

As emphasized at the beginning of Chap. 2, the two important types of random variables are discrete and continuous. In this chapter, we study the second general type of random variable that arises in many applied problems. Sections 3.1 and 3.2 present the basic definitions and properties of continuous random variables, their probability distributions, and their various expected values. The normal distribution, arguably the most important and useful model in all of probability and statistics, is introduced in Sect. 3.3. Sections 3.4 and 3.5 discuss some other continuous distributions that are often used in applied work. In Sect. 3.6, we introduce a method for assessing whether given sample data is consistent with a specified distribution. Section 3.7 presents methods for obtaining the distribution of a rv Y from the distribution of X when the two are related by some equation $Y = g(X)$. The last section of this chapter is dedicated to the simulation of continuous rvs.

3.1 Probability Density Functions and Cumulative Distribution Functions

A discrete random variable (rv) is one whose possible values either constitute a finite set or else can be listed in an infinite sequence (a list in which there is a first element, a second element, etc.). A random variable whose set of possible values is an entire interval of numbers is not discrete.

Recall from the beginning of Chap. 2 that a random variable X is **continuous** if (1) its possible values comprise either a single interval on the number line (for some $A < B$, any number x between A and B is a possible value) or a union of disjoint intervals, and (2) $P(X = c) = 0$ for any number c that is a possible value of X .

Example 3.1 If in the study of the ecology of a lake, we make depth measurements at randomly chosen locations, then $X =$ the depth at such a location is a continuous rv. Here A is the minimum depth in the region being sampled, and B is the maximum depth. ■

Example 3.2 If a chemical compound is randomly selected and its pH X is determined, then X is a continuous rv because any pH value between 0 and 14 is possible. If more is known about the compound selected for analysis, then the set of possible values might be a subinterval of $[0, 14]$, such as $5.5 \leq x \leq 6.5$, but X would still be continuous. ■

Example 3.3 Let X represent the amount of time a randomly selected customer spends waiting for a haircut. Your first thought might be that X is a continuous random variable, since a measurement is required to determine its value. However, there are customers lucky enough to have no wait whatsoever before climbing into the barber or stylist's chair. So it must be the case that $P(X = 0) > 0$. Conditional on no chairs being empty, however, the waiting time will be continuous since X could then assume any value between some minimum possible time A and a maximum possible time B . This random variable is neither purely discrete nor purely continuous but instead is a mixture of the two types. ■

One might argue that although in principle variables such as height, weight, and temperature are continuous, in practice the limitations of our measuring instruments restrict us to a discrete (though sometimes very finely subdivided) world. However, continuous models often approximate real-world situations very well, and continuous mathematics (the calculus) is frequently easier to work with than the mathematics of discrete variables and distributions.

3.1.1 Probability Distributions for Continuous Variables

Suppose the variable X of interest is the depth of a lake at a randomly chosen point on the surface. Let $M =$ the maximum depth (in meters), so that any number in the interval $[0, M]$ is a possible value of X . If we “discretize” X by measuring depth to the nearest meter, then possible values are nonnegative integers less than or equal to M . The resulting discrete distribution of depth can be pictured using a probability histogram. If we draw the histogram so that the area of the rectangle above any possible integer k is the proportion of the lake whose depth is (to the nearest meter) k , then the total area of all rectangles is 1. A possible histogram appears in Fig. 3.1a.

If depth is measured much more precisely and the same measurement axis as in Fig. 3.1a is used, each rectangle in the resulting probability histogram is much narrower, although the total area of all rectangles is still 1. A possible histogram is pictured in Fig. 3.1b; it has a much smoother appearance than the histogram in Fig. 3.1a. If we continue in this way to measure depth more and more finely, the resulting sequence of histograms approaches a smooth curve, as pictured in Fig. 3.1c. Because for each histogram the total area of all rectangles equals 1, the total area under the smooth curve is also 1. The probability that the depth at a randomly chosen point is between a and b is just the area under the smooth curve between a and b . It is exactly a smooth curve of the type pictured in Fig. 3.1c that specifies a continuous probability distribution.

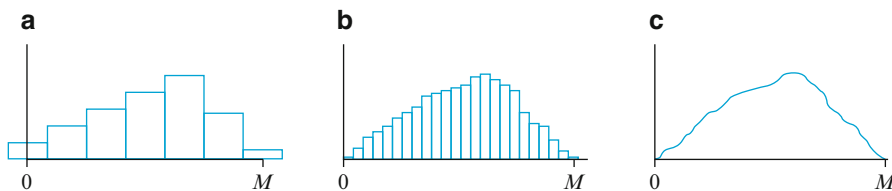


Fig. 3.1 (a) Probability histogram of depth measured to the nearest meter; (b) probability histogram of depth measured to the nearest centimeter; (c) a limit of a sequence of discrete histograms

DEFINITION

Let X be a continuous rv. Then a **probability distribution** or **probability density function** (pdf) of X is a function $f(x)$ such that for any two numbers a and b with $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

That is, the probability that X takes on a value in the interval $[a, b]$ is the area above this interval and under the graph of the density function, as illustrated in Fig. 3.2. The graph of $f(x)$ is often referred to as the *density curve*.

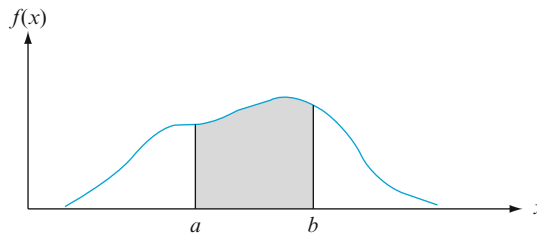


Fig. 3.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x) dx = [\text{area under the entire graph of } f(x)] = 1$

Example 3.4 The direction of an imperfection with respect to a reference line on a circular object such as a tire, brake rotor, or flywheel is often subject to uncertainty. Consider the reference line connecting the valve stem on a tire to the center point, and let X be the angle measured clockwise to the location of an imperfection. One possible pdf for X is

$$f(x) = \begin{cases} \frac{1}{360} & 0 \leq x < 360 \\ 0 & \text{otherwise} \end{cases}$$

The pdf is graphed in Fig. 3.3. Clearly $f(x) \geq 0$. The area under the density curve is just the area of a rectangle: (height)(base) = $\left(\frac{1}{360}\right)(360) = 1$. The probability that the angle is between 90° and 180° is

$$P(90 \leq X \leq 180) = \int_{90}^{180} \frac{1}{360} dx = \frac{x}{360} \Big|_{x=90}^{x=180} = \frac{1}{4} = .25$$

The probability that the angle of occurrence is within 90° of the reference line is

$$P(0 \leq X \leq 90) + P(270 \leq X < 360) = .25 + .25 = .50$$

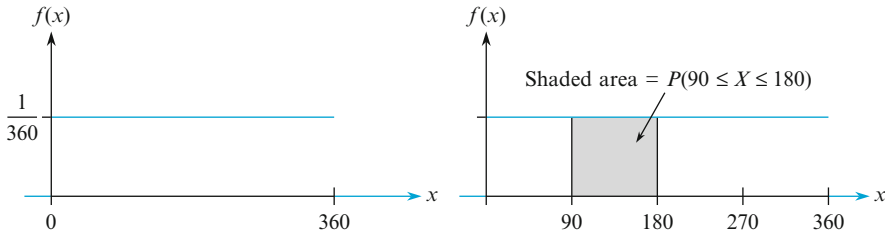


Fig. 3.3 The pdf and probability for Example 3.4 ■

Because the pdf in Fig. 3.3 is completely “level” (i.e., has a uniform height) on the interval $[0, 360]$, X is said to have a *uniform* distribution.

DEFINITION

A continuous rv X is said to have a **uniform distribution** on the interval $[A, B]$ if the pdf of X is

$$f(x; A, B) = \begin{cases} \frac{1}{B - A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

The statement that X has a uniform distribution on $[A, B]$ will be denoted $X \sim \text{Unif}[A, B]$.

The graph of any uniform pdf looks like the graph in Fig. 3.3 except that the interval of positive density is $[A, B]$ rather than $[0, 360]$.

In the discrete case, a probability mass function (pmf) tells us how little “blobs” of probability mass of various magnitudes are distributed along the measurement axis. In the continuous case, probability density is “smeared” in a continuous fashion along the interval of possible values. When density is smeared evenly over the interval, a uniform pdf, as in Fig. 3.3, results.

When X is a discrete random variable, each possible value is assigned positive probability. This is not true of a continuous random variable, because the area under a density curve that lies above any single value is zero:

$$P(X = c) = P(c \leq X \leq c) = \int_c^c f(x) dx = 0$$

The fact that $P(X = c) = 0$ when X is continuous has an important practical consequence: The probability that X lies in some interval between a and b does not depend on whether the lower limit a or the upper limit b is included in the probability calculation:

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) \quad (3.1)$$

In contrast, if X were discrete and both a and b were possible values of X (e.g., $X \sim \text{Bin}(20, .3)$ and $a = 5, b = 10$), then all four of the probabilities in Eq. (3.1) would be different. This also means that whether we include the endpoints of the range of values for a continuous rv X is somewhat arbitrary; for example, the pdf in Example 3.4 could be defined to be positive on $(0, 360)$ or $[0, 360]$ rather than $[0, 360)$, and the same applies for a uniform distribution on $[A, B]$ in general.

The zero probability condition has a physical analog. Consider a solid circular rod (with cross-sectional area of 1 in^2 for simplicity). Place the rod alongside a measurement axis and suppose that the density of the rod at any point x is given by the value $f(x)$ of a density function. Then if the rod is sliced

at points a and b and this segment is removed, the amount of mass removed is $\int_a^b f(x)dx$; however, if the rod is sliced just at the point c , no mass is removed. Mass is assigned to interval segments of the rod but not to individual points.

So, if $P(X = c) = 0$ when X is a continuous rv, then what does $f(c)$ represent? After all, if X were discrete, its pmf evaluated at $x = c$, $p(c)$, would indicate the probability that X equals c . To help understand what $f(c)$ means, consider a small window near $x = c$ —say, $[c, c + \Delta x]$. Using a rectangle to approximate the area under $f(x)$ between c and $c + \Delta x$ (the usual “Riemann approximation” idea from calculus), one obtains $\int_c^{c + \Delta x} f(x)dx \approx \Delta x \cdot f(c)$, from which

$$f(c) \approx \frac{\int_c^{c+\Delta x} f(x)dx}{\Delta x} = \frac{P(c \leq X \leq c + \Delta x)}{\Delta x}$$

This indicates that $f(c)$ is not a probability, but rather roughly the probability of an interval *divided by the length of the chosen interval*. If we associate mass with probability and remember that interval length is the one-dimensional analog of volume, then f represents their quotient, mass per volume, more commonly known as *density* (hence, the name pdf). The height of the function $f(x)$ at a particular point reflects how “dense” the values of X are near that point—taller sections of $f(x)$ contain more probability within a fixed interval length than do shorter sections.

Example 3.5 “Time headway” in traffic flow is the elapsed time between the time that one car finishes passing a fixed point and the instant that the next car begins to pass that point. Let X = the time headway for two randomly chosen consecutive cars on a freeway during a period of heavy flow. The following pdf of X is essentially the one suggested in “The Statistical Properties of Freeway Traffic” (*Transp. Res.*, 11: 221–228):

$$f(x) = \begin{cases} .15e^{-.15(x-.5)} & x \geq .5 \\ 0 & \text{otherwise} \end{cases}$$

The graph of $f(x)$ is given in Fig. 3.4; there is no density associated with headway times less than .5, and headway density decreases rapidly (exponentially fast) as x increases from .5. The fact that the graph of $f(x)$ is taller near $x = .5$ and shorter near, say, $x = 10$ indicates that time headway values are more dense near the left boundary, i.e., there is a higher proportion of time headways in the interval $[.5, 1.5]$ than in $[10, 11]$, even though these two intervals have the same length.

Clearly, $f(x) \geq 0$; to show that $\int_{-\infty}^{\infty} f(x)dx = 1$ we use the calculus result $\int_a^{\infty} e^{-kx}dx = (1/k)e^{-ka}$.

Then

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{.5} 0 dx + \int_{.5}^{\infty} .15e^{-.15(x-.5)}dx \\ &= .15e^{-.075} \int_{.5}^{\infty} e^{-.15x}dx = .15e^{.075} \cdot \frac{1}{.15}e^{-.15(.5)} = 1 \end{aligned}$$

The probability that headway time is at most 5 seconds is

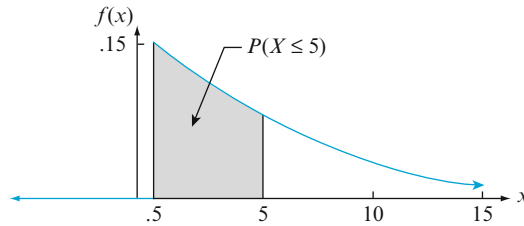


Fig. 3.4 The density curve for headway time in Example 3.5

$$\begin{aligned}
 P(X \leq 5) &= \int_{-\infty}^5 f(x) \, dx = \int_{.5}^5 .15e^{-.15(x-.5)} \, dx = .15e^{.075} \int_{.5}^5 e^{-.15x} \, dx \\
 &= .15e^{.075} \cdot \left. \frac{-1}{.15} e^{-.15x} \right|_{x=.5}^{x=5} \\
 &= e^{.075} (-e^{-.75} + e^{-.075}) = 1.078(-.472 + .928) = .491
 \end{aligned}$$

Since X is a continuous rv, .491 also equals $P(X < 5)$, the probability that headway time is (strictly) less than 5 s. The difference between these two events is $\{X = 5\}$, i.e., that headway time is exactly 5 s, which has probability zero: $P(X = 5) = \int_5^5 f(x) \, dx = 0$.

This last statement may feel uncomfortable to you: Is there really zero chance that the headway time between two cars is exactly 5 s? If time is treated as continuous, then “exactly 5 s” means $X = 5.000\dots$, with an endless repetition of 0s. That is to say, X isn’t rounded to the nearest second (or even tenth of a second); we are asking for the probability that X equals one specific number, 5.000..., out of the (uncountably) infinite collection of possible values of X . ■

Unlike discrete distributions such as the binomial, hypergeometric, and negative binomial, the distribution of any given continuous rv cannot usually be derived using simple probabilistic arguments. Instead, one must make a judicious choice of pdf based on prior knowledge and available data. Fortunately, some general pdf families have been found to fit well in a wide variety of experimental situations; several of these are discussed later in the chapter.

Just as in the discrete case, it is often helpful to think of the population of interest as consisting of X values rather than individuals or objects. The pdf is then a model for the distribution of values in this numerical population, and from this model various population characteristics (such as the mean) can be calculated.

Several of the most important concepts introduced in the study of discrete distributions also play an important role for continuous distributions. Definitions analogous to those in Chap. 2 involve replacing summation by integration.

3.1.2 The Cumulative Distribution Function

The cumulative distribution function (cdf) $F(x)$ for a discrete rv X gives, for any specified number x , the probability $P(X \leq x)$. It is obtained by summing the pmf $p(y)$ over all possible values y satisfying $y \leq x$. The cdf of a continuous rv gives the same probabilities $P(X \leq x)$ and is obtained by integrating the pdf $f(y)$ between the limits $-\infty$ and x .

DEFINITION

The **cumulative distribution function** $F(x)$ for a continuous rv X is defined for every number x by

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y)dy$$

For each x , $F(x)$ is the area under the density curve to the left of x . This is illustrated in Fig. 3.5, where $F(x)$ increases smoothly as x increases.

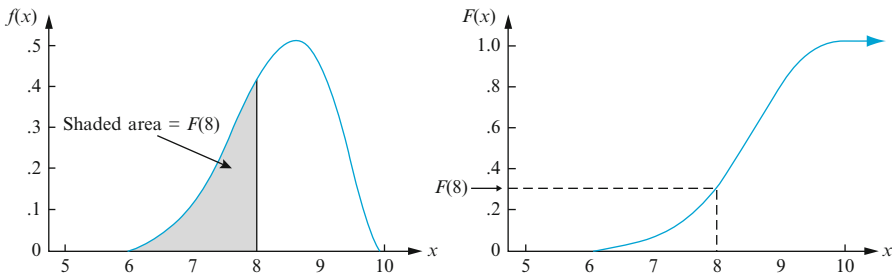


Fig. 3.5 A pdf and associated cdf

Example 3.6 Let X , the thickness of a membrane, have a uniform distribution on $[A, B]$. The density function is shown in Fig. 3.6.

