F over a finite interval $[a, b]$ if F is given by the following formulas:

$$F(t) = \begin{cases} 0 & \text{if } t \leq a, \\ \dfrac{t-a}{b-a} & \text{if } a < t < b, \\ 1 & \text{if } t \geq b. \end{cases}$$



(a) The distribution function F. (b) The density function f.

FIGURE 14.6 A uniform distribution over an interval $[a, b]$ and the corresponding density function.

This is a continuous distribution whose graph is shown in Figure 14.6(a).

The derivative $F'(t)$ exists everywhere except at the points a and b, and we can write

$$F(t) = \int_{-\infty}^{t} f(u)\, du ,$$

where f is the density function, defined as follows:

$$f(t) = \begin{cases} 1/(b - a) & \text{if } a < t < b, \\ 0 & \text{otherwise.} \end{cases}$$

The graph off is shown in Figure 14.6(b).

The next theorem characterizes uniform distributions in another way.

THEOREM 14.7. *Let X be a one-dimensional random variable with all its values in a finite interval $[a, b]$, and let F be the distribution function of X. Then F is uniform over $[a, b]$ if and only if*

$$P(X \in I) = P(X \in J) \tag{14.17}$$

for every pair of subintervals Z and J of $[a, b]$ having the same length, in which case we have

$$P(X \in I) = \frac{h}{b - a},$$

where h is the length of I.

Proof. Assume first that X has a uniform distribution over $[a, b]$. If $[c, c + h]$ is any subinterval of $[a, b]$ of length h we have

$$P(c \le X \le c + h) = F(c + h) - F(c) = \frac{c + h - a}{b - a} - \frac{c - a}{b - a} = \frac{h}{b - a}.$$

This shows that $P(X \in I) = P(X \in J) = h/(b - a)$ for every pair of subintervals Z and J of $[a, b]$ of length h.

To prove the converse, assume that X satisfies (14.17). First we note that $F(t) = 0$ if $t < a$ and $F(t) = 1$ if $t > b$, since X has all its values in $[a, b]$.

Introduce a new function g defined on the half-open interval $(0, b - a]$ by the equation

$$g(u) = P(a < X \le a + u) \quad \text{if } 0 < u \le b - a. \tag{14.18}$$

Using additivity and property (14.17) we find

$$\begin{aligned} g(u + v) &= P(a < X \le a + u + v) \\ &= P(a < X \le a + u) + P(a + u < X \le a + u + v) \\ &= g(u) + P(a < X \le a + v) = g(u) + g(v), \end{aligned}$$

provided that $0 < u + v \le b - a$. That is, g satisfies the functional equation

$$g(u + v) = g(u) + g(v)$$

for all u and v such that $u > 0$, $v > 0$, $u + v \le b - a$. This is known as ***Cauchy's functional equation.*** In a moment we shall prove that every nonnegative solution of Cauchy's functional equation is given by

$$g(u) = \frac{u}{b - a}\, g(b - a) \qquad \text{for } 0 < u \le b - a .$$

Using this in Equation (14.18) we find that for $0 < u \le b - a$ we have

$$P(a < X \le a + u) = \frac{u}{b - a}\, P(a < X \le b) = \frac{u}{b - a}$$

since $P(a < X \le b) = 1$. In other words,

$$F(a + u) - F(a) = \frac{u}{b - a} \qquad \text{if } 0 < u \le b - a .$$

We put $t = a + u$ and rewrite this as

$$F(t) - F(a) = \frac{t - a}{b - a} \qquad \text{if } a < t \le b .$$

But $F(a) = 0$ since F is continuous from the right. Hence

$$F(t) = \frac{t - a}{b - a} \qquad \text{if } a \le t \le b ,$$

which proves that F is uniform on $[a, b]$.