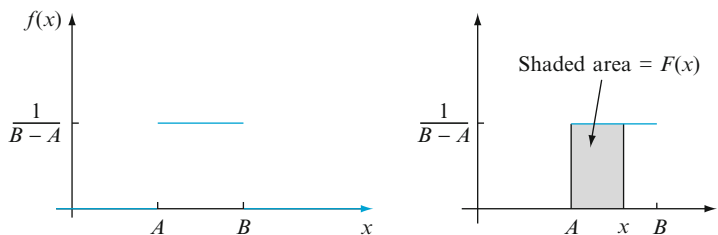
For $x < A$, $F(x) = 0$, since there is no area under the graph of the density function to the left of such an x . For $x \geq B$, $F(x) = 1$, since all the area is accumulated to the left of such an x . Finally, for $A \leq x < B$,

$$F(x) = \int_{-\infty}^x f(y)dy = \int_A^x \frac{1}{B-A} dy = \frac{1}{B-A} \cdot y \Big|_{y=A}^{y=x} = \frac{x-A}{B-A}$$

The entire cdf is

$$F(x) = \begin{cases} 0 & x < A \\ \frac{x-A}{B-A} & A \leq x < B \\ 1 & x \geq B \end{cases}$$

Fig. 3.6 The pdf for a uniform distribution



The graph of this cdf appears in Fig. 3.7.

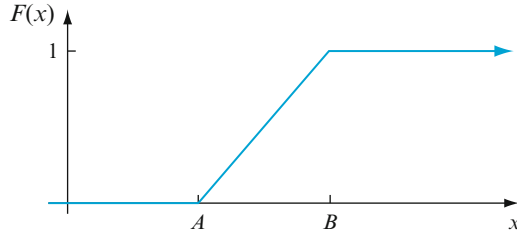


Fig. 3.7 The cdf for a uniform distribution

3.1.3 Using $F(x)$ to Compute Probabilities

The importance of the cdf here, just as for discrete rvs, is that probabilities of various intervals can be computed from a formula or table for $F(x)$.

PROPOSITION

Let X be a continuous rv with pdf $f(x)$ and cdf $F(x)$. Then for any number a ,

$$P(X > a) = 1 - F(a)$$

and for any two numbers a and b with $a < b$,

$$P(a \leq X \leq b) = F(b) - F(a)$$

Figure 3.8 illustrates the second part of this proposition; the desired probability is the shaded area under the density curve between a and b , and it equals the difference between the two shaded cumulative areas. This is different from what is appropriate for a discrete integer-valued rv (e.g., binomial or Poisson): $P(a \leq X \leq b) = F(b) - F(a - 1)$ when a and b are integers.

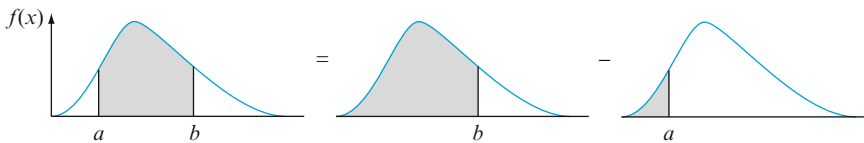


Fig. 3.8 Computing $P(a \leq X \leq b)$ from cumulative probabilities

Example 3.7 Suppose the pdf of the magnitude X of a dynamic load on a bridge (in newtons) is given by

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

For any number x between 0 and 2,

$$F(x) = \int_{-\infty}^x f(y)dy = \int_0^x \left(\frac{1}{8} + \frac{3}{8}y\right)dy = \frac{x}{8} + \frac{3x^2}{16}$$

Thus

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{8} + \frac{3x^2}{16} & 0 \leq x \leq 2 \\ 1 & 2 < x \end{cases}$$

The graphs of $f(x)$ and $F(x)$ are shown in Fig. 3.9. The probability that the load is between 1 and 1.5 N is

$$\begin{aligned} P(1 \leq X \leq 1.5) &= F(1.5) - F(1) = \left[\frac{1}{8}(1.5) + \frac{3}{16}(1.5)^2\right] - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right] \\ &= \frac{19}{64} = .297 \end{aligned}$$

The probability that the load exceeds 1 N is

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1) = 1 - \left[\frac{1}{8}(1) + \frac{3}{16}(1)^2\right] = \frac{11}{16} = .688$$

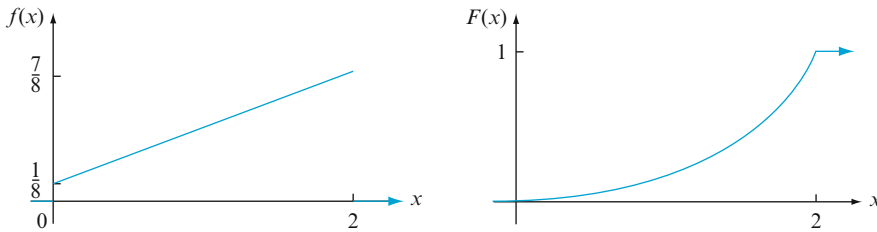


Fig. 3.9 The pdf and cdf for Example 3.7

The beauty of the cdf in the continuous case is that once it is available, any probability involving X can easily be calculated without any further integration.

3.1.4 Obtaining $f(x)$ from $F(x)$

For X discrete, the pmf is obtained from the cdf by taking the difference between two $F(x)$ values. The continuous analog of a difference is a derivative. The following result is a consequence of the Fundamental Theorem of Calculus.

PROPOSITION

If X is a continuous rv with pdf $f(x)$ and cdf $F(x)$, then at every x at which the derivative $F'(x)$ exists, $F'(x) = f(x)$.

Example 3.8 (Example 3.6 continued) When $X \sim \text{Unif}[A, B]$, $F(x)$ is differentiable except at $x = A$ and $x = B$, where the graph of $F(x)$ has sharp corners. Since $F(x) = 0$ for $x < A$ and $F(x) = 1$ for $x > B$, $F'(x) = 0 = f(x)$ for such x . For $A < x < B$,

$$F(x) = \frac{d}{dx} \left(\frac{x-A}{B-A} \right) = \frac{1}{B-A} = f(x) \quad \blacksquare$$

3.1.5 Percentiles of a Continuous Distribution

When we say that an individual's test score was at the 85th percentile of the population, we mean that 85% of all population scores were below that score and 15% were above. Similarly, the 40th percentile is the score that exceeds 40% of all scores and is exceeded by 60% of all scores.

DEFINITION

Let p be a number between 0 and 1. The **(100p)th percentile** of the distribution of a continuous rv X , denoted by η_p , is defined implicitly by the equation

$$p = F(\eta_p) = \int_{-\infty}^{\eta_p} f(y) dy \quad (3.2)$$

Assuming we can find the inverse of $F(x)$, this can also be written as

$$\eta_p = F^{-1}(p)$$

In particular, the **median** of a continuous distribution is the 50th percentile, $\eta_{.5}$ or $F^{-1}(.5)$. That is, half the area under the density curve is to the left of the median and half is to the right of the median. We will occasionally denote the median of a distribution simply as η (i.e., without the .5 subscript).

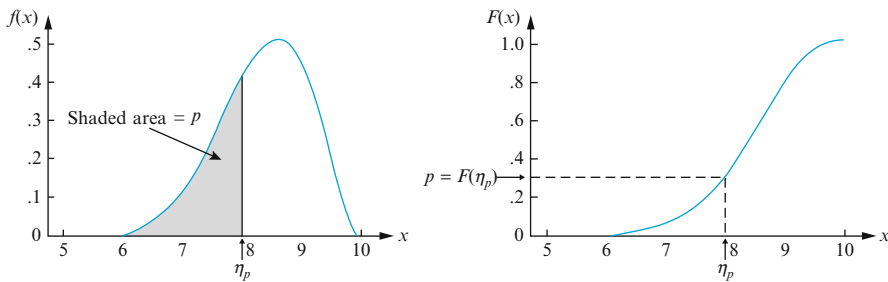


Fig. 3.10 The (100p)th percentile of a continuous distribution

According to Expression (3.2), η_p is that value on the measurement axis such that 100p% of the area under the graph of $f(x)$ lies to the left of η_p and 100(1 - p)% lies to the right. Thus $\eta_{.75}$, the 75th percentile, is such that the area under the graph of $f(x)$ to the left of $\eta_{.75}$ is .75. Figure 3.10 illustrates the definition.

Example 3.9 The distribution of the amount of gravel (in tons) sold by a construction supply company in a given week is a continuous rv X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The cdf of sales for any x between 0 and 1 is

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)dy = \frac{3}{2} \left(y - \frac{y^3}{3} \right) \Big|_{y=0}^{y=x} = \frac{3}{2} \left(x - \frac{x^3}{3} \right)$$

The graphs of both $f(x)$ and $F(x)$ appear in Fig. 3.11. The $(100p)$ th percentile of this distribution satisfies the equation

$$p = F(\eta_p) = \frac{3}{2} \left(\eta_p - \frac{\eta_p^3}{3} \right)$$

that is,

$$\eta_p^3 - 3\eta_p + 2p = 0$$

For the median, $p = .5$ and the equation to be solved is $\eta^3 - 3\eta + 1 = 0$; the solution is $\eta = .347$. If the distribution remains the same from week to week, then in the long run 50% of all weeks will result in sales of less than .347 tons and 50% in more than .347 tons.

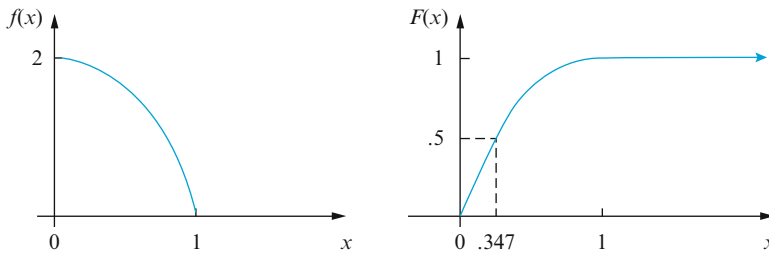


Fig. 3.11 The pdf and cdf for Example 3.9



A continuous distribution whose pdf is **symmetric**—which means that the graph of the pdf to the left of some point is a mirror image of the graph to the right of that point—has median η equal to the point of symmetry, since half the area under the curve lies to either side of this point. Figure 3.12 gives several examples. The amount of error in a measurement of a physical quantity is often assumed to have a symmetric distribution.

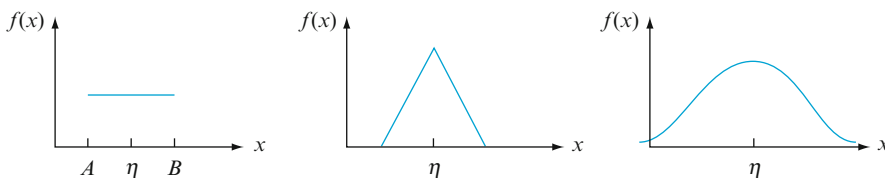


Fig. 3.12 Medians of symmetric distributions

3.1.6 Exercises: Section 3.1 (1–18)

1. The current in a certain circuit as measured by an ammeter is a continuous random variable X with the following density function:

$$f(x) = \begin{cases} .075x + .2 & 3 \leq x \leq 5 \\ 0 & \text{otherwise} \end{cases}$$

- Graph the pdf and verify that the total area under the density curve is indeed 1.
 - Calculate $P(X \leq 4)$. How does this probability compare to $P(X < 4)$?
 - Calculate $P(3.5 \leq X \leq 4.5)$ and $P(X > 4.5)$.
2. Suppose the reaction temperature X (in $^{\circ}\text{C}$) in a chemical process has a uniform distribution with $A = -5$ and $B = 5$.
- Compute $P(X < 0)$.
 - Compute $P(-2.5 < X < 2.5)$.
 - Compute $P(-2 \leq X \leq 3)$.
 - For k satisfying $-5 < k < k + 4 < 5$, compute $P(k < X < k + 4)$. Interpret this in words.
3. Suppose the error involved in making a measurement is a continuous rv X with pdf

$$f(x) = \begin{cases} .09375(4 - x^2) & -2 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

- Sketch the graph of $f(x)$.
 - Compute $P(X > 0)$.
 - Compute $P(-1 < X < 1)$.
 - Compute $P(X < -.5 \text{ or } X > .5)$.
4. Let X denote the vibratory stress (psi) on a wind turbine blade at a particular wind speed in a wind tunnel. The article “Blade Fatigue Life Assessment with Application to VAWTS” (*J. Solar Energy Engr.*, 1982: 107–111) proposes the *Rayleigh* distribution, with pdf

$$f(x; \theta) = \begin{cases} \frac{x}{\theta^2} \cdot e^{-x^2/(2\theta^2)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

as a model for X , where θ is a positive constant.

- Verify that $f(x; \theta)$ is a legitimate pdf.
 - Suppose $\theta = 100$ (a value suggested by a graph in the article). What is the probability that X is at most 200? Less than 200? At least 200?
 - What is the probability that X is between 100 and 200 (again assuming $\theta = 100$)?
 - Give an expression for the cdf of X .
5. A college professor never finishes his lecture before the end of the hour and always finishes his lectures within 2 min after the hour. Let X = the time that elapses between the end of the hour and the end of the lecture and suppose the pdf of X is

$$f(x) = \begin{cases} kx^2 & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Find the value of k and draw the corresponding density curve. [*Hint*: Total area under the graph of $f(x)$ is 1.]
 - (b) What is the probability that the lecture ends within 1 min of the end of the hour?
 - (c) What is the probability that the lecture continues beyond the hour for between 60 and 90 s?
 - (d) What is the probability that the lecture continues for at least 90 s beyond the end of the hour?
6. The actual tracking weight of a stereo cartridge that is set to track at 3 g on a particular changer can be regarded as a continuous rv X with pdf

$$f(x) = \begin{cases} k[1 - (x - 3)^2] & 2 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Sketch the graph of $f(x)$.
 - (b) Find the value of k .
 - (c) What is the probability that the actual tracking weight is greater than the prescribed weight?
 - (d) What is the probability that the actual weight is within .25 g of the prescribed weight?
 - (e) What is the probability that the actual weight differs from the prescribed weight by more than .5 g?
7. The article “Second Moment Reliability Evaluation vs. Monte Carlo Simulations for Weld Fatigue Strength” (*Quality and Reliability Engr. Intl.*, 2012: 887-896) considered the use of a uniform distribution with $A = .20$ and $B = 4.25$ for the diameter X of a certain type of weld (mm).
- (a) Determine the pdf of X and graph it.
 - (b) What is the probability that diameter exceeds 3 mm?
 - (c) What is the probability that diameter is within 1 mm of the mean diameter?
 - (d) For any value a satisfying $.20 < a < a + 1 < 4.25$, what is $P(a < X < a + 1)$?
8. Commuting to work requires getting on a bus near home and then transferring to a second bus. If the waiting time (in minutes) at each stop has a Unif[0, 5] distribution, then it can be shown that the total waiting time Y has the pdf

$$f(y) = \begin{cases} \frac{1}{25}y & 0 \leq y < 5 \\ \frac{2}{5} - \frac{1}{25}y & 5 \leq y \leq 10 \\ 0 & y < 0 \text{ or } y > 10 \end{cases}$$

- (a) Sketch the pdf of Y .
- (b) Verify that $\int_{-\infty}^{\infty} f(y)dy = 1$.
- (c) What is the probability that total waiting time is at most 3 min?
- (d) What is the probability that total waiting time is at most 8 min?
- (e) What is the probability that total waiting time is between 3 and 8 min?
- (f) What is the probability that total waiting time is either less than 2 min or more than 6 min?

9. Consider again the pdf of $X =$ time headway given in Example 3.5. What is the probability that time headway is
- At most 6 s?
 - More than 6 s? At least 6 s?
 - Between 5 and 6 s?
10. A family of pdfs that has been used to approximate the distribution of income, city population size, and size of firms is the *Pareto* family. The family has two parameters, k and θ , both > 0 , and the pdf is

$$f(x; k, \theta) = \begin{cases} \frac{k \cdot \theta^k}{x^{k+1}} & x \geq \theta \\ 0 & x < \theta \end{cases}$$

- Sketch the graph of $f(x; k, \theta)$.
 - Verify that the total area under the graph equals 1.
 - If the rv X has pdf $f(x; k, \theta)$, obtain an expression for the cdf of X .
 - For $\theta < a < b$, obtain an expression for the probability $P(a \leq X \leq b)$.
 - Find an expression for the $(100p)$ th percentile η_p .
11. Let X denote the amount of time a book on 2-h reserve is actually checked out, and suppose the cdf is

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

Use this to compute the following:

- $P(X \leq 1)$
 - $P(.5 \leq X \leq 1)$
 - $P(X > 1.5)$
 - The median checkout duration η [Hint: Solve $F(\eta) = .5$.]
 - $F'(x)$ to obtain the density function $f(x)$
12. The cdf for $X =$ measurement error of Exercise 3 is

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{2} + \frac{3}{32} \left(4x - \frac{x^3}{3} \right) & -2 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

- Compute $P(X < 0)$.
 - Compute $P(-1 < X < 1)$.
 - Compute $P(X > .5)$.
 - Verify that $f(x)$ is as given in Exercise 3 by obtaining $F'(x)$.
 - Verify that $\eta = 0$.
13. Example 3.5 introduced the concept of time headway in traffic flow and proposed a particular distribution for $X =$ the headway between two randomly selected consecutive car. Suppose that in a different traffic environment, the distribution of time headway has the form

$$f(x) = \begin{cases} \frac{k}{x^4} & x > 1 \\ 0 & x \leq 1 \end{cases}$$

- (a) Determine the value of k for which $f(x)$ is a legitimate pdf.
 (b) Obtain the cumulative distribution function.
 (c) Use the cdf from (b) to determine the probability that headway exceeds 2 s and also the probability that headway is between 2 and 3 s.
14. Let X denote the amount of space occupied by an article placed in a 1-ft³ packing container. The pdf of X is

$$f(x) = \begin{cases} 90x^8(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Graph the pdf. Then obtain the cdf of X and graph it.
 (b) What is $P(X \leq .5)$ [i.e., $F(.5)$]?
 (c) Using part (a), what is $P(.25 < X \leq .5)$? What is $P(.25 \leq X \leq .5)$?
 (d) What is the 75th percentile of the distribution?
15. Answer parts (a)–(d) of Exercise 14 for the random variable X , lecture time past the hour, given in Exercise 5.
16. The article “A Model of Pedestrians’ Waiting Times for Street Crossings at Signalized Intersections” (*Transportation Research*, 2013: 17–28) suggested that under some circumstances the distribution of waiting time X could be modeled with the following pdf:

$$f(x; \theta, \tau) = \begin{cases} \frac{\theta}{\tau} (1 - x/\tau)^{\theta-1} & 0 \leq x < \tau \\ 0 & \text{otherwise} \end{cases}$$

where $\theta, \tau > 0$.

- (a) Graph $f(x; \theta, 80)$ for the three cases $\theta = 4, 1,$ and $.5$ (these graphs appear in the cited article) and comment on their shapes.
 (b) Obtain the cumulative distribution function of X .
 (c) Obtain an expression for the median of the waiting time distribution.
 (d) For the case $\theta = 4$ and $\tau = 80$, calculate $P(50 \leq X \leq 70)$ without doing any additional integration.
17. Let X be a continuous rv with cdf

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{4} \left[1 + \ln\left(\frac{4}{x}\right) \right] & 0 < x \leq 4 \\ 1 & x > 4 \end{cases}$$

[This type of cdf is suggested in the article “Variability in Measured Bedload-Transport Rates” (*Water Resources Bull.*, 1985: 39–48) as a model for a hydrologic variable.] What is

- (a) $P(X \leq 1)$?
 (b) $P(1 \leq X \leq 3)$?
 (c) The pdf of X ?

18. Let X be the temperature in $^{\circ}\text{C}$ at which a chemical reaction takes place, and let Y be the temperature in $^{\circ}\text{F}$ (so $Y = 1.8X + 32$).
- If the median of the X distribution is η , show that $1.8\eta + 32$ is the median of the Y distribution.
 - How is the 90th percentile of the Y distribution related to the 90th percentile of the X distribution? Verify your conjecture.
 - More generally, if $Y = aX + b$, how is any particular percentile of the Y distribution related to the corresponding percentile of the X distribution?

3.2 Expected Values and Moment Generating Functions

In Sect. 3.1 we saw that the transition from a discrete cdf to a continuous cdf entails replacing summation by integration. The same thing is true in moving from expected values of discrete variables to those of continuous variables.

3.2.1 Expected Values

For a discrete random variable X , the mean μ_X or $E(X)$ was defined as a weighted average and obtained by summing $x \cdot p(x)$ over possible X values. Here we replace summation by integration and the pmf by the pdf to get a continuous weighted average.

DEFINITION

The **expected value** or **mean value** of a continuous rv X with pdf $f(x)$ is

$$\mu = \mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Example 3.10 (Example 3.9 continued) The pdf of weekly gravel sales X was

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

so

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^1 x \cdot \frac{3}{2}(1-x^2) dx = \frac{3}{2} \int_0^1 (x-x^3) dx = \frac{3}{2} \left(\frac{x^2}{2} - \frac{x^4}{4} \right) \Big|_{x=0}^{x=1} = \frac{3}{8}$$

If gravel sales are determined week after week according to the given pdf, then the long-run average value of sales per week will be $.375$ ton. ■

Similar to the interpretation in the discrete case, the mean value μ can be regarded as the balance point (or fulcrum or center of mass) of a continuous distribution. In Example 3.10, if a piece of cardboard were cut out in the shape of the region under the density curve $f(x)$, then it would balance if supported at $\mu = 3/8$ along the bottom edge. When a pdf $f(x)$ is symmetric, then it will balance at its

point of symmetry, which must be the mean μ . Recall from Sect. 3.1 that the median is also the point of symmetry; in general, if a distribution is symmetric and the mean exists, then it is equal to the median.

Often we wish to compute the expected value of some function $h(X)$ of the rv X . If we think of $h(X)$ as a new rv Y , methods from Sect. 3.7 can be used to derive the pdf of Y , and $E(Y)$ can be computed from the definition. Fortunately, as in the discrete case, there is an easier way to compute $E[h(X)]$.

PROPOSITION

If X is a continuous rv with pdf $f(x)$ and $h(X)$ is any function of X , then

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

This is sometimes called the *Law of the Unconscious Statistician*.

Importantly, except in the case where $h(x)$ is a linear function (see later in this section), $E[h(X)]$ is *not* equal to $h(\mu_X)$, the function h evaluated at the mean of X .

Example 3.11 The variation in a certain electrical current source X (in milliamps) can be modeled by the pdf

$$f(x) = \begin{cases} 1.25 - .25x & 2 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

The average current from this source is

$$E(X) = \int_2^4 x(1.25 - .25x)dx = \frac{17}{6} = 2.833\text{mA}$$

If this current passes through a $220\text{-}\Omega$ resistor, the resulting power (in microwatts) is given by the expression $h(X) = (\text{current})^2(\text{resistance}) = 220X^2$. The expected power is given by

$$E(h(X)) = E(220X^2) = \int_2^4 220x^2(1.25 - .25x)dx = \frac{5500}{3} = 1833.3\mu\text{W}$$

Notice that the expected power is *not* equal to $220(2.833)^2$, a common error that results from substituting the mean current μ_X into the power formula. ■

Example 3.12 Two species are competing in a region for control of a limited amount of a resource. Let X = the proportion of the resource controlled by species 1 and suppose X has pdf

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

which is a uniform distribution on $[0, 1]$. (In her book *Ecological Diversity*, E. C. Pielou calls this the “broken-stick” model for resource allocation, since it is analogous to breaking a stick at a randomly chosen point.) Then the species that controls the majority of this resource controls the amount

$$h(X) = \max(X, 1 - X) = \begin{cases} 1 - X & \text{if } 0 \leq X < \frac{1}{2} \\ X & \text{if } \frac{1}{2} \leq X \leq 1 \end{cases}$$

The expected amount controlled by the species having majority control is then

$$\begin{aligned} E[h(X)] &= \int_{-\infty}^{\infty} \max(x, 1 - x) \cdot f(x) dx = \int_0^1 \max(x, 1 - x) \cdot 1 dx \\ &= \int_0^{1/2} (1 - x) \cdot 1 dx + \int_{1/2}^1 x \cdot 1 dx = \frac{3}{4} \end{aligned} \quad \blacksquare$$

In the discrete case, the variance of X was defined as the expected squared deviation from μ and was calculated by summation. Here again integration replaces summation.

DEFINITION

The **variance** of a continuous random variable X with pdf $f(x)$ and mean value μ is

$$\sigma_X^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx = E[(X - \mu)^2]$$

The **standard deviation** of X is $\sigma_X = \text{SD}(X) = \sqrt{\text{Var}(X)}$.

As in the discrete case, σ_X^2 is the expected or average squared deviation about the mean μ , and σ_X can be interpreted roughly as the size of a representative deviation from the mean value μ . Note that σ_X has the same units as X itself.

Example 3.13 Let $X \sim \text{Unif}[A, B]$. Since a uniform distribution is symmetric, the mean of X is at the density curve's point of symmetry, which is clearly the midpoint $(A + B)/2$. This can be verified by integration:

$$\mu = \int_A^B x \cdot \frac{1}{B - A} dx = \frac{1}{B - A} \frac{x^2}{2} \Big|_A^B = \frac{1}{B - A} \frac{B^2 - A^2}{2} = \frac{A + B}{2}$$

The variance of X is then given by

$$\begin{aligned} \sigma^2 &= \int_A^B (x - \mu)^2 \cdot \frac{1}{B - A} dx = \frac{1}{B - A} \int_A^B \left(x - \frac{A + B}{2}\right)^2 dx \\ &= \frac{1}{B - A} \int_{-(B-A)/2}^{(B-A)/2} u^2 du && \text{substitute } u = x - \frac{A + B}{2} \\ &= \frac{2}{B - A} \int_0^{(B-A)/2} u^2 du && \text{symmetry} \\ &= \frac{2}{B - A} \frac{u^3}{3} \Big|_0^{(B-A)/2} = \frac{2}{B - A} \frac{(B - A)^3}{2^3 \cdot 3} = \frac{(B - A)^2}{12} \end{aligned}$$

The standard deviation of X is the square root of the variance: $\sigma = (B-A)/\sqrt{12}$. Notice that the standard deviation of a $\text{Unif}[A, B]$ distribution is proportional to the length of the interval, $B - A$, which matches our intuitive notion that a larger standard deviation corresponds to greater “spread” in a distribution. ■

Section 2.3 presented several properties of expected value, variance, and standard deviation for discrete random variables. Those same properties hold for the continuous case; proofs of these results are obtained by replacing summation with integration in the proofs presented in Chap. 2.

PROPOSITION

Let X be a continuous rv with pdf $f(x)$, mean μ , and standard deviation σ . Then the following properties hold.

- (variance shortcut) $\text{Var}(X) = E(X^2) - \mu^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - \left(\int_{-\infty}^{\infty} x \cdot f(x)dx\right)^2$
- (Chebyshev’s inequality) For any constant $k \geq 1$,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- (linearity of expectation) For any functions $h_1(X)$ and $h_2(X)$ and any constants a_1, a_2 , and b ,

$$E[a_1h_1(X) + a_2h_2(X) + b] = a_1E[h_1(X)] + a_2E[h_2(X)] + b$$

- (rescaling) For any constants a and b ,

$$E(aX + b) = a\mu + b \quad \text{Var}(aX + b) = a^2\sigma^2 \quad \sigma_{aX+b} = |a|\sigma$$

Example 3.14 (Example 3.10 continued) For $X =$ weekly gravel sales, we computed $E(X) = 3/8$. Since

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot \frac{3}{2}(1-x^2)dx = \frac{3}{2} \int_0^1 (x^2 - x^4)dx = \frac{1}{5},$$

$$\text{Var}(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320} = .059 \quad \text{and} \quad \sigma_X = \sqrt{.059} = .244$$

Suppose the amount of gravel actually received by customers in a week is $h(X) = X - .02X^2$; the second term accounts for the small amount that is lost in transport. Then the average weekly amount received by customers is

$$E(X - .02X^2) = E(X) - .02E(X^2) = \frac{3}{8} - .02 \cdot \frac{1}{5} = .371 \text{ tons} \quad \blacksquare$$

Example 3.15 When a dart is thrown at a circular target, consider the location of the landing point relative to the bull’s eye. Let X be the angle in degrees measured from the horizontal, and assume that $X \sim \text{Unif}[0, 360)$. By Example 3.13, $E(X) = 180$ and $\text{SD}(X) = 360/\sqrt{12}$. Define Y to be the angle

measured in radians between $-\pi$ and π , so $Y = (2\pi/360)X - \pi$. Then, applying the rescaling properties with $a = 2\pi/360$ and $b = -\pi$,

$$E(Y) = \frac{2\pi}{360} \cdot E(X) - \pi = \frac{2\pi}{360} 180 - \pi = 0$$

and

$$\sigma_Y = \left| \frac{2\pi}{360} \right| \cdot \sigma_X = \frac{2\pi}{360} \frac{360}{\sqrt{12}} = \frac{2\pi}{\sqrt{12}} \quad \blacksquare$$

3.2.2 Moment Generating Functions

Moments and moment generating functions for discrete random variables were introduced in Sect. 2.7. These concepts carry over to the continuous case.

DEFINITION

The **moment generating function** (mgf) of a continuous random variable X is

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

As in the discrete case, the moment generating function exists iff $M_X(t)$ is defined for an interval that includes zero as well as positive and negative values of t .

Just as before, when $t = 0$ the value of the mgf is always 1:

$$M_X(0) = E(e^{0X}) = \int_{-\infty}^{\infty} e^{0x} f(x) dx = \int_{-\infty}^{\infty} f(x) dx = 1.$$

Example 3.16 At a store, the checkout time X in minutes has the pdf $f(x) = 2e^{-2x}$, $x \geq 0$; $f(x) = 0$ otherwise. Then

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_0^{\infty} e^{tx} (2e^{-2x}) dx = \int_0^{\infty} 2e^{-(2-t)x} dx \\ &= -\frac{2}{2-t} e^{-(2-t)x} \Big|_0^{\infty} = \frac{2}{2-t} - \frac{2}{2-t} \lim_{x \rightarrow \infty} e^{-(2-t)x} \end{aligned}$$

The limit above exists (in fact, it equals zero) provided the coefficient on x is negative, i.e., $-(2-t) < 0$. This is equivalent to $t < 2$. The mgf exists because it is defined for an interval of values including 0 in its interior, specifically $(-\infty, 2)$. For t in that interval, the mgf of X is $M_X(t) = 2/(2-t)$.

Notice that $M_X(0) = 2/(2-0) = 1$. Of course, from the calculation preceding this example we know that $M_X(0) = 1$ must always be the case, but it is useful as a check to set $t = 0$ and see if the result is 1. ■

Recall that in Sect. 2.7 we had a uniqueness property for the mgfs of discrete distributions. This proposition is equally valid in the continuous case: two distributions have the same pdf if and only if they have the same moment generating function, assuming that the mgf exists. For example, if a

random variable X is known to have mgf $M_X(t) = 2/(2-t)$ for $t < 2$, then from Example 3.16 it must necessarily be the case that the pdf of X is $f(x) = 2e^{-2x}$ for $x \geq 0$ and $f(x) = 0$ otherwise.

In the discrete case we also had a theorem on how to get moments from the mgf, and this theorem applies also in the continuous case: the r th moment of a continuous rv with mgf $M_X(t)$ is given by

$$E(X^r) = M_X^{(r)}(0),$$

the r th derivative of the mgf with respect to t evaluated at $t = 0$, if the mgf exists.

Example 3.17 (Example 3.16 continued) The mgf of the rv $X =$ checkout time at the store was found to be $M_X(t) = 2/(2-t) = 2(2-t)^{-1}$ for $t < 2$. To find the mean and standard deviation, first compute the derivatives:

$$M_X'(t) = -2(2-t)^{-2}(-1) = \frac{2}{(2-t)^2}$$

$$M_X''(t) = \frac{d}{dt} \left[2(2-t)^{-2} \right] = -4(2-t)^{-3}(-1) = \frac{4}{(2-t)^3}$$

Setting t to 0 in the first derivative gives the expected checkout time as

$$E(X) = M_X^{(1)}(0) = M_X'(0) = .5 \text{ min.}$$

Setting t to 0 in the second derivative gives the second moment

$$E(X^2) = M_X^{(2)}(0) = M_X''(0) = .5,$$

from which the variance of the checkout time is $\text{Var}(X) = \sigma^2 = E(X^2) - [E(X)]^2 = .5 - .5^2 = .25$ and the standard deviation is $\sigma = \sqrt{.25} = .5 \text{ min.}$ ■

We will sometimes need to transform X using a linear function $Y = aX + b$. As discussed in the discrete case, if X has the mgf $M_X(t)$ and $Y = aX + b$, then $M_Y(t) = e^{bt}M_X(at)$.