THEOREM 14.8. SOLUTION OF CAUCHY'S FUNCTIONAL EQUATION. *Let g be a real-valued function defined on a half-open interval $(0, c]$ and satisfying the following two properties:*

(a) *$g(u + v) = g(u) + g(v)$ whenever u, v, and $u + v$ are in $(0, c]$,*

and

(b) *g is nonnegative on $(0, c]$.*

Then g is given by the formula

$$g(u) = \frac{u}{c}\, g(c) \qquad \text{for } 0 < u \le c .$$

Proof. By introducing a change of scale we can reduce the proof to the special case in which $c = 1$. In fact, let

$$G(x) = g(cx) \qquad \text{for } 0 < x \le 1 .$$

Then G is nonnegative and satisfies the Cauchy functional equation

$$G(x + y) = G(x) + G(y)$$

whenever x, y, and $x + y$ are in $(0, 1]$. If we prove that

$$G(x) = xG(1) \quad \text{for } 0 < x \leq 1 \tag{14.19}$$

it follows that $g(cx) = xg(c)$, or that $g(u) = (u/c)g(c)$ for $0 < u \leq c$.

If x is in $(0, 1]$ then $x/2$ is also in $(0, 1]$ and we have

$$G(x) = G\left(\frac{x}{2}\right) + G\left(\frac{x}{2}\right) = 2G\left(\frac{x}{2}\right).$$

By induction, for each x in $(0, 1]$ we have

$$G(x) = nG\left(\frac{x}{n}\right) \quad \text{for} \quad n = 1, 2, 3, \ldots. \tag{14.20}$$

Similarly, if y and my are in $(0, 1]$ we have

$$G(my) = mG(y) \quad \text{for } m = 1, 2, 3, \ldots.$$

Taking $y = x/n$ and using (14.20) we obtain

$$G\left(\frac{m}{n}x\right) = \frac{m}{n}G(x)$$

if x and mx/n are in $(0, 1]$. In other words, we have

$$G(rx) = rG(x) \tag{14.21}$$

for every positive rational number r such that x and rx are in $(0, 1]$.

Now take any x in the open interval $(0, 1)$ and let $\{r_n\}$ and $\{R_n\}$ be two sequences of rational numbers in $(0, 1]$ such that

$$r_n < x < R_n \quad \text{and such that} \quad \lim_{n\to\infty} r_n = \lim_{n\to\infty} R_n = x.$$

Cauchy's functional equation and the nonnegative property of G show that $G(x + y) \geq G(x)$ so G is monotonic increasing in $(0, 1]$. Therefore

$$G(r_n) \leq G(x) \leq G(R_n).$$

Using (14.21) we rewrite this as

$$r_nG(1) \leq G(x) \leq R_nG(1).$$

Letting $n \to \infty$ we find $xG(1) \leq G(x) \leq xG(1)$, so $G(x) = xG(1)$, which proves (14.19).

Note: Uniform distributions are often used in experiments whose outcomes are points selected at random from an interval $[a, b]$, or in experiments involving an interval $[a, b]$ as a target, where aiming is impossible. The terms "at random" and "aiming is impossible" are usually interpreted to mean that if I is any subinterval of $[a, b]$ then the probability $P(X \in I)$ depends only on the length of I and not on its location in $[a, b]$. Theorem 14.7 shows that uniform distributions are the only distributions with this property.

We turn now to the probability questions asked at the beginning of this chapter.

EXAMPLE. A line segment is broken into two pieces, with the point of subdivision chosen at random. Let X denote the random variable which measures the ratio of the length of the left-hand piece to that of the right-hand piece. Determine the probability distribution function F_X.

Solution. Use the interval $[0, 1]$ to represent the line segment and let the point of subdivision be described by the random variable $Y(\omega) = \omega$ for each ω in $(0, 1)$. Since the point of subdivision is chosen at random we assume that Y has a uniform distribution function F_Y over $[0, 1]$. Hence

$$F_Y(t) = t \qquad \text{for } 0 \le t \le 1.$$

If the segment is broken at ω, then $\omega/(1 - \omega)$ is the ratio of the length of the left-hand piece to that of the right-hand piece. Therefore $X(\omega) = \omega/(1 - \omega)$.

If $t < 0$ we have $F_X(t) = 0$ since the ratio $X(\omega)$ cannot be negative. If $t \ge 0$, the inequality $X(\omega) \le t$ is equivalent to $\omega/(1 - \omega) \le t$, which is equivalent to $\omega \le t/(1 + t)$. Therefore

$$F_X(t) = P(X \le t) = P\left(Y \le \frac{t}{1+t}\right) = F_Y\left(\frac{t}{1+t}\right) = \frac{t}{1+t}$$

since $0 \le t/(1 + t) < 1$.