Example 3.18 Let $X \sim \text{Unif}[A, B]$. As verified in Exercise 32, the moment generating function of X is

$$M_X(t) = \begin{cases} \frac{e^{Bt} - e^{At}}{(B-A)t} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

In particular, consider the situation in Example 3.15. Let X , the angle measured in degrees, be uniform on $[0, 360]$, so $A = 0$ and $B = 360$. Then

$$M_X(t) = \frac{e^{360t} - 1}{360t} \quad t \neq 0, \quad M_X(0) = 1$$

Now let $Y = (2\pi/360)X - \pi$, so Y is the angle measured in radians between $-\pi$ and π . Using the mgf rule for linear transformations with $a = 2\pi/360$ and $b = -\pi$, we get

$$\begin{aligned}
 M_Y(t) &= e^{bt}M_X(at) = e^{-\pi t}M_X\left(\frac{2\pi t}{360}\right) \\
 &= e^{-\pi t}\frac{e^{360(2\pi/360)t} - 1}{360\left(\frac{2\pi t}{360}\right)} \\
 &= \frac{e^{\pi t} - e^{-\pi t}}{2\pi t} \quad t \neq 0, \quad M_Y(0) = 1
 \end{aligned}$$

This matches the general form of the moment generating function for a uniform random variable with $A = -\pi$ and $B = \pi$. Thus, by the mgf uniqueness property, $Y \sim \text{Unif}[-\pi, \pi]$. ■

3.2.3 Exercises: Section 3.2 (19–38)

19. Reconsider the distribution of checkout duration X described in Exercise 11. Compute the following:
 - (a) $E(X)$
 - (b) $\text{Var}(X)$ and $\text{SD}(X)$
 - (c) If the borrower is charged an amount $h(X) = X^2$ when checkout duration is X , compute the expected charge $E[h(X)]$.
20. The article “Modeling Sediment and Water Column Interactions for Hydrophobic Pollutants” (*Water Res.*, 1984: 1169–1174) suggests the uniform distribution on the interval $[7.5, 20]$ as a model for depth (cm) of the bioturbation layer in sediment in a certain region.
 - (a) What are the mean and variance of depth?
 - (b) What is the cdf of depth?
 - (c) What is the probability that observed depth is at most 10? Between 10 and 15?
 - (d) What is the probability that the observed depth is within 1 standard deviation of the mean value?
Within 2 standard deviations?
21. For the distribution of Exercise 14,
 - (a) Compute $E(X)$ and $\text{SD}(X)$.
 - (b) What is the probability that X is more than 1 standard deviation from its mean value?
22. Consider the pdf given in Exercise 6.
 - (a) Obtain and graph the cdf of X .
 - (b) From the graph of $f(x)$, what is the median, η ?
 - (c) Compute $E(X)$ and $\text{Var}(X)$.
23. Let $X \sim \text{Unif}[A, B]$.
 - (a) Obtain an expression for the $(100p)$ th percentile.
 - (b) Obtain an expression for the median, η . How does this compare to the mean μ , and why does that make sense for this distribution?
 - (c) For n a positive integer, compute $E(X^n)$.
24. Consider the pdf for total waiting time Y for two buses

$$f(y) = \begin{cases} \frac{1}{25}y & 0 \leq y < 5 \\ \frac{2}{5} - \frac{1}{25}y & 5 \leq y \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

introduced in Exercise 8.

- Compute and sketch the cdf of Y . [*Hint*: Consider separately $0 \leq y < 5$ and $5 \leq y \leq 10$ in computing $F(y)$. A graph of the pdf should be helpful.]
 - Obtain an expression for the $(100p)$ th percentile. [*Hint*: Consider separately $0 < p < .5$ and $.5 \leq p < 1$.]
 - Compute $E(Y)$ and $\text{Var}(Y)$. How do these compare with the expected waiting time and variance for a single bus when the time is uniformly distributed on $[0, 5]$?
 - Explain how symmetry can be used to obtain $E(Y)$.
25. An ecologist wishes to mark off a circular sampling region having radius 10 m. However, the radius of the resulting region is actually a random variable R with pdf

$$f(r) = \begin{cases} \frac{3}{4} [1 - (10 - r)^2] & 9 \leq r \leq 11 \\ 0 & \text{otherwise} \end{cases}$$

What is the expected area of the resulting circular region?

26. The weekly demand for propane gas (in 1000s of gallons) from a particular facility is an rv X with pdf

$$f(x) = \begin{cases} 2 \left(1 - \frac{1}{x^2}\right) & 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

- Compute the cdf of X .
 - Obtain an expression for the $(100p)$ th percentile. What is the value of the median, η ?
 - Compute $E(X)$. How do the mean and median of this distribution compare?
 - Compute $\text{Var}(X)$ and $\text{SD}(X)$.
 - If 1.5 thousand gallons are in stock at the beginning of the week and no new supply is due in during the week, how much of the 1.5 thousand gallons is expected to be left at the end of the week? [*Hint*: Let $h(x)$ = amount left when demand is x .]
27. If the temperature at which a compound melts is a random variable with mean value 120°C and standard deviation 2°C , what are the mean temperature and standard deviation measured in $^\circ\text{F}$? [*Hint*: $^\circ\text{F} = 1.8^\circ\text{C} + 32$.]
28. Let X have the Pareto pdf introduced in Exercise 10:

$$f(x; k, \theta) = \begin{cases} \frac{k \cdot \theta^k}{x^{k+1}} & x \geq \theta \\ 0 & x < \theta \end{cases}$$

- If $k > 1$, compute $E(X)$.
- What can you say about $E(X)$ if $k = 1$?
- If $k > 2$, show that $\text{Var}(X) = k\theta^2(k-1)^{-2}(k-2)^{-1}$.

- (d) If $k = 2$, what can you say about $\text{Var}(X)$?
- (e) What conditions on k are necessary to ensure that $E(X^n)$ is finite?
29. The time (min) between successive visits to a particular Web site has pdf $f(x) = 4e^{-4x}$, $x \geq 0$; $f(x) = 0$ otherwise. Use integration by parts to obtain $E(X)$ and $\text{SD}(X)$.
30. Consider the weights, in grams, of walnuts harvested at a nearby farm. Suppose this weight distribution can be modeled by the following pdf:

$$f(x) = \begin{cases} .5 - \frac{x}{8} & 0 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Show that $E(X) = 4/3$ and $\text{Var}(X) = 8/9$.
- (b) The *skewness coefficient* is defined as $E[(X - \mu)^3]/\sigma^3$. Show that its value for the given pdf is .566. What would the skewness be for a perfectly symmetric pdf?
31. The *delta method* provides approximations to the mean and variance of a nonlinear function $h(X)$ of a rv X . These approximations are based on a first-order Taylor series expansion of $h(x)$ about $x = \mu$, the mean of X :

$$h(X) \approx h_1(X) = h(\mu) + h'(\mu)(X - \mu)$$

- (a) Show that $E[h_1(X)] = h(\mu)$. (This is the delta method approximation to $E[h(X)]$.)
- (b) Show that $\text{Var}[h_1(X)] = [h'(\mu)]^2 \text{Var}(X)$. (This is the delta method approximation to $\text{Var}[h(X)]$.)
- (c) If the voltage v across a medium is fixed but current I is random, then resistance will also be a random variable related to I by $R = v/I$. If $\mu_I = 20$ and $\sigma_I = .5$, calculate approximations to μ_R and σ_R .
- (d) Let R have the distribution in Exercise 25, whose mean and variance are 10 and $1/5$, respectively. Let $h(R) = \pi R^2$, the area of the ecologist's sampling region. How does $E[h(R)]$ from Exercise 25 compare to the delta method approximation $h(10)$?
- (e) It can be shown that $\text{Var}[h(R)] = 14008\pi^2/175$. Compute the delta method approximation to $\text{Var}[h(R)]$ using the formula in (b). How good is the approximation?
32. Let $X \sim \text{Unif}[A, B]$, so its pdf is $f(x) = 1/(B - A)$, $A \leq x \leq B$, $f(x) = 0$ otherwise. Show that the moment generating function of X is

$$M_X(t) = \begin{cases} \frac{e^{Bt} - e^{At}}{(B - A)t} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

33. Let $X \sim \text{Unif}[0, 1]$. Find a linear function $Y = g(X)$ such that the interval $[0, 1]$ is transformed into $[-5, 5]$. Use the relationship for linear functions $M_{aX+b}(t) = e^{bt}M_X(at)$ to obtain the mgf of Y from the mgf of X . Compare your answer with the result of Exercise 32, and use this to obtain the pdf of Y .
34. If the pdf of a measurement error X is $f(x) = .5e^{-|x|}$, $-\infty < x < \infty$, show that $M_X(t) = 1/(1 - t^2)$ for $|t| < 1$.
35. Consider the rv $X = \text{time headway}$ in Example 3.5.
- (a) Find the moment generating function and use it to find the mean and variance.
- (b) Now consider a random variable whose pdf is

$$f(x) = \begin{cases} .15e^{-.15x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the moment generating function and use it to find the mean and variance. Compare with (a), and explain the similarities and differences.

- (c) Let $Y = X - .5$ and use the relationship for linear functions $M_{aX+b}(t) = e^{bt}M_X(at)$ to obtain the mgf of Y from (a). Compare with the result of (b) and explain.
36. Define $L_X(t) = \ln[M_X(t)]$. It was shown in Exercise 120 of Chap. 2 that $L'_X(0) = E(X)$ and $L''_X(0) = \text{Var}(X)$.
- (a) Determine $M_X(t)$ for the pdf in Exercise 29, and use this mgf to obtain $E(X)$ and $\text{Var}(X)$. How does this compare, in terms of difficulty, with the integration by parts required in that exercise?
- (b) Determine $L_X(t)$ for this same distribution, and use $L_X(t)$ to obtain $E(X)$ and $\text{Var}(X)$. How does the computational effort here compare with that of (a)?
37. Let X be a nonnegative, continuous rv with pdf $f(x)$ and cdf $F(x)$.
- (a) Show that, for any constant $t > 0$,

$$\int_t^\infty x \cdot f(x) dx \geq t \cdot P(X > t) = t \cdot [1 - F(t)]$$

- (b) Assume the mean of X is finite (i.e., the integral defining μ converges). Use part (a) to show that

$$\lim_{t \rightarrow \infty} t \cdot [1 - F(t)] = 0$$

[Hint: Write the integral for μ as the sum of two other integrals, one from 0 to t and another from t to ∞ .]

38. Let X be a nonnegative, continuous rv with cdf $F(x)$.
- (a) Assuming the mean μ of X is finite, show that

$$\mu = \int_0^\infty [1 - F(x)] dx$$

[Hint: Apply integration by parts to the integral above, and use the result of the previous exercise.] This is the continuous analog of the result established in Exercise 48 of Chap. 2.

- (b) A similar argument can be used to show that the k th moment of X is given by

$$E(X^k) = k \int_0^\infty x^{k-1} [1 - F(x)] dx$$

and that $E(X^k)$ exists iff $t^k[1 - F(t)] \rightarrow 0$ as $t \rightarrow \infty$. (This was the topic of a 2012 article in *The American Statistician*.) Suppose the lifetime X , in weeks, of a low-grade transistor under continuous use has cdf $F(x) = 1 - (x + 1)^{-3}$ for $x > 0$. Without finding the pdf of X , determine its mean and its standard deviation.

3.3 The Normal (Gaussian) Distribution

The normal distribution, often called the Gaussian distribution by engineers, is the most important one in all of probability and statistics. Many numerical populations have distributions that can be fit very closely by an appropriate normal curve. Examples include heights, weights, and other physical characteristics, measurement errors in scientific experiments, measurements on fossils, reaction times in psychological experiments, measurements of intelligence and aptitude, scores on various tests, and numerous economic measures and indicators. Even when the underlying distribution is discrete, the normal curve often gives an excellent approximation. In addition, even when individual variables themselves are not normally distributed, sums and averages of the variables will, under suitable conditions, have approximately a normal distribution; this is the content of the Central Limit Theorem discussed in Chap. 4.

DEFINITION

A continuous rv X is said to have a **normal distribution** (or **Gaussian distribution**) with parameters μ and σ , where $-\infty < \mu < \infty$ and $\sigma > 0$, if the pdf of X is

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)} \quad -\infty < x < \infty \quad (3.3)$$

The statement that X is normally distributed with parameters μ and σ is often abbreviated $X \sim N(\mu, \sigma)$.

Figure 3.13 presents graphs of $f(x; \mu, \sigma)$ for several different (μ, σ) pairs. Each resulting density curve is symmetric about μ and bell-shaped, so the center of the bell (point of symmetry) is both the mean of the distribution and the median. The value of σ is the distance from μ to the inflection points of the curve (the points at which the curve changes between turning downward to turning upward). Large values of σ yield density curves that are quite spread out about μ , whereas small values of σ yield density curves with a high peak above μ and most of the area under the density curve quite close to μ . Thus a large σ implies that a value of X far from μ may well be observed, whereas such a value is quite unlikely when σ is small.

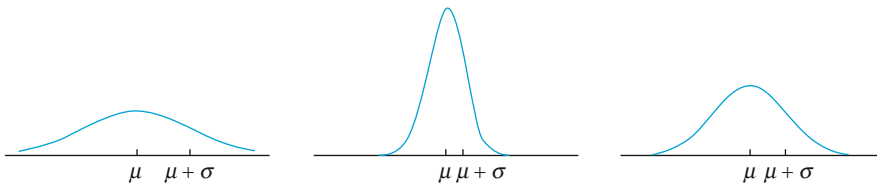


Fig. 3.13 Normal density curves

Clearly $f(x; \mu, \sigma) \geq 0$, but a somewhat complicated calculus argument is required to prove that $\int_{-\infty}^{\infty} f(x; \mu, \sigma) dx = 1$ (see Exercise 66). It can be shown using calculus (Exercise 67) or moment generating functions (Exercise 68) that $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$, so the parameters μ and σ are the mean and the standard deviation, respectively, of X .

3.3.1 The Standard Normal Distribution

To compute $P(a \leq X \leq b)$ when $X \sim N(\mu, \sigma)$, we must evaluate

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx \quad (3.4)$$

None of the standard integration techniques can be used here, and there is no closed-form expression for the integral. Table 3.1 at the end of this section provides the code for performing such normal distribution calculations in both Matlab and R. For the purpose of hand calculation of normal distribution probabilities, we now introduce a special normal distribution.

DEFINITION

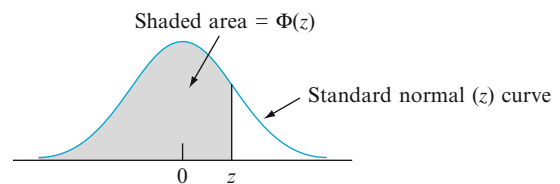
The normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**. A random variable that has a standard normal distribution is called a **standard normal random variable** and will be denoted by Z . The pdf of Z is

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

The cdf of Z is $P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$, which we will denote by $\Phi(z)$.

The standard normal distribution does not frequently serve as a model for a naturally arising population, since few variables have mean 0 and standard deviation 1. Instead, it is a reference distribution from which information about other normal distributions can be obtained. Appendix Table A.3 gives values of $\Phi(z)$ for $z = -3.49, -3.48, \dots, 3.48, 3.49$ and is referred to as the *standard normal table* or *z table*. Figure 3.14 illustrates the type of cumulative area (probability) tabulated in Table A.3. From this table, various other probabilities involving Z can be calculated.

Fig. 3.14 Standard normal cumulative areas tabulated in Appendix Table A.3



Example 3.19 Here we demonstrate how the z table is used to calculate various probabilities involving a standard normal rv.

- $P(Z \leq 1.25) = \Phi(1.25)$, a probability that is tabulated in Table A.3 at the intersection of the row marked 1.2 and the column marked .05. The number there is .8944, so $P(Z \leq 1.25) = .8944$. See Fig. 3.15a. In Matlab, we may type `normcdf(1.25, 0, 1)`; in R, use `pnorm(1.25, 0, 1)` or just `pnorm(1.25)`.
- $P(Z > 1.25) = 1 - P(Z \leq 1.25) = 1 - \Phi(1.25)$, the area under the standard normal curve to the right of 1.25 (an upper-tail area). Since $\Phi(1.25) = .8944$, it follows that $P(Z > 1.25) = .1056$. Since Z is a continuous rv, $P(Z \geq 1.25)$ also equals .1056. See Fig. 3.15b.

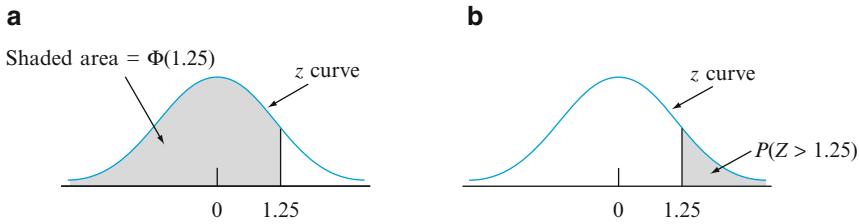


Fig. 3.15 Normal curve areas (probabilities) for Example 3.19

- (c) $P(Z \leq -1.25) = \Phi(-1.25)$, a lower-tail area. Directly from the z table, $\Phi(-1.25) = .1056$. By symmetry of the normal curve, this is identical to the probability in (b).
- (d) $P(-.38 \leq Z \leq 1.25)$ is the area under the standard normal curve above the interval $[-.38, 1.25]$. From Sect. 3.1, if Z is a continuous rv with cdf $F(z)$, then $P(a \leq Z \leq b) = F(b) - F(a)$. This gives $P(-.38 \leq Z \leq 1.25) = \Phi(1.25) - \Phi(-.38) = .8944 - .3520 = .5424$. (See Fig. 3.16.) To evaluate this probability in Matlab, type `normcdf(1.25, 0, 1) - normcdf(-.38, 0, 1)`; in R, type `pnorm(1.25, 0, 1) - pnorm(-.38, 0, 1)` or just `pnorm(1.25) - pnorm(-.38)`.

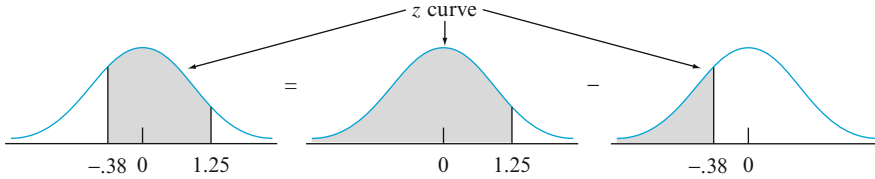


Fig. 3.16 $P(-.38 \leq Z \leq 1.25)$ as the difference between two cumulative areas

From Sect. 3.1, we have that the $(100p)$ th percentile of the standard normal distribution, for any p between 0 and 1, is the solution to the equation $\Phi(z) = p$. So, we may write the $(100p)$ th percentile of the standard normal distribution as $\eta_p = \Phi^{-1}(p)$. Matlab, R, or the z table can be used to obtain this percentile.

Example 3.20 The 99th percentile of the standard normal distribution, $\Phi^{-1}(.99)$, is the value on the horizontal axis such that the area under the curve to the left of the value is .9900, as illustrated in Fig. 3.17. To solve the “inverse” problem $\Phi(z) = p$, the standard normal table is used in an inverse fashion: Find in the middle of the table .9900; the row and column in which it lies identify the 99th z percentile. Here .9901 lies in the row marked 2.3 and column marked .03, so $\Phi(2.33) = .9901 \approx .99$ and the 99th percentile is approximately $z = 2.33$. By symmetry, the first percentile is the negative of the 99th percentile, so it equals -2.33 (1% lies below the first and above the 99th). See Fig. 3.18.

Fig. 3.17 Finding the 99th percentile

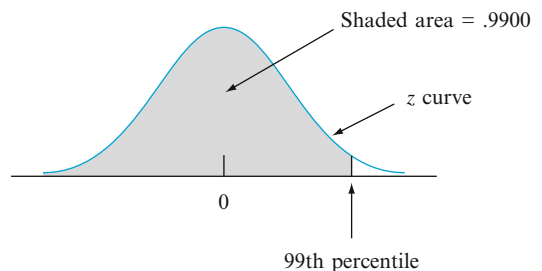
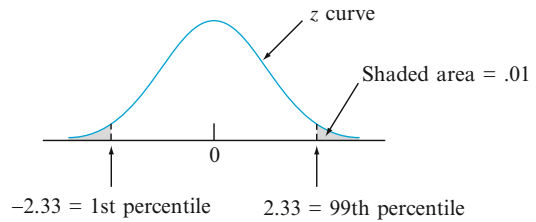


Fig. 3.18 The relationship between the 1st and 99th percentiles



To find the 99th percentile of the standard normal distribution in Matlab, use the command `norminv(.99, 0, 1)`; in R, `qnorm(.99, 0, 1)` or just `qnorm(.99)` produces that same value of roughly $z = 2.33$. ■

3.3.2 Non-standardized Normal Distributions

When $X \sim N(\mu, \sigma)$, probabilities involving X may be computed by “standardizing.” A **standardized variable** has the form $(X - \mu)/\sigma$. Subtracting μ shifts the mean from μ to zero, and then dividing by σ scales the variable so that the standard deviation is 1 rather than σ .

Standardizing amounts to nothing more than calculating a distance from the mean and then reexpressing the distance as some number of standard deviations. For example, if $\mu = 100$ and $\sigma = 15$, then $x = 130$ corresponds to $z = (130 - 100)/15 = 30/15 = 2.00$. That is, 130 is 2 standard deviations above (to the right of) the mean value. Similarly, standardizing 85 gives $(85 - 100)/15 = -1.00$, so 85 is 1 standard deviation below the mean. According to the next proposition, the z table applies to *any* normal distribution provided that we think in terms of number of standard deviations away from the mean value.

PROPOSITION

If $X \sim N(\mu, \sigma)$, then the “standardized” rv Z defined by

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

has a standard normal distribution. Thus

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right),$$

$$P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right), \quad P(X \geq b) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right),$$

and the $(100p)$ th percentile of the $N(\mu, \sigma)$ distribution is given by

$$\eta_p = \mu + \Phi^{-1}(p) \cdot \sigma.$$

Conversely, if $Z \sim N(0, 1)$ and μ and σ are constants (with $\sigma > 0$), then the “un-standardized” rv $X = \mu + \sigma Z$ has a normal distribution with mean μ and standard deviation σ .

Proof Let $X \sim N(\mu, \sigma)$ and define $Z = (X - \mu)/\sigma$ as in the statement of the proposition. Then the cdf of Z is given by

$$\begin{aligned}
 F_z(z) &= P(Z \leq z) \\
 &= P\left(\frac{X - \mu}{\sigma} \leq z\right) \\
 &= P(X \leq \mu + z\sigma) = \int_{-\infty}^{\mu+z\sigma} f(x; \mu, \sigma) dx = \int_{-\infty}^{\mu+z\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx
 \end{aligned}$$

Now make the substitution $u = (x - \mu)/\sigma$. The new limits of integration become $-\infty$ to z , and the differential dx is replaced by σdu , resulting in

$$F_z(z) = \int_{-\infty}^z \frac{1}{\sigma\sqrt{2\pi}} e^{-u^2/2} \sigma du = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = \Phi(z)$$

Thus, the cdf of $(X - \mu)/\sigma$ is the standard normal cdf, which establishes that $(X - \mu)/\sigma \sim N(0, 1)$.

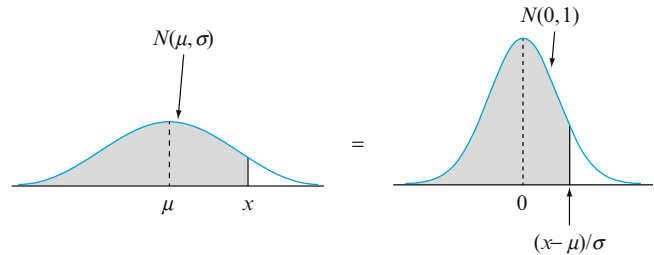
The probability formulas in the statement of the proposition follow directly from this main result, as does the formula for the $(100p)$ th percentile:

$$\begin{aligned}
 p &= P(X \leq \eta_p) = P\left(\frac{X - \mu}{\sigma} \leq \frac{\eta_p - \mu}{\sigma}\right) = \Phi\left(\frac{\eta_p - \mu}{\sigma}\right) \Rightarrow \frac{\eta_p - \mu}{\sigma} = \Phi^{-1}(p) \Rightarrow \\
 \eta_p &= \mu + \Phi^{-1}(p) \cdot \sigma
 \end{aligned}$$

The converse statement $Z \sim N(0, 1) \Rightarrow \mu + \sigma Z \sim N(\mu, \sigma)$ is derived similarly. ■

The key idea of this proposition is that by standardizing, any probability involving X can be expressed as a probability involving a standard normal rv Z , so that the z table can be used. This is illustrated in Fig. 3.19.

Fig. 3.19 Equality of nonstandard and standard normal curve areas



Software eliminates the need for standardizing X , although the standard normal distribution is still important in its own right. Table 3.1 at the end of this section details the relevant R and Matlab commands, which are also illustrated in the following examples.

Example 3.21 The time that it takes a driver to react to the brake lights on a decelerating vehicle is critical in avoiding rear-end collisions. The article “Fast-Rise Brake Lamp as a Collision-Prevention Device” (*Ergonomics*, 1993: 391–395) suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled with a normal distribution having mean value 1.25 s and standard deviation of .46 s. What is the probability that reaction time is between 1.00 s and 1.75 s? If we let X denote reaction time, then standardizing gives $1.00 \leq X \leq 1.75$ if and only if

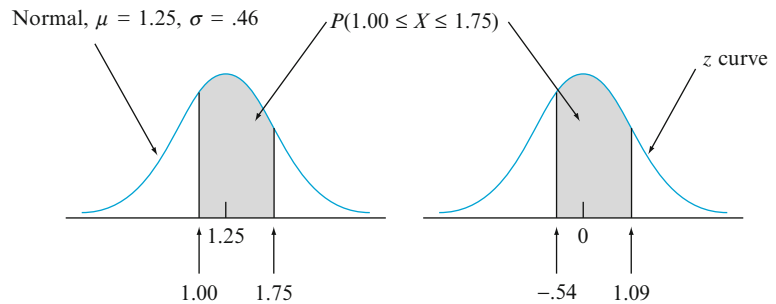
$$\frac{1.00 - 1.25}{.46} \leq \frac{X - 1.25}{.46} \leq \frac{1.75 - 1.25}{.46}$$

The middle expression, by the previous proposition, is a standard normal rv. Thus

$$\begin{aligned}
 P(1.00 \leq X \leq 1.75) &= P\left(\frac{1.00 - 1.25}{.46} \leq Z \leq \frac{1.75 - 1.25}{.46}\right) \\
 &= P(-.54 \leq Z \leq 1.09) = \Phi(1.09) - \Phi(-.54) \\
 &= .8621 - .2946 = .5675
 \end{aligned}$$

This is illustrated in Fig. 3.20. The same answer may be produced in Matlab with the command `normcdf(1.75, 1.25, .46) - normcdf(1.00, 1.25, .46)`; Matlab gives the answer .5681, which is more accurate than the value .5675 above (due to rounding the z -values to two decimal places). The analogous R command is `pnorm(1.75, 1.25, .46) - pnorm(1.00, 1.25, .46)`.

Fig. 3.20 Normal curves for Example 3.21



Similarly, if we view 2 s as a critically long reaction time, the probability that actual reaction time will exceed this value is

$$P(X > 2) = P\left(Z > \frac{2 - 1.25}{.46}\right) = P(Z > 1.63) = 1 - \Phi(1.63) = .0516$$

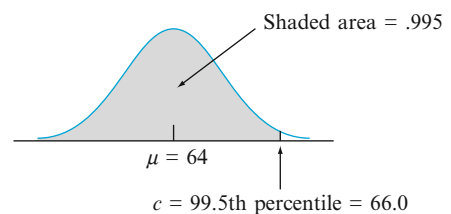
This probability is determined in Matlab and R by executing the commands `1 - normcdf(2, 1.25, .46)` and `1 - pnorm(2, 1.25, .46)`, respectively. ■

Example 3.22 The amount of distilled water dispensed by a machine is normally distributed with mean value 64 oz and standard deviation .78 oz. What container size c will ensure that overflow occurs only .5% of the time? If X denotes the amount dispensed, the desired condition is that $P(X > c) = .005$, or, equivalently, that $P(X \leq c) = .995$. Thus c is the 99.5th percentile of the normal distribution with $\mu = 64$ and $\sigma = .78$. The 99.5th percentile of the standard normal distribution is $\Phi^{-1}(.995) \approx 2.58$, so

$$c = \eta_{.995} = 64 + (2.58)(.78) = 64 + 2.0 = 66.0 \text{ oz}$$

This is illustrated in Fig. 3.21.

Fig. 3.21 Distribution of amount dispensed for Example 3.22



The Matlab and R commands to calculate this percentile are `norminv(.995, 64, .78)` and `qnorm(.995, 64, .78)`, respectively. ■

Example 3.23 The return on a diversified investment portfolio is normally distributed. What is the probability that the return is within 1 standard deviation of its mean value? This question can be answered without knowing either μ or σ , as long as the distribution is known to be normal; in other words, the answer is the same for *any* normal distribution. Going one standard deviation below μ lands us at $\mu - \sigma$, while $\mu + \sigma$ is one standard deviation above the mean. Thus

$$\begin{aligned} P\left(X \text{ is within one standard deviation of its mean}\right) &= P(\mu - \sigma \leq X \leq \mu + \sigma) \\ &= P\left(\frac{\mu - \sigma - \mu}{\sigma} \leq Z \leq \frac{\mu + \sigma - \mu}{\sigma}\right) \\ &= P(-1 \leq Z \leq 1) \\ &= \Phi(1) - \Phi(-1) = .6826 \end{aligned}$$

The probability that X is within 2 standard deviations of the mean is $P(-2 \leq Z \leq 2) = .9544$ and the probability that X is within 3 standard deviations of the mean is $P(-3 \leq Z \leq 3) = .9973$. ■

The results of Example 3.23 are often reported in percentage form and referred to as the *empirical rule* (because empirical evidence has shown that histograms of real data can very frequently be approximated by normal curves).

EMPIRICAL RULE

If the population distribution of a variable is (approximately) normal, then

1. Roughly 68% of the values are within 1 SD of the mean.
2. Roughly 95% of the values are within 2 SDs of the mean.
3. Roughly 99.7% of the values are within 3 SDs of the mean.

3.3.3 The Normal MGF

The moment generating function provides a straightforward way to establish several important results concerning the normal distribution.

PROPOSITION

The moment generating function of a normally distributed random variable X is

$$M_X(t) = e^{\mu t + \sigma^2 t^2 / 2}$$

Proof Consider first the special case of a standard normal rv Z . Then

$$M_Z(t) = E(e^{tZ}) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z^2 - 2tz)/2} dz$$

Completing the square in the exponent, we have

$$M_Z(t) = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z^2 - 2tz + t^2)/2} dz = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2/2} dz$$

The last integral is the area under a normal density curve with mean t and standard deviation 1, so the value of the integral is 1. Therefore, $M_Z(t) = e^{t^2/2}$.

Now let X be any normal rv with mean μ and standard deviation σ . Then, by the proposition earlier in this section, $(X - \mu)/\sigma = Z$, where Z is standard normal. Rewrite this relationship as $X = \mu + \sigma Z$, and use the property $M_{aY+b}(t) = e^{bt} M_Y(at)$:

$$M_X(t) = M_{\mu+\sigma Z}(t) = e^{t\mu} M_Z(\sigma t) = e^{t\mu} e^{\sigma^2 t^2/2} = e^{t\mu + \sigma^2 t^2/2} \quad \blacksquare$$

The normal mgf can be used to establish that μ and σ are indeed the mean and standard deviation of X , as claimed earlier (Exercise 68). Also, by the mgf uniqueness property, any rv X whose moment generating function has the form specified above is necessarily normally distributed. For example, if it is known that the mgf of X is $M_X(t) = e^{8t^2}$, then X must be a normal rv with mean $\mu = 0$ and standard deviation $\sigma = 4$ (since the $N(0, 4)$ distribution has e^{8t^2} as its mgf).

It was established earlier in this section that if $X \sim N(\mu, \sigma)$ and $Z = (X - \mu)/\sigma$, then $Z \sim N(0, 1)$, and vice versa. This standardizing transformation is actually a special case of a much more general property.

PROPOSITION

Let $X \sim N(\mu, \sigma)$. Then for any constants a and b with $a \neq 0$, $aX + b$ is also normally distributed. That is, any linear rescaling of a normal rv is normal.

The proof of this proposition uses mgfs and is left as an exercise (Exercise 70). This proposition provides a much easier proof of the earlier relationship between X and Z . The rescaling formulas and this proposition combine to give the following statement: if X is normally distributed and $Y = aX + b$ (with $a \neq 0$), then Y is also normal, with mean $\mu_Y = a\mu_X + b$ and standard deviation $\sigma_Y = |a|\sigma_X$.

3.3.4 The Normal Distribution and Discrete Populations

The normal distribution is often used as an approximation to the distribution of values in a discrete population. In such situations, extra care must be taken to ensure that probabilities are computed in an accurate manner.

Example 3.24 IQ (as measured by a standard test) is known to be approximately normally distributed with $\mu = 100$ and $\sigma = 15$. What is the probability that a randomly selected individual has an IQ of at least 125? Letting $X =$ the IQ of a randomly chosen person, we wish $P(X \geq 125)$. The temptation here is to standardize $X \geq 125$ immediately as in previous examples. However, the IQ population is actually discrete, since IQs are integer-valued, so the normal curve is an approximation to a discrete probability histogram, as pictured in Fig. 3.22.