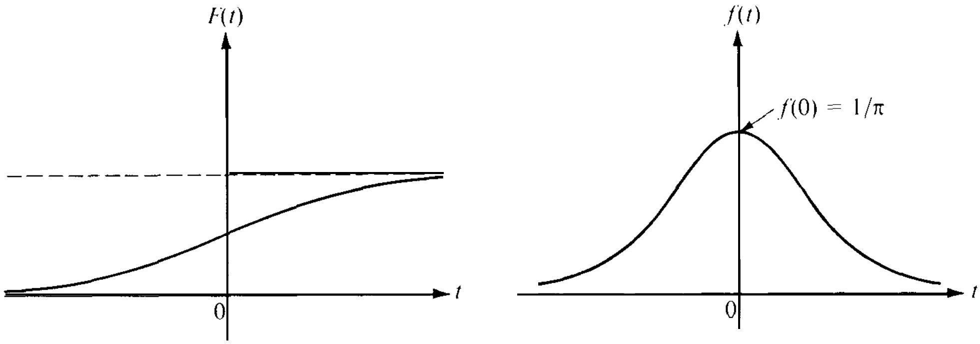
Now we can calculate various probabilities. For example, the probability that the two pieces have equal length is $P(X = 1) = 0$. In fact, since F_X is a continuous distribution, the probability that X takes any particular value is zero.

The probability that the left-hand segment is at least twice as long as the right-hand segment is $P(X \ge 2) = 1 - P(X < 2) = 1 - \frac{2}{3} = \frac{1}{3}$. Similarly, the probability that the right-hand segment is at least twice as long as the left-hand segment is $P(X \le \frac{1}{2}) = \frac{1}{3}$. The probability that the longer segment is at least twice as long as the shorter segment is $P(X \ge 2) + P(X \le \frac{1}{2}) = \frac{2}{3}$.

14.11 Cauchy's distribution

A random variable X is said to have a Cauchy distribution F if

$$F(t) = \frac{1}{2} + \frac{1}{\pi}\arctan t$$



(a) The distribution function F. (b) The density function f.

FIGURE 14.7 Cauchy's distribution function and the corresponding density function.

for all real t. This function has a continuous derivative everywhere; a continuous density function f is given by the formula

$$f(t) = \frac{1}{\pi(1 + t^2)}.$$

The graphs of F and f are shown in Figures 14.7(a) and (b), respectively.

The following experiment leads to a Cauchy distribution. A pointer pivoted at the point (-1, 0) on the x-axis is spun and allowed to come to rest. An outcome of the experiment is θ, the angle of inclination from the x-axis made by a line drawn through the pointer; θ is measured so that $-\frac{1}{2}\pi < \theta \le \frac{1}{2}\pi$. Let X be the random variable defined by $X(\theta) = \theta$, and let Y be the random variable which measures the y-intercept of the line through the pointer. If θ is the angle described above, then

$$Y(\theta) = \tan\theta.$$

We shall prove that Y has a Cauchy distribution F_Y if X has a uniform distribution over $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$.

If $a < t$ let $\alpha = \arctan a$ and let $\theta = \arctan t$. Then we have

$$F_Y(t) - F_Y(a) = P(a < Y \le t) = P(\alpha < X \le \theta) = \int_\alpha^\theta f_X(u)\, du = \frac{\theta - \alpha}{\pi}.$$

Since $\alpha \to -\frac{1}{2}\pi$ as $a \to -\infty$ we find

$$F_Y(t) = \frac{\theta + \frac{1}{2}\pi}{\pi} = \frac{1}{\pi}\arctan t + \frac{1}{2}.$$

This shows that Y has a Cauchy distribution, as asserted.

14.12 Exercises

1. A random variable X has a continuous distribution function F, where

$$F(t) = \begin{cases} 0 & \text{if } t \le 0, \\ ct & \text{if } 0 \le t \le 1, \\ 1 & \text{if } t > 1. \end{cases}$$

(a) Determine the constant c and describe the density function f.
(b) Compute the probabilities $P(X = \frac{1}{3})$, $P(X < \frac{1}{3})$, $P(|X| < \frac{1}{3})$.

2. Let $f(t) = c\,|\sin t|$ for $|t| < \pi/2$ and let $f(t) = 0$ otherwise. Determine the value of the constant c so that f will be the density of a continuous distribution function F. Also, describe F and sketch its graph.
3. Solve Exercise 2 if $f(t) = c(4t - 2t^2)$ for $0 \le t \le 2$, and $f(t) = 0$ otherwise.
4. The time in minutes that a person has to wait for a bus is known to be a random variable with density function f given by the following formulas :

$$f(t) = \tfrac{1}{2} \quad \text{for } 0 < t < 1, \qquad f(t) = a \quad \text{for } 2 < t < 4, \qquad f(t) = 0 \quad \text{otherwise.}$$

Calculate the probability that the time a person has to wait is (a) more than one minute; (b) more than two minutes; (c) more than three minutes.

5. A random variable X has a continuous distribution function F and a probability density f. The density has the following properties : $f(t) = 0$ if $t < \frac{1}{4}$, $f(\frac{1}{4}) = 1$, $f(t)$ is linear if $\frac{1}{4} \le t \le \frac{1}{2}$, $f(1 - t) = f(t)$ for all t.
(a) Make a sketch of the graph of f.
(b) Give a set of formulas for determining F and sketch its graph.
(c) Compute the following probabilities: $P(X < 1)$, $P(X < \frac{3}{4})$, $P(X < \frac{1}{2})$, $P(X \le \frac{1}{4})$, $P(\frac{1}{2} < x < \frac{5}{8})$.

6. A random variable X has a uniform distribution over $[-3, 3]$.
(a) Compute $P(X = 2)$, $P(X < 2)$, $P(|X| < 2)$, $P(|X - 2| < 2)$.
(b) Find a t for which $P(X > t) = \frac{1}{3}$.

7. The Lethe Subway Company schedules a northbound train every 30 minutes at a certain station. A man enters the station at a random time. Let the random variable X count the number of minutes he has to wait for the next train. Assume X has a uniform distribution over the interval [0, 30]. (This is how we interpret the statement that he enters the station at "random time.")
(a) For each k = 5, 10, 15, 20, 25, 30, compute the probability that he has to wait at least k minutes for the next train.
(b) A competitor, the Styx Subway Company, is allowed to schedule a northbound train every 30 minutes at the same station, but at least 5 minutes must elapse between the arrivals of competitive trains. Assume the passengers come into the station at random times and always board the first train that arrives. Show that the Styx Company can arrange its schedule so that it receives five times as many passengers as its competitor.

8. Let X be a random variable with a uniform distribution F_X over the interval [0, 1]. Let $Y = aX + b$, where $a > 0$. Determine the distribution function F_Y and sketch its graph.
9. A roulette wheel carries the integers from 0 to 36, distributed among 37 arcs of equal length. The wheel is spun and allowed to come to rest, and the point on the circumference next to a fixed pointer is recorded. Consider this point as a random variable X with a uniform distribution. Calculate the probability that X lies in an arc containing (a) the integer 0; (b) an integer n in the interval $11 \le n \le 20$; (c) an odd integer.

10. A random variable is said to have a Cauchy distribution with parameters a and b, where $a > 0$, if its density function is given by

$$f(t) = \frac{1}{\pi} \frac{a}{a^2 + (t - b)^2}.$$

Verify that the integral of f from $-$ co to + co is 1, and determine the distribution function F.

11. Let $f_1(t) = 1$ for $0 < t < 1$, and let $f_1(t) = 0$ otherwise. Define a sequence of functions $\{f_n\}$ by the recursion formula

$$f_{n+1}(x) = \int_{-\infty}^{\infty} f_1(x - t) f_n(t)\, dt.$$

(a) Prove that $f_{n+1}(x) = \int_{x-1}^{x} f_n(t)\, dt$.

(b) Make a sketch showing the graphs of f_1, f_2, and f_3.