The rectangles of the histogram are *centered* at integers, so IQs of at least 125 correspond to rectangles beginning at 124.5, as shaded in Fig. 3.22. Thus we really want the area under the approximating normal curve to the right of 124.5. Standardizing this value gives $P(Z \geq 1.63) = .0516$. If we had standardized $X \geq 125$, we would have obtained $P(Z \geq 1.67) = .0475$. The difference is not great, but the answer .0516 is more accurate. Similarly, $P(X = 125)$ would be approximated by the area between 124.5 and 125.5, since the area under the normal curve above the single value 125 is zero.

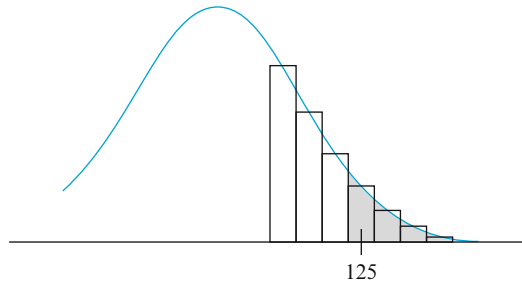


Fig. 3.22 A normal approximation to a discrete distribution ■

The correction for discreteness of the underlying distribution in Example 3.24 is often called a **continuity correction**; it adjusts for the use of a continuous distribution in approximating a probability involving a discrete rv. It is useful in the following application of the normal distribution to the computation of binomial probabilities. The normal distribution was actually created as an approximation to the binomial distribution (by Abraham de Moivre in the 1730s).

3.3.5 Approximating the Binomial Distribution

Recall that the mean value and standard deviation of a binomial random variable X are $\mu = np$ and $\sigma = \sqrt{npq}$, respectively. Figure 3.23a displays a probability histogram for the binomial distribution with $n = 20$, $p = .6$ [so $\mu = 20(.6) = 12$ and $\sigma = \sqrt{20(.6)(.4)} = 2.19$]. A normal curve with mean value and standard deviation equal to the corresponding values for the binomial distribution has been superimposed on the probability histogram. Although the probability histogram is a bit skewed (because $p \neq .5$), the normal curve gives a very good approximation, especially in the middle part of the picture. The area of any rectangle (probability of any particular X value) except those in the extreme tails can be accurately approximated by the corresponding normal curve area. For example, $P(X = 10) = \binom{20}{10} (.6)^{10} (.4)^{10} = .117$, whereas the area under the normal curve between 9.5 and 10.5 is $P(-1.14 \leq Z \leq -.68) = .120$.

On the other hand, a normal distribution is a poor approximation to a discrete distribution that is heavily skewed. For example, Figure 3.23b shows a probability histogram for the $\text{Bin}(20, .1)$

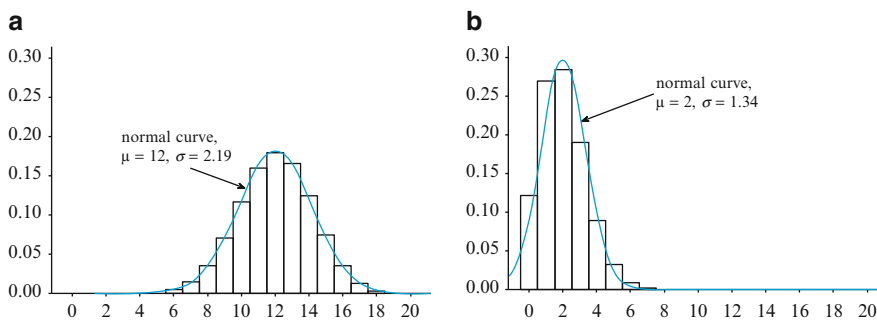


Fig. 3.23 Binomial probability histograms with normal approximation curves superimposed: (a) $n = 20$ and $p = .6$ (a good fit); (b) $n = 20$ and $p = .1$ (a poor fit)

distribution and the normal pdf with the same mean and standard deviation ($\mu = 2$ and $\sigma = 1.34$). Clearly, we would not want to use this normal curve to approximate binomial probabilities, even with a continuity correction.

PROPOSITION

Let X be a binomial rv based on n trials with success probability p . Then if the binomial probability histogram is not too skewed, X has approximately a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$. In particular, for $x =$ a possible value of X ,

$$\begin{aligned} P(X \leq x) = B(x; n, p) &\approx (\text{area under the normal curve to the left of } x + .5) \\ &= \Phi\left(\frac{x + .5 - np}{\sqrt{npq}}\right) \end{aligned}$$

In practice, the approximation is adequate provided that both $np \geq 10$ and $nq \geq 10$.

If either $np < 10$ or $nq < 10$, the binomial distribution may be too skewed for the (symmetric) normal curve to give accurate approximations.

Example 3.25 Suppose that 25% of all licensed drivers in a state do not have insurance. Let X be the number of uninsured drivers in a random sample of size 50 (somewhat perversely, a success is an uninsured driver), so that $p = .25$. Then $\mu = 12.5$ and $\sigma = 3.062$. Since $np = 50(.25) = 12.5 \geq 10$ and $nq = 37.5 \geq 10$, the approximation can safely be applied:

$$\begin{aligned} P(X \leq 10) = B(10; 50, .25) &\approx \Phi\left(\frac{10 + .5 - 12.5}{3.062}\right) \\ &= \Phi(-.6532) = .2568 \end{aligned}$$

Similarly, the probability that between 5 and 15 (inclusive) of the selected drivers are uninsured is

$$\begin{aligned} P(5 \leq X \leq 15) &= B(15; 50, .25) - B(4; 50, .25) \\ &\approx \Phi\left(\frac{15.5 - 12.5}{3.062}\right) - \Phi\left(\frac{4.5 - 12.5}{3.062}\right) = .8319 \end{aligned}$$

The exact probabilities are .2622 and .8348, respectively, so the approximations are quite good. In the last calculation, the probability $P(5 \leq X \leq 15)$ is being approximated by the area under the normal curve between 4.5 and 15.5—the continuity correction is used for both the upper and lower limits. ■

The wide availability of software for doing binomial probability calculations, even for large values of n , has considerably diminished the importance of the normal approximation. However, it is important for another reason. When the objective of an investigation is to make an inference about a population proportion p , interest will focus on the sample proportion of successes $\hat{P} = X/n$ rather than on X itself. Because this proportion is just X multiplied by the constant $1/n$, the earlier rescaling proposition tells us that \hat{P} will also have approximately a normal distribution (with mean $\mu = p$ and standard deviation $\sigma = \sqrt{pq/n}$) provided that both $np \geq 10$ and $nq \geq 10$. This normal approximation is the basis for several inferential procedures to be discussed in Chap. 5.

It is quite difficult to give a direct proof of the validity of this normal approximation (the first one goes back about 270 years to de Moivre). In Chap. 4, we'll see that it is a consequence of an important general result called the Central Limit Theorem.

3.3.6 Normal Distribution Calculations with Software

Many software packages, including Matlab and R, have built-in functions to determine both probabilities under a normal curve and quantiles (aka percentiles) of any given normal distribution. Table 3.1 summarizes the relevant code in both packages.

Table 3.1 Normal probability and quantile calculations in Matlab and R

Function:	cdf	quantile, i.e., the $(100p)$ th percentile
Notation:	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	$\eta_p = \mu + \Phi^{-1}(p) \cdot \sigma$
Matlab:	<code>normcdf(x, μ, σ)</code>	<code>norminv(p, μ, σ)</code>
R:	<code>pnorm(x, μ, σ)</code>	<code>qnorm(p, μ, σ)</code>

In the special case of a standard normal distribution, R (but not Matlab) will allow the user to drop the last two arguments, μ and σ . That is, the R commands `pnorm(x)` and `pnorm(x, 0, 1)` yield the same result for any number x , and a similar comment applies to `qnorm`. Both software packages also have built-in function calls for the normal pdf: `normpdf(x, μ, σ)` and `dnorm(x, μ, σ)`, respectively. However, these two commands are generally only used when one desires to *graph* a normal density curve (x vs. $f(x; \mu, \sigma)$), since the pdf evaluated at particular x does not represent a probability, as discussed in Sect. 3.1.

3.3.7 Exercises: Section 3.3 (39–70)

39. Let Z be a standard normal random variable and obtain each of the following probabilities, drawing pictures wherever appropriate.
- $P(0 \leq Z \leq 2.17)$
 - $P(0 \leq Z \leq 1)$
 - $P(-2.50 \leq Z \leq 0)$
 - $P(-2.50 \leq Z \leq 2.50)$
 - $P(Z \leq 1.37)$
 - $P(-1.75 \leq Z)$
 - $P(-1.50 \leq Z \leq 2.00)$
 - $P(1.37 \leq Z \leq 2.50)$
 - $P(1.50 \leq Z)$
 - $P(|Z| \leq 2.50)$
40. In each case, determine the value of the constant c that makes the probability statement correct.
- $\Phi(c) = .9838$
 - $P(0 \leq Z \leq c) = .291$
 - $P(c \leq Z) = .121$
 - $P(-c \leq Z \leq c) = .668$
 - $P(c \leq |Z|) = .016$

41. Find the following percentiles for the standard normal distribution. Interpolate where appropriate.
- 91st
 - 9th
 - 75th
 - 25th
 - 6th
42. Suppose the force acting on a column that helps to support a building is a normally distributed random variable X with mean value 15.0 kips and standard deviation 1.25 kips. Compute the following probabilities.
- $P(X \leq 15)$
 - $P(X \leq 17.5)$
 - $P(X \geq 10)$
 - $P(14 \leq X \leq 18)$
 - $P(|X - 15| \leq 3)$
43. Mopeds (small motorcycles with an engine capacity below 50 cc) are very popular in Europe because of their mobility, ease of operation, and low cost. The article “Procedure to Verify the Maximum Speed of Automatic Transmission Mopeds in Periodic Motor Vehicle Inspections” (*J. of Automobile Engr.*, 2008: 1615-1623) described a rolling bench test for determining maximum vehicle speed. A normal distribution with mean value 46.8 km/h and standard deviation 1.75 km/h is postulated. Consider randomly selecting a single such moped.
- What is the probability that maximum speed is at most 50 km/h?
 - What is the probability that maximum speed is at least 48 km/h?
 - What is the probability that maximum speed differs from the mean value by at most 1.5 standard deviations?
44. Let X be the birth weight, in grams, of a randomly selected full-term baby. The article “Fetal Growth Parameters and Birth Weight: Their Relationship to Neonatal Body Composition” (*Ultrasound in Obstetrics and Gynecology*, 2009: 441–446) suggests that X is normally distributed with mean 3500 and standard deviation 600.
- Sketch the relevant density curve, including tick marks on the horizontal scale.
 - What is $P(3000 < X < 4500)$, and how does this compare to $P(3000 \leq X \leq 4500)$?
 - What is the probability that the weight of such a newborn is less than 2500 g?
 - What is the probability that the weight of such a newborn exceeds 6000 g (roughly 13.2 lb)?
 - How would you characterize the most extreme .1% of all birth weights?
 - Use the rescaling proposition from this section to determine the distribution of birth weight expressed in pounds (shape, mean, and standard deviation), and then recalculate the probability from part (c). How does this compare to your previous answer?
45. Based on extensive data from an urban freeway near Toronto, Canada, “it is assumed that free speeds can best be represented by a normal distribution” (“Impact of Driver Compliance on the Safety and Operational Impacts of Freeway Variable Speed Limit Systems,” *J. of Transp. Engr.*, 2011: 260–268). The mean and standard deviation reported in the article were 119 km/h and 13.1 km/h, respectively.
- What is the probability that the speed of a randomly selected vehicle is between 100 and 120 km/h?
 - What speed characterizes the fastest 10% of all speeds?

- (c) The posted speed limit was 100 km/h. What percentage of vehicles was traveling at speeds exceeding this posted limit?
 - (d) If five vehicles are randomly and independently selected, what is the probability that at least one is not exceeding the posted speed limit?
 - (e) What is the probability that the speed of a randomly selected vehicle exceeds 70 miles/h?
46. The defect length of a corrosion defect in a pressurized steel pipe is normally distributed with mean value 30 mm and standard deviation 7.8 mm (suggested in the article “Reliability Evaluation of Corroding Pipelines Considering Multiple Failure Modes and Time-Dependent Internal Pressure,” *J. of Infrastructure Systems*, 2011: 216–224).
- (a) What is the probability that defect length is at most 20 mm? Less than 20 mm?
 - (b) What is the 75th percentile of the defect length distribution, i.e., the value that separates the smallest 75% of all lengths from the largest 25%?
 - (c) What is the 15th percentile of the defect length distribution?
 - (d) What values separate the middle 80% of the defect length distribution from the smallest 10% and the largest 10%?
47. The plasma cholesterol level (mg/dL) for patients with no prior evidence of heart disease who experience chest pain is normally distributed with mean 200 and standard deviation 35. Consider randomly selecting an individual of this type. What is the probability that the plasma cholesterol level
- (a) Is at most 250?
 - (b) Is between 300 and 400?
 - (c) Differs from the mean by at least 1.5 standard deviations?
48. Suppose the diameter at breast height (in.) of trees of a certain type is normally distributed with $\mu = 8.8$ and $\sigma = 2.8$, as suggested in the article “Simulating a Harvester-Forwarder Softwood Thinning” (*Forest Products J.*, May 1997: 36–41).
- (a) What is the probability that the diameter of a randomly selected tree will be at least 10 in.? Will exceed 10 in.?
 - (b) What is the probability that the diameter of a randomly selected tree will exceed 20 in.?
 - (c) What is the probability that the diameter of a randomly selected tree will be between 5 and 10 in.?
 - (d) What value c is such that the interval $(8.8 - c, 8.8 + c)$ includes 98% of all diameter values?
 - (e) If four trees are independently selected, what is the probability that at least one has a diameter exceeding 10 in.?
49. There are two machines available for cutting corks intended for use in wine bottles. The first produces corks with diameters that are normally distributed with mean 3 cm and standard deviation .1 cm. The second machine produces corks with diameters that have a normal distribution with mean 3.04 cm and standard deviation .02 cm. Acceptable corks have diameters between 2.9 and 3.1 cm. Which machine is more likely to produce an acceptable cork?
50. Human body temperatures for healthy individuals have approximately a normal distribution with mean 98.25 °F and standard deviation .75 °F. (The past accepted value of 98.6 °F was obtained by converting the Celsius value of 37°, which is correct to the nearest integer.)
- (a) Find the 90th percentile of the distribution.
 - (b) Find the 5th percentile of the distribution.
 - (c) What temperature separates the coolest 25% from the others?
51. The article “Monte Carlo Simulation—Tool for Better Understanding of LRFD” (*J. Struct. Engr.*, 1993: 1586–1599) suggests that yield strength (ksi) for A36 grade steel is normally distributed with $\mu = 43$ and $\sigma = 4.5$.

- (a) What is the probability that yield strength is at most 40? Greater than 60?
- (b) What yield strength value separates the strongest 75% from the others?
52. The automatic opening device of a military cargo parachute has been designed to open when the parachute is 200 m above the ground. Suppose opening altitude actually has a normal distribution with mean value 200 m and standard deviation 30 m. Equipment damage will occur if the parachute opens at an altitude of less than 100 m. What is the probability that there is equipment damage to the payload of at least one of five independently dropped parachutes?
53. The temperature reading from a thermocouple placed in a constant-temperature medium is normally distributed with mean μ , the actual temperature of the medium, and standard deviation σ . What would the value of σ have to be to ensure that 95% of all readings are within $.1^\circ$ of μ ?
54. Vehicle speed on a particular bridge in China can be modeled as normally distributed (“Fatigue Reliability Assessment for Long-Span Bridges under Combined Dynamic Loads from Winds and Vehicles,” *J. of Bridge Engr.*, 2013: 735–747).
- (a) If 5% of all vehicles travel less than 39.12 mph and 10% travel more than 73.24 mph, what are the mean and standard deviation of vehicle speed? [Note: The resulting values should agree with those given in the cited article.]
- (b) What is the probability that a randomly selected vehicle’s speed is between 50 and 65 mph?
- (c) What is the probability that a randomly selected vehicle’s speed exceeds the speed limit of 70 mph?
55. If adult female heights are normally distributed, what is the probability that the height of a randomly selected woman is
- (a) Within 1.5 SDs of its mean value?
- (b) Farther than 2.5 SDs from its mean value?
- (c) Between 1 and 2 SDs from its mean value?
56. A machine that produces ball bearings has initially been set so that the true average diameter of the bearings it produces is .500 in. A bearing is acceptable if its diameter is within .004 in. of this target value. Suppose, however, that the setting has changed during the course of production, so that the bearings have normally distributed diameters with mean value .499 in. and standard deviation .002 in. What percentage of the bearings produced will not be acceptable?
57. The Rockwell hardness of a metal is determined by impressing a hardened point into the surface of the metal and then measuring the depth of penetration of the point. Suppose the Rockwell hardness of an alloy is normally distributed with mean 70 and standard deviation 3. (Rockwell hardness is measured on a continuous scale.)
- (a) If a specimen is acceptable only if its hardness is between 67 and 75, what is the probability that a randomly chosen specimen has an acceptable hardness?
- (b) If the acceptable range of hardness is $(70 - c, 70 + c)$, for what value of c would 95% of all specimens have acceptable hardness?
- (c) If the acceptable range is as in part (a) and the hardness of each of ten randomly selected specimens is independently determined, what is the expected number of acceptable specimens among the ten?
- (d) What is the probability that at most eight of ten independently selected specimens have a hardness of less than 73.84? [Hint: $Y =$ the number among the ten specimens with hardness less than 73.84 is a binomial variable; what is p ?]
58. The weight distribution of parcels sent in a certain manner is normal with mean value 12 lb and standard deviation 3.5 lb. The parcel service wishes to establish a weight value c beyond which there will be a surcharge. What value of c is such that 99% of all parcels are at least 1 lb under the surcharge weight?