12. Refer to Exercise 11. Prove that each functionf,, is a probability density.

14.13 Exponential distributions

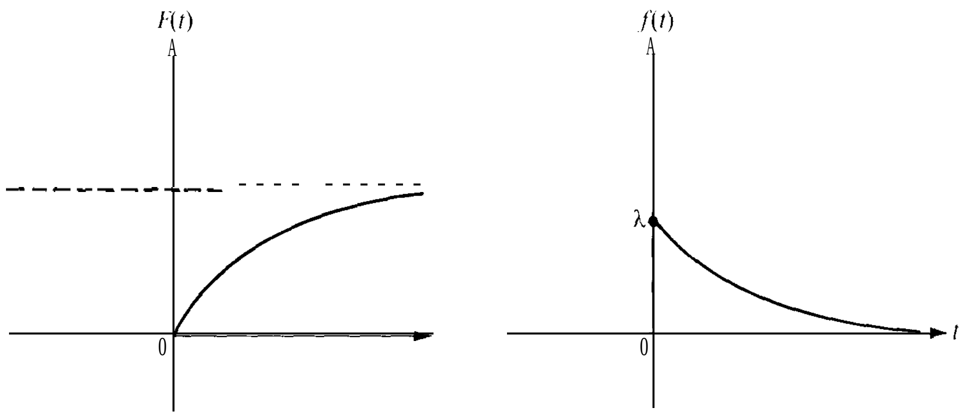
Let λ be a positive constant. A one-dimensional random variable X is said to have an exponential distribution F with parameter λ if

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

A corresponding density functionf is given by the formulas

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0, \\ 0 & \text{for } t < 0. \end{cases}$$

The graphs of F and f are like those shown in Figure 14.8.



(a) The distribution function F. (b) The density functionf.

FIGURE 14.8 An exponential distribution and the corresponding density function.

Exponential distributions have a characteristic property which suggests their use in certain problems involving radioactive decay, traffic accidents, and failure of electronic equipment such as vacuum tubes. This property is analogous to that which characterizes uniform distributions and can be described as follows.

Let X denote the observed waiting time until a piece of equipment fails, and let F be the distribution function of X. We assume that $F(t) = 0$ for $t \leq 0$, and for the moment we put no further restrictions on F. If $t > 0$, then $X \leq t$ is the event "failure occurs in the interval $[0, t]$." Hence $X > t$ is the complementary event, "no failure occurs in the interval $[0, t]$."

Suppose that no failure occurs in the interval $[0, t]$. What is the probability of continued survival in the interval $[t, t + s]$? This is a question in conditional probabilities. We wish to determine $P(X > t + s \mid X > t)$, the conditional probability that there is no failure in the interval $[0, t + s]$, given that there is no failure in the interval $[0, t]$.

From the definition of conditional probability we have

$$(14.22) \qquad P(X > t + s \mid X > t) = \frac{P[(X > t + s) \cap (X > t)]}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)} .$$

Suppose now that F is an exponential distribution with parameter $\lambda > 0$. Then $F(t) = 1 - e^{-\lambda t}$ for $t > 0$, and $P(X > t) = 1 - P(X \leq t) = e^{-\lambda t}$. Hence Equation (14.22) becomes

$$P(X > t + s \mid X > t) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s).$$

In other words, if the piece of equipment survives in the interval $[0, t]$, then the probability of continued survival in the interval $[t, t + s]$ is equal to the probability of survival in the interval $[0, s]$ having the same length. That is, the probability of survival depends only on the length of the time interval and not on the age of the equipment. Expressed in terms of the distribution function F, this property states that

$$(14.23) \qquad \frac{1 - F(t + s)}{1 - F(t)} = 1 - F(s) \qquad \text{for all } t > 0 \text{ and } s > 0.$$

The next theorem shows that exponential distributions are the only probability distributions with this property.

THEOREM 14.9. *Let F be a probability distribution function satisfying the functional equation (14.23), where $F(t) < 1$ for $t > 0$. Then there is a positive constant $\lambda > 0$ such that*

$$F(t) = 1 - e^{-\lambda t} \qquad \textit{for all } t > 0.$$

Proof. Let $g(t) = -\log [1 - F(t)]$ for $t > 0$. Then $1 - F(t) = e^{-g(t)}$, so to prove the theorem it suffices to prove that $g(t) = \lambda t$ for some $\lambda > 0$.