59. Suppose Appendix Table A.3 contained $\Phi(z)$ only for $z \geq 0$. Explain how you could still compute
- $P(-1.72 \leq Z \leq -.55)$
 - $P(-1.72 \leq Z \leq .55)$
- Is it necessary to tabulate $\Phi(z)$ for z negative? What property of the standard normal curve justifies your answer?
60. Chebyshev's inequality (Sect. 3.2) states that for any number k satisfying $k \geq 1$, $P(|X - \mu| \geq k\sigma)$ is no more than $1/k^2$. Obtain this probability in the case of a normal distribution for $k = 1, 2$, and 3 , and compare to Chebyshev's upper bound.
61. Let X denote the number of flaws along a 100-m reel of magnetic tape (an integer-valued variable). Suppose X has approximately a normal distribution with $\mu = 25$ and $\sigma = 5$. Use the continuity correction to calculate the probability that the number of flaws is
- Between 20 and 30, inclusive.
 - At most 30. Less than 30.
62. Let X have a binomial distribution with parameters $n = 25$ and p . Calculate each of the following probabilities using the normal approximation (with the continuity correction) for the cases $p = .5, .6$, and $.8$ and compare to the exact probabilities calculated from Appendix Table A.1.
- $P(15 \leq X \leq 20)$
 - $P(X \leq 15)$
 - $P(20 \leq X)$
63. Suppose that 10% of all steel shafts produced by a process are nonconforming but can be reworked (rather than having to be scrapped). Consider a random sample of 200 shafts, and let X denote the number among these that are nonconforming and can be reworked. What is the (approximate) probability that X is
- At most 30?
 - Less than 30?
 - Between 15 and 25 (inclusive)?
64. Suppose only 70% of all drivers in a state regularly wear a seat belt. A random sample of 500 drivers is selected. What is the probability that
- Between 320 and 370 (inclusive) of the drivers in the sample regularly wear a seat belt?
 - Fewer than 325 of those in the sample regularly wear a seat belt? Fewer than 315?
65. In response to concerns about nutritional contents of fast foods, McDonald's announced that it would use a new cooking oil for its french fries that would decrease substantially trans fatty acid levels and increase the amount of more beneficial polyunsaturated fat. The company claimed that 97 out of 100 people cannot detect a difference in taste between the new and old oils. Assuming that this figure is correct (as a long-run proportion), what is the approximate probability that in a random sample of 1000 individuals who have purchased fries at McDonald's,
- At least 40 can taste the difference between the two oils?
 - At most 5% can taste the difference between the two oils?
66. The following proof that the normal pdf integrates to 1 comes courtesy of Professor Robert Young, Oberlin College. Let $f(z)$ denote the standard normal pdf, and consider the function of two variables

$$g(x, y) = f(x) \cdot f(y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

Let V denote the volume under $g(x, y)$ above the xy -plane.

- Let A denote the area under the standard normal curve. By setting up the double integral for the volume underneath $g(x, y)$, show that $V = A^2$.

- (b) Using the rotational symmetry of $g(x, y)$, V can be determined by adding up the volumes of shells from rotation about the y -axis:

$$V = \int_0^{\infty} 2\pi r \cdot \frac{1}{2\pi} e^{-r^2/2} dr$$

Show this integral equals 1, then use (a) to establish that the area under the standard normal curve is 1.

- (c) Show that $\int_{-\infty}^{\infty} f(x; \mu, \sigma) dx = 1$. [Hint: Write out the integral, and then make a substitution to reduce it to the standard normal case. Then invoke (b).]
67. Suppose $X \sim N(\mu, \sigma)$.
- (a) Show via integration that $E(X) = \mu$. [Hint: Make the substitution $u = (x - \mu)/\sigma$, which will create two integrals. For one, use the symmetry of the pdf; for the other, use the fact that the standard normal pdf integrates to 1.]
- (b) Show via integration that $\text{Var}(X) = \sigma^2$. [Hint: Evaluate the integral for $E[(X - \mu)^2]$ rather than using the variance shortcut formula. Use the same substitution as in part (a).]
68. The moment generating function can be used to find the mean and variance of the normal distribution.
- (a) Use derivatives of $M_X(t)$ to verify that $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$.
- (b) Repeat (a) using $L_X(t) = \ln[M_X(t)]$, and compare with part (a) in terms of effort. (Refer back to Exercise 36 for properties of the function $L_X(t)$.)
69. There is no nice formula for the standard normal cdf $\Phi(z)$, but several good approximations have been published in articles. The following is from “Approximations for Hand Calculators Using Small Integer Coefficients” (*Math. Comput.*, 1977: 214–222). For $0 < z \leq 5.5$,

$$P(Z \geq z) = 1 - \Phi(z) \approx .5 \exp \left\{ - \left[\frac{(83z + 351)z + 562}{(703/z) + 165} \right] \right\}$$

The relative error of this approximation is less than .042%. Use this to calculate approximations to the following probabilities, and compare whenever possible to the probabilities obtained from Appendix Table A.3.

- (a) $P(Z \geq 1)$
- (b) $P(Z < -3)$
- (c) $P(-4 < Z < 4)$
- (d) $P(Z > 5)$
70. (a) Use mgfs to show that if X has a normal distribution with parameters μ and σ , then $Y = aX + b$ (a linear function of X) also has a normal distribution. What are the parameters of the distribution of Y [i.e., $E(Y)$ and $\text{SD}(Y)$]?
- (b) If when measured in $^{\circ}\text{C}$, temperature is normally distributed with mean 115 and standard deviation 2, what can be said about the distribution of temperature measured in $^{\circ}\text{F}$?

3.4 The Exponential and Gamma Distributions

The graph of any normal pdf is bell-shaped and thus symmetric. In many practical situations, the variable of interest to the experimenter might have a skewed distribution. A family of pdfs that yields a wide variety of skewed distributional shapes is the gamma family. We first consider a special case, the exponential distribution, and then generalize later in the section.

3.4.1 The Exponential Distribution

The family of exponential distributions provides probability models that are widely used in engineering and science disciplines.

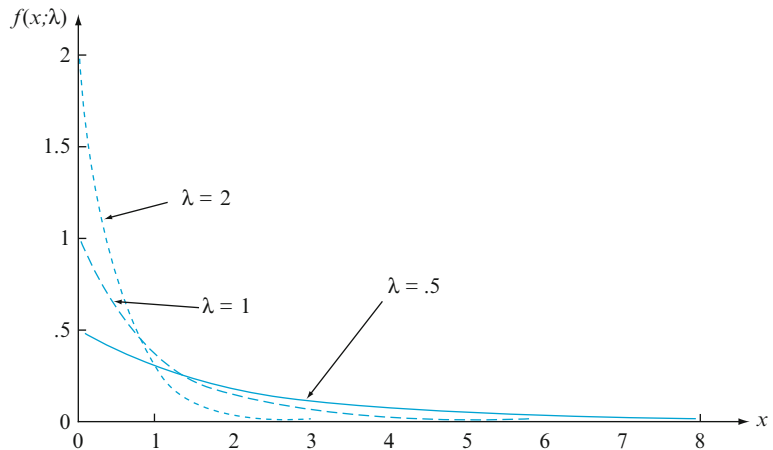
DEFINITION

X is said to have an **exponential distribution** with parameter λ ($\lambda > 0$) if the pdf of X is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Some sources write the exponential pdf in the form $(1/\beta)e^{-x/\beta}$, so that $\beta = 1/\lambda$. Graphs of several exponential pdfs appear in Fig. 3.24.

Fig. 3.24 Exponential density curves



The expected value of an exponentially distributed random variable X is

$$E(X) = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx$$

Obtaining this expected value requires integration by parts. The variance of X can be computed using the shortcut formula $\text{Var}(X) = E(X^2) - [E(X)]^2$; evaluating $E(X^2)$ uses integration by parts twice in succession. In contrast, the exponential cdf is easily obtained by integrating the pdf. The results of these integrations are as follows.

PROPOSITION

Let X be an exponential variable with parameter λ . Then the cdf of X is

$$F(x; \lambda) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

The mean and standard deviation of X are both equal to $1/\lambda$.

Under the alternative parameterization, the exponential cdf becomes $1 - e^{-x/\beta}$ for $x > 0$, and the mean and standard deviation are both equal to β .

Example 3.23 The response time X at an on-line computer terminal (the elapsed time between the end of a user's inquiry and the beginning of the system's response to that inquiry) has an exponential distribution with expected response time equal to 5 s. Then $E(X) = 1/\lambda = 5$, so $\lambda = .2$. The probability that the response time is at most 10 s is

$$P(X \leq 10) = F(10; 2) = 1 - e^{-(.2)(10)} = 1 - e^{-2} = 1 - .135 = .865$$

The probability that response time is between 5 and 10 s is

$$P(5 \leq X \leq 10) = F(10; 2) - F(5; 2) = (1 - e^{-2}) - (1 - e^{-1}) = .233 \quad \blacksquare$$

The exponential distribution is frequently used as a model for the distribution of times between the occurrence of successive events, such as customers arriving at a service facility or calls coming in to a call center. The reason for this is that the exponential distribution is closely related to the Poisson distribution introduced in Chap. 2. We will explore this relationship fully in Sect. 7.5 (Poisson Processes).

Another important application of the exponential distribution is to model the distribution of component lifetimes. A partial reason for the popularity of such applications is the “**memoryless**” **property** of the exponential distribution. Suppose component lifetime is exponentially distributed with parameter λ . After putting the component into service, we leave for a period of t_0 hours and then return to find the component still working; what now is the probability that it lasts at least an additional t hours? In symbols, we wish $P(X \geq t + t_0 | X \geq t_0)$. By the definition of conditional probability,

$$P(X \geq t + t_0 | X \geq t_0) = \frac{P[(X \geq t + t_0) \cap (X \geq t_0)]}{P(X \geq t_0)}$$

But the event $X \geq t_0$ in the numerator is redundant, since both events can occur if and only if $X \geq t + t_0$. Therefore,

$$P(X \geq t + t_0 | X \geq t_0) = \frac{P(X \geq t + t_0)}{P(X \geq t_0)} = \frac{1 - F(t + t_0; \lambda)}{1 - F(t_0; \lambda)} = \frac{e^{-\lambda(t+t_0)}}{e^{-\lambda t_0}} = e^{-\lambda t}$$

This conditional probability is identical to the original probability $P(X \geq t)$ that the component lasted t hours. Thus *the distribution of additional lifetime is exactly the same as the original distribution of lifetime*, so at each point in time the component shows no effect of wear. In other words, the distribution of remaining lifetime is independent of current age (we wish that were true of us!).

Although the memoryless property can be justified at least approximately in many applied problems, in other situations components deteriorate with age or occasionally improve with age (at least up to a certain point). More general lifetime models are then furnished by the gamma,

Weibull, and lognormal distributions (the latter two are discussed in the next section). Lifetime distributions are at the heart of *reliability* models, which we'll consider in depth in Sect. 4.8.

3.4.2 The Gamma Distribution

To define the family of gamma distributions, which generalizes the exponential distribution, we first need to introduce a function that plays an important role in many branches of mathematics.

DEFINITION

For $\alpha > 0$, the **gamma function** $\Gamma(\alpha)$ is defined by

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

The most important properties of the gamma function are the following:

1. For any $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ (via integration by parts)
2. For any positive integer n , $\Gamma(n) = (n - 1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

The following proposition will prove useful for several computations that follow.

PROPOSITION

For any $\alpha, \beta > 0$,

$$\int_0^{\infty} x^{\alpha-1} e^{-x/\beta} dx = \beta^{\alpha} \Gamma(\alpha) \quad (3.5)$$

Proof Make the substitution $u = x/\beta$, so that $x = \beta u$ and $dx = \beta du$:

$$\int_0^{\infty} x^{\alpha-1} e^{-x/\beta} dx = \int_0^{\infty} (\beta u)^{\alpha-1} e^{-u} \beta du = \beta^{\alpha} \int_0^{\infty} u^{\alpha-1} e^{-u} du = \beta^{\alpha} \Gamma(\alpha)$$

The last equality comes from the definition of the gamma function. ■

With the preceding proposition in mind, we make the following definition.

DEFINITION

A continuous random variable X is said to have a **gamma distribution** if the pdf of X is

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^{\alpha} \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

where the parameters α and β satisfy $\alpha > 0$, $\beta > 0$. When $\beta = 1$, X is said to have a **standard gamma distribution**, and its pdf may be denoted $f(x; \alpha)$.

The exponential distribution results from taking $\alpha = 1$ and $\beta = 1/\lambda$.

It's clear that $f(x; \alpha, \beta) \geq 0$ for all x ; the previous proposition guarantees that this function integrates to 1, as required. Figure 3.25a illustrates the graphs of the gamma pdf for several (α, β) pairs, whereas Fig. 3.25b presents graphs of the standard gamma pdf. For the standard pdf, when $\alpha \leq 1$, $f(x; \alpha)$ is strictly decreasing as x increases; when $\alpha > 1$, $f(x; \alpha)$ rises to a maximum and then decreases. Because of this difference, α is referred to as a *shape parameter*. The parameter β in Eq. (3.6) is called the *scale parameter* because values other than 1 either stretch or compress the pdf in the x direction.

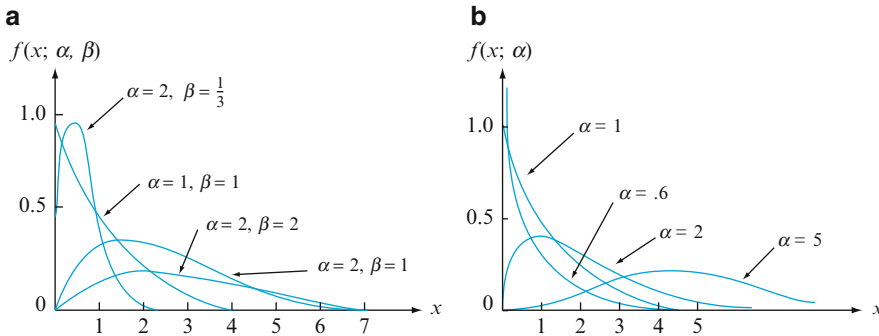


Fig. 3.25 (a) Gamma density curves; (b) standard gamma density curves

The mean and variance of a gamma random variable are

$$E(X) = \mu = \alpha\beta \quad \text{Var}(X) = \sigma^2 = \alpha\beta^2$$

These can be calculated directly from the gamma pdf using integration by parts, or by employing properties of the gamma function along with Expression (3.5); see Exercise 83. Notice these are consistent with the aforementioned mean and variance of the exponential distribution: with $\alpha = 1$ and $\beta = 1/\lambda$ we obtain $E(X) = 1(1/\lambda) = 1/\lambda$ and $\text{Var}(X) = 1(1/\lambda)^2 = 1/\lambda^2$.

In the special case where the shape parameter α is a positive integer, n , the gamma distribution is sometimes rewritten with the substitution $\lambda = 1/\beta$, and the resulting pdf is

$$f(x; n, 1/\lambda) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad x > 0$$

This is often called an *Erlang distribution*, and it plays a central role in the study of Poisson processes (again, see Sect. 7.5; notice that the $n = 1$ case of the Erlang distribution is actually the exponential pdf). In Chap. 4, it will be shown that the sum of n independent exponential rvs follows this Erlang distribution.

When X is a standard gamma rv, the cdf of X , which for $x > 0$ is

$$G(x; \alpha) = P(X \leq x) = \int_0^x \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-y} dy \tag{3.7}$$

is called the **incomplete gamma function**. (In mathematics literature, the incomplete gamma function sometimes refers to Eq. (3.7) without the denominator $\Gamma(\alpha)$ in the integrand.) In Appendix Table A.4, we present a small tabulation of $G(x; \alpha)$ for $\alpha = 1, 2, \dots, 10$ and $x = 1, 2, \dots, 15$.

Table 3.2 at the end of this section provides the Matlab and R commands related to the gamma cdf, which are illustrated in the following examples.