Now g is nonnegative and satisfies Cauchy's functional equation,

$$g(t + s) = g(t) + g(s)$$

for all $t > 0$ and $s > 0$. Therefore, applying Theorem 14.8 with $c = 1$, we deduce that $g(t) = tg(1)$ for $0 < t \leq 1$. Let $\lambda = g(1)$. Then $\lambda = -\log [1 - F(1)] > 0$, and hence $g(t) = \lambda t$ for $0 < t \leq 1$.

To prove that $g(t) = \lambda t$ for all $t > 0$, let $G(t) = g(t) - \lambda t$. The function G also satisfies Cauchy's functional equation, Moreover, G is periodic with period 1 because $G(t + 1) = G(t) + G(1)$ and $G(1) = 0$. Since G is identically 0 in (0, 1] the periodicity shows that $G(t) = 0$ for all $t > 0$. In other words, $g(t) = \lambda t$ for all $t > 0$, which completes the proof.

EXAMPLE 1. Let X be a random variable which measures the lifetime (in hours) of a certain type of vacuum tube. Assume X has an exponential distribution with parameter $\lambda = 0.001$. The manufacturer wishes to guarantee these tubes for T hours. Determine T so that $P(X > T) = 0.95$.

Solution. The distribution function is given by $F(t) = 1 - e^{-\lambda t}$ for $t > 0$, where $\lambda = 0.001$. Since $P(X > T) = 1 - F(T) = e^{-\lambda T}$, we choose T to make $e^{-\lambda T} = 0.95$. Hence $T = -(\log 0.95)/\lambda = -1000 \log 0.95 = 51.25+$.

EXAMPLE 2. Consider the random variable of Example 1, but with an unspecified value of λ. The following argument suggests a reasonable procedure for determining λ. Start with an initial number of vacuum tubes at time $t = 0$, and let $g(t)$ denote the number of tubes still functioning t hours later. The ratio $[g(0) - g(t)]/g(0)$ is the fraction of the original number that has failed in time t. Since the probability that a particular tube fails in time t is $1 - e^{-\lambda t}$, it seems reasonable to expect that the equation

$$\frac{g(0) - g(t)}{g(0)} = 1 - e^{-\lambda t} \tag{14.24}$$

should be a good approximation to reality. If we assume (14.24) we obtain

$$g(t) = g(0)e^{-\lambda t}.$$

In other words, under the hypothesis (14.24), the number $g(t)$ obeys an exponential decay law with decay constant λ. The decay constant can be computed in terms of the half-life. If t_1 is the half-life then $\frac{1}{2} = g(t_1)/g(0) = e^{-\lambda t_1}$, so $\lambda = (\log 2)/t_1$. For example, if the half-life of a large sample of tubes is known to be 693 hours, we obtain $\lambda = (\log 2)/693 = 0.001$.

14.14 Normal distributions

Let m and σ be fixed real numbers, with $\sigma > 0$. A random variable X is said to have a *normal distribution* with mean m and variance σ^2 if the density function f is given by the formula

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(t-m)/\sigma]^2/2}$$

for all real t. The corresponding distribution function F is, of course, the integral

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{t} e^{-[(u-m)/\sigma]^2/2}\, du \cdot$$

Table 14.1 Values of the standard normal distribution function

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2}\, du .$$

t	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	0.5596	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	0.8577	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
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

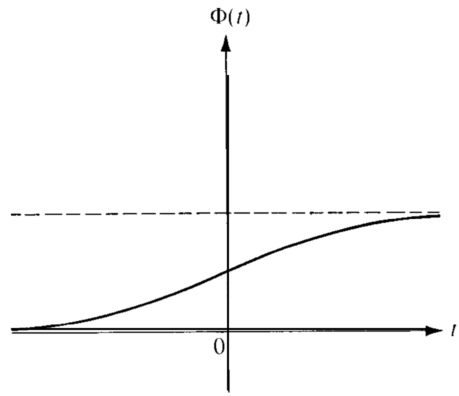


FIGURE 14.9 The standard normal distribution function: $m = 0$, $\sigma = 1$.

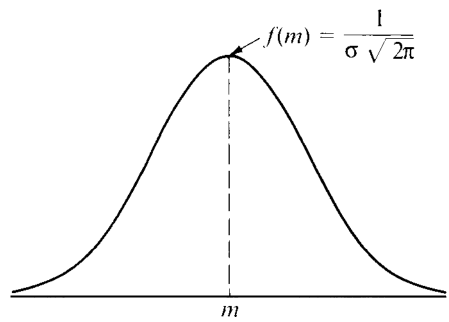


FIGURE 14.10 The density function of a normal distribution with mean m and variance σ^2.

It is clear that this function F is monotonic increasing, continuous everywhere, and tends to 0 as $t \to -\infty$. Also, it can be shown that $F(t) \to 1$ as $t \to +\infty$. (See Exercise 7 of Section 14.16.)

The special case $m = 0$, $\sigma = 1$ is called the *standard* normal distribution. In this case the function F is usually denoted by the letter Φ. Thus,

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2}\, du .$$

The general case can be reduced to the standard case by introducing the change of variable $v = (u - m)/\sigma$ in the integral for F. This leads to the formula

$$F(t) = \Phi\left(\frac{t-m}{\sigma}\right).$$

A four-place table of values of $\Phi(t)$ for values of t spaced at intervals of length 0.01 is given in Table 14.1 for $t = 0.00$ to $t = 3.69$. The graph of Φ is shown in Figure 14.9. The graph of the density f is a famous "bell-shaped" curve, shown in Figure 14.10. The top of the bell is directly above the mean m. For large values of σ the curve tends to flatten out; for small σ it has a sharp peak, as in Figure 14.10.

Normal distributions are among the most important of all continuous distributions. Many random variables that occur in nature behave as though their distribution functions are normal or approximately normal. Examples include the measurement of the height of people in a large population, certain measurements on large populations of living organisms encountered in biology, and the errors of observation encountered when making large numbers of measurements. In physics, Maxwell's law of velocities implies that the distribution function of the velocity in any given direction of a molecule of mass M in a gas at absolute temperature T is normal with mean 0 and variance $M/(kT)$, where k is a constant (Boltzmann's constant).

The normal distribution is also of theoretical importance because it can be used to approximate the distributions of many random phenomena. One example is the binomial

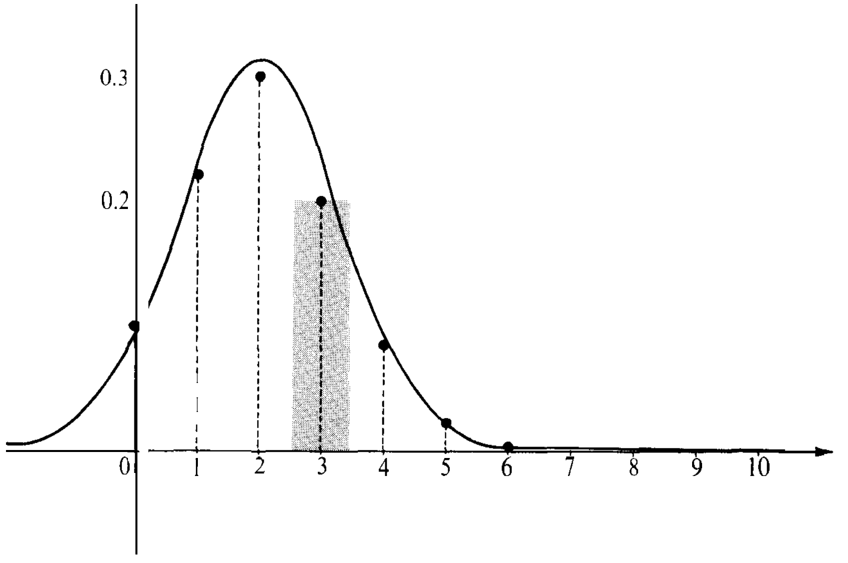


FIGURE 14.11 The density function of a normal distribution considered as an approximation to the probability mass function of a binomial distribution.

distribution with parameters n and p. If X is a random variable having a binomial distribution with parameters n andp, the probability $P(a \le X \le b)$ is given by the sum

$$\sum_{k=a}^{b} \binom{n}{k} p^k q^{n-k},$$

where $q = 1 - p$. For a large n, laborious computations are needed to evaluate this sum. In practice these computations are avoided by use of the approximate formula

$$\sum_{k=a}^{b} \binom{n}{k} p^k q^{n-k} \sim \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right), \tag{14.25}$$

where the symbol $\sim$ means that the two sides of (14.25) are asymptotically equal; that is, the ratio of the left member to the right member approaches the limit 1 as $n \to \infty$. The limit relation expressed in (14.25) is a special case of the so-called *central limit theorem* of the calculus of probabilities. This theorem (discussed in more detail in Section 14.30) explains the theoretical importance of normal distributions.

Figure 14.11 illustrates approximate formula (14.25) and shows that it can be accurate even for a relatively small value of n. The dotted lines are the ordinates of the probability mass function p of a binomial distribution with parameters $n = 10$ and $p = \frac{1}{5}$. These ordinates were computed from the formula

$$p(t) = P(X = t) = \binom{10}{t}\left(\frac{1}{5}\right)^t \left(\frac{4}{5}\right)^{10-t} \quad \text{for} \quad t = 0, 1, 2, \ldots, 10.$$

The ordinates for $t = 7, 8, 9$, and 10 are not shown because their numerical values are too near zero. For example, $p(10) = (\frac{1}{5})^{10} = 2^{10}/10^{10} = 0.0000001024$. The smooth curve is the graph of the density function f of a normal distribution (with mean $m = np = 2$ and variance $\sigma^2 = npq = 1.6$). To compute the probability $P(a \le t \le b)$ from the mass function p we add the function values $p(t)$ at the mass points in the interval $a \le t \le b$. Each value $p(t)$ may be interpreted as the area of a rectangle of height $p(t)$ located over an interval of unit length centered about the mass point t. (An example, centered about $t = 3$, is shown in Figure 14.11.) The approximate formula in (14.25) is the result of replacing the areas of these rectangles by the area of the ordinate set off over the interval $[a - \frac{1}{2}, b + \frac{1}{2}]$.

14.15 Remarks on more general distributions

In the foregoing sections we have discussed examples of discrete and continuous distributions. The values of a discrete distribution are computed by adding the values of the corresponding probability mass function. The values of a continuous distribution with a density are computed by integrating the density function. There are, of course, distributions that are neither discrete nor continuous. Among these are the so-called "mixed" types in which the mass distribution is partly discrete and partly continuous. (An example is shown in Figure 14.3.)

A distribution function F is called *mixed* if it can be expressed as a linear combination of the form

$$F(t) = c_1F_1(t) + c_2F_2(t), \tag{14.26}$$

where F_1 is discrete and F_2 is continuous. The constants c_1 and c_2 must satisfy the relations

$$0 < c_1 < 1, \qquad 0 < c_2 < 1, \qquad c_1 + c_2 = 1.$$

Properties of mixed distributions may be found by studying those that are discrete or continuous and then appealing to the linearity expressed in Equation (14.26).

A general kind of integral, known as the *Riemann-Stieltjes integral,* makes possible a simultaneous treatment of the discrete, continuous, and mixed cases.7 Although this integral unifies the theoretical discussion of distribution functions, in any specific problem the computation of probabilities must be reduced to ordinary summation and integration. In this introductory account we shall not attempt to describe the Riemann-Stieltjes integral. Consequently, most of the topics we discuss come in pairs, one for the discrete case and one for the continuous case. However, we shall only give complete details for one case, leaving the untreated case for the reader to work out.

Even the Riemann-Stieltjes integral is inadequate for treating the *most general* distribution functions. But a more powerful concept, called the Lebesgue-Stieltjes integral,‡ does give a satisfactory treatment of all cases. The advanced theory of probability cannot be undertaken without a knowledge of the Lebesgue-Stieltjes integral.

† A discussion of the Riemann-Stieltjes integral may be found in Chapter 9 of the author's Mathematical *Analysis,* Addison-Wesley Publishing Company, Reading, Mass. 1957.

‡ See any book on measure theory.