Example 3.27 Suppose the reaction time X (in seconds) of a randomly selected individual to a certain stimulus has a standard gamma distribution with $\alpha = 2$. Since X is continuous,

$$\begin{aligned} P(3 \leq X \leq 5) &= P(X \leq 5) - P(X \leq 3) = G(5; 2) - G(3; 2) \\ &= .960 - .801 = .159 \end{aligned}$$

This probability can be obtained in Matlab with `gamcdf(5, 2, 1) - gamcdf(3, 2, 1)` and in R with `pgamma(5, 2) - pgamma(3, 2)`.

The probability that the reaction time is more than 4 s is

$$P(X > 4) = 1 - P(X \leq 4) = 1 - G(4; 2) = 1 - .908 = .092 \quad \blacksquare$$

The incomplete gamma function can also be used to compute probabilities involving gamma distributions for any $\beta > 0$.

PROPOSITION

Let X have a gamma distribution with parameters α and β . Then for any $x > 0$, the cdf of X is given by

$$P(X \leq x) = G\left(\frac{x}{\beta}; \alpha\right),$$

the incomplete gamma function evaluated at x/β .

The proof is similar to that of Eq. (3.5).

Example 3.28 Suppose the survival time X in weeks of a randomly selected male mouse exposed to 240 rads of gamma radiation has, rather fittingly, a gamma distribution with $\alpha = 8$ and $\beta = 15$. (Data in *Survival Distributions: Reliability Applications in the Biomedical Services*, by A. J. Gross and V. Clark, suggest $\alpha \approx 8.5$ and $\beta \approx 13.3$.) The expected survival time is $E(X) = (8)(15) = 120$ weeks, whereas $SD(X) = \sqrt{(8)(15)^2} = \sqrt{1800} = 42.43$ weeks. The probability that a mouse survives between 60 and 120 weeks is

$$\begin{aligned} P(60 \leq X \leq 120) &= P(X \leq 120) - P(X \leq 60) \\ &= G(120/15; 8) - G(60/15; 8) \\ &= G(8; 8) - G(4; 8) = .547 - .051 = .496 \end{aligned}$$

In Matlab, the command `gamcdf(120, 8, 15) - gamcdf(60, 8, 15)` yields the desired probability; the corresponding R code is `pgamma(120, 8, 1/15) - pgamma(60, 8, 1/15)`.

The probability that a mouse survives at least 30 weeks is

$$P(X \geq 30) = 1 - P(X < 30) = 1 - P(X \leq 30) = 1 - G(30/15; 8) = .999 \quad \blacksquare$$

3.4.3 The Gamma MGF

The integral proposition earlier in this section makes it easy to determine the mean and variance of a gamma rv. However, the moment generating function of the gamma distribution — and, as a special case, of the exponential model — will prove critical in establishing some of the more advanced properties of these distributions in Chap. 4.

Proposition

The moment generating function of a gamma random variable is

$$M_X(t) = \frac{1}{(1 - \beta t)^\alpha} \quad t < 1/\beta$$

Proof By definition, the mgf is

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-(t+1/\beta)x} dx$$

Now use Expression (3.5): provided $-t + 1/\beta > 0$, i.e., $t < 1/\beta$,

$$\frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-(t+1/\beta)x} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \cdot \Gamma(\alpha) \left(\frac{1}{-t + 1/\beta} \right)^\alpha = \frac{1}{(1 - \beta t)^\alpha} \quad \blacksquare$$

The exponential mgf can then be determined with the substitution $\alpha = 1$, $\beta = 1/\lambda$:

$$M_X(t) = \frac{1}{(1 - (1/\lambda)t)^1} = \frac{\lambda}{\lambda - t} \quad t < \lambda$$

3.4.4 Gamma and Exponential Calculations with Software

Table 3.2 summarizes the syntax for gamma and exponential probability calculations in Matlab and R, which follows the pattern of the other distributions. In a sense, the exponential commands are redundant, since they are just a special case ($\alpha = 1$) of the gamma distribution.

Notice that Matlab and R parameterize the distributions differently: in Matlab, both the gamma and exponential functions require β (that is, $1/\lambda$) as the last input, whereas the R functions take as their last input the “rate” parameter $\lambda = 1/\beta$. So, for the gamma rv with parameters $\alpha = 8$ and $\beta = 15$ from Example 3.28, the probability $P(X \leq 30)$ would be evaluated as `gamcdf(30, 8, 15)` in Matlab but `pgamma(30, 8, 1/15)` in R. This inconsistency of gamma inputs can be remedied by using a name assignment in the last argument in R; specifically, `pgamma(30, 8, scale=15)` will instruct R to

Table 3.2 Matlab and R code for gamma and exponential calculations

	Gamma	Exponential
Function:	<code>cdf</code>	<code>cdf</code>
Notation:	$G(x/\beta; \alpha)$	$F(x; \lambda) = 1 - e^{-\lambda x}$
Matlab:	<code>gamcdf(x, α, β)</code>	<code>expcdf(x, 1/λ)</code>
R:	<code>pgamma(x, α, 1/β)</code>	<code>pexp(x, λ)</code>

use $\beta = 15$ in its gamma probability calculation and produce the same answer as the previous expressions. Interestingly, as of this writing the same option does not exist in the `pexp` function.

To graph gamma or exponential distributions, one can request their pdfs by replacing `cdf` with `pdf` (in Matlab) or the leading letter `p` with `d` (in R). To find quantiles of either of these distributions, the appropriate replacements are `inv` and `q`, respectively. For example, the 75th percentile of the gamma distribution from Example 3.28 can be determined with `gamainv(.75, 8, 15)` in Matlab or `qgamma(.75, 8, scale=15)` in R (both give 145.2665 weeks).

3.4.5 Exercises: Section 3.4 (71–83)

71. Let X = the time between two successive arrivals at the drive-up window of a local bank. If X has an exponential distribution with $\lambda = 1$, compute the following:
 - (a) The expected time between two successive arrivals
 - (b) The standard deviation of the time between successive arrivals
 - (c) $P(X \leq 4)$
 - (d) $P(2 \leq X \leq 5)$
72. Let X denote the distance (m) that an animal moves from its birth site to the first territorial vacancy it encounters. Suppose that for banner-tailed kangaroo rats, X has an exponential distribution with parameter $\lambda = .01386$ (as suggested in the article “Competition and Dispersal from Multiple Nests,” *Ecology*, 1997: 873–883).
 - (a) What is the probability that the distance is at most 100 m? At most 200 m? Between 100 and 200 m?
 - (b) What is the probability that distance exceeds the mean distance by more than 2 standard deviations?
 - (c) What is the value of the median distance?
73. In studies of anticancer drugs it was found that if mice are injected with cancer cells, the survival time can be modeled with the exponential distribution. Without treatment the expected survival time was 10 h. What is the probability that
 - (a) A randomly selected mouse will survive at least 8 h? At most 12 h? Between 8 and 12 h?
 - (b) The survival time of a mouse exceeds the mean value by more than 2 standard deviations? More than 3 standard deviations?
74. Data collected at Toronto Pearson International Airport suggests that an exponential distribution with mean value 2.725 h is a good model for rainfall duration (*Urban Stormwater Management Planning with Analytical Probabilistic Models*, 2000, p.69).
 - (a) What is the probability that the duration of a particular rainfall event at this location is at least 2 h? At most 3 h? Between 2 and 3 h?
 - (b) What is the probability that rainfall duration exceeds the mean value by more than 2 standard deviations? What is the probability that it is less than the mean value by more than one standard deviation?
75. Evaluate the following:
 - (a) $\Gamma(6)$
 - (b) $\Gamma(5/2)$
 - (c) $G(4; 5)$ (the incomplete gamma function)
 - (d) $G(5; 4)$
 - (e) $G(0; 4)$

76. Let X have a standard gamma distribution with $\alpha = 7$. Evaluate the following:
- $P(X \leq 5)$
 - $P(X < 5)$
 - $P(X > 8)$
 - $P(3 \leq X \leq 8)$
 - $P(3 < X < 8)$
 - $P(X < 4 \text{ or } X > 6)$
77. Suppose that when a type of transistor is subjected to an accelerated life test, the lifetime X (in weeks) has a gamma distribution with mean 24 weeks and standard deviation 12 weeks.
- What is the probability that a transistor will last between 12 and 24 weeks?
 - What is the probability that a transistor will last at most 24 weeks? Is the median of the lifetime distribution less than 24? Why or why not?
 - What is the 99th percentile of the lifetime distribution?
 - Suppose the test will actually be terminated after t weeks. What value of t is such that only .5% of all transistors would still be operating at termination?
78. The two-parameter gamma distribution can be generalized by introducing a third parameter γ , called a *threshold* or *location* parameter: replace x in Eq. (3.6) by $x - \gamma$ and $x \geq 0$ by $x \geq \gamma$. This amounts to shifting the density curves in Fig. 3.25 so that they begin their ascent or descent at γ rather than 0. The article “Bivariate Flood Frequency Analysis with Historical Information Based on Copulas” (*J. of Hydrologic Engr.*, 2013: 1018–1030) employs this distribution to model $X = 3$ -day flood volume (10^8 m^3). Suppose that values of the parameters are $\alpha = 12$, $\beta = 7$, $\gamma = 40$ (very close to estimates in the cited article based on past data).
- What are the mean value and standard deviation of X ?
 - What is the probability that flood volume is between 100 and 150?
 - What is the probability that flood volume exceeds its mean value by more than one standard deviation?
 - What is the 95th percentile of the flood volume distribution?
79. If X has an exponential distribution with parameter λ , derive an expression for the $(100p)$ th percentile of the distribution. Then specialize to obtain the median.
80. A system consists of five identical components connected in series as shown:



As soon as one component fails, the entire system will fail. Suppose each component has a lifetime that is exponentially distributed with $\lambda = .01$ and that components fail independently of one another. Define events $A_i = \{i\text{th component lasts at least } t \text{ hours}\}$, $i = 1, \dots, 5$, so that the A_i s are independent events. Let X = the time at which the system fails—that is, the shortest (minimum) lifetime among the five components.

- The event $\{X \geq t\}$ is equivalent to what event involving A_1, \dots, A_5 ?
 - Using the independence of the five A_i s, compute $P(X \geq t)$. Then obtain $F(t) = P(X \leq t)$ and the pdf of X . What type of distribution does X have?
 - Suppose there are n components, each having exponential lifetime with parameter λ . What type of distribution does X have?
81. Based on an analysis of sample data, the article “Pedestrians’ Crossing Behaviors and Safety at Unmarked Roadways in China” (*Accident Analysis and Prevention*, 2011: 1927–1936) proposed the pdf $f(x) = .15e^{-.15(x-1)}$ when $x \geq 1$ as a model for the distribution of $X =$ time (sec) spent at the median line. This is an example of a *shifted exponential* distribution, i.e., an exponential model beginning at an x -value other than 0.

- (a) What is the probability that waiting time is at most 5 s? More than 5 s?
- (b) What is the probability that waiting time is between 2 and 5 s?
- (c) What is the mean waiting time?
- (d) What is the standard deviation of waiting times?

[Hint: For (c) and (d), you can either use integration or write $X = Y + 1$, where Y has an exponential distribution with parameter $\lambda = .15$. Then, apply rescaling properties of mean and standard deviation.]

82. The *double exponential distribution* has pdf

$$f(x) = .5\lambda e^{-\lambda|x|} \quad \text{for } -\infty < x < \infty$$

The article “Microwave Observations of Daily Antarctic Sea-Ice Edge Expansion and Contribution Rates” (*IEEE Geosci. and Remote Sensing Letters*, 2006: 54-58) states that “the distribution of the daily sea-ice advance/retreat from each sensor is similar and is approximately double exponential.” The standard deviation is given as 40.9 km.

- (a) What is the mean of a random variable with pdf $f(x)$? [Hint: Draw a picture of the density curve.]
 - (b) What is the value of the parameter λ when $\sigma_X = 40.9$?
 - (c) What is the probability that the extent of daily sea-ice change is within 1 standard deviation of the mean value?
83. (a) Find the mean and variance of the gamma distribution using integration and Expression (3.5) to obtain $E(X)$ and $E(X^2)$.
- (b) Use the gamma mgf to find the mean and variance.

3.5 Other Continuous Distributions

The normal, gamma (including exponential), and uniform families of distributions provide a wide variety of probability models for continuous variables, but there are many practical situations in which no member of these families fits a set of observed data very well. Statisticians and other investigators have developed other families of distributions that are often appropriate in practice.

3.5.1 The Weibull Distribution

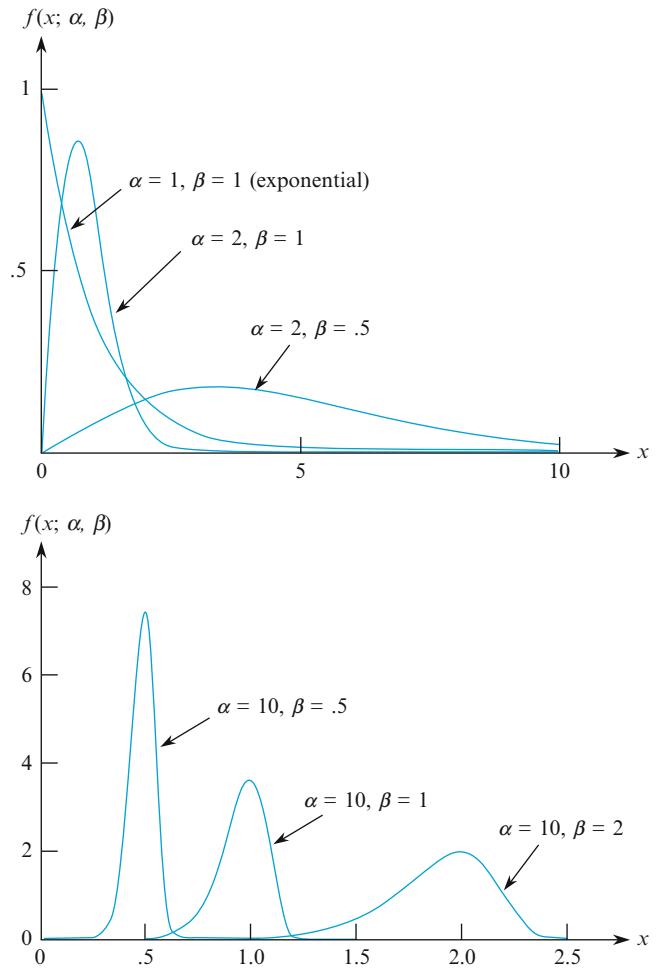
The family of Weibull distributions was introduced by the Swedish physicist Waloddi Weibull in 1939; his 1951 article “A Statistical Distribution Function of Wide Applicability” (*J. Appl. Mech.*, 18: 293–297) discusses a number of applications.

DEFINITION

A random variable X is said to have a **Weibull distribution** with parameters α and β ($\alpha > 0, \beta > 0$) if the pdf of X is

$$f(x; \alpha, \beta) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (3.8)$$

Fig. 3.26 Weibull density curves



In some situations there are theoretical justifications for the appropriateness of the Weibull distribution, but in many applications $f(x; \alpha, \beta)$ simply provides a good fit to observed data for particular values of α and β . When $\alpha = 1$, the pdf reduces to the exponential distribution (with $\lambda = 1/\beta$), so the exponential distribution is a special case of both the gamma and Weibull distributions. However, there are gamma distributions that are not Weibull distributions and vice versa, so one family is not a subset of the other. Both α and β can be varied to obtain a number of different distributional shapes, as illustrated in Fig. 3.26. Note that β is a scale parameter, so different values stretch or compress the graph in the x -direction; α is referred to as a shape parameter. Integrating to obtain $E(X)$ and $E(X^2)$ yields

$$\mu = \beta \Gamma\left(1 + \frac{1}{\alpha}\right) \quad \sigma^2 = \beta^2 \left\{ \Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right) \right]^2 \right\}$$

The computation of μ and σ^2 thus necessitate using the gamma function from Sect. 3.4. (The moment generating function of the Weibull distribution is very complicated, and so we do not include it here.) On the other hand, the integration $\int_0^x f(y; \alpha, \beta) dy$ is easily carried out to obtain the cdf of X :

$$F(x; \alpha, \beta) = \begin{cases} 0 & x < 0 \\ 1 - e^{-(x/\beta)^\alpha} & x \geq 0 \end{cases} \quad (3.9)$$

Example 3.29 In recent years the Weibull distribution has been used to model engine emissions of various pollutants. Let X denote the amount of NO_x emission (g/gal) from a randomly selected four-stroke engine of a certain type, and suppose that X has a Weibull distribution with $\alpha = 2$ and $\beta = 10$ (suggested by information in the article “Quantification of Variability and Uncertainty in Lawn and Garden Equipment NO_x and Total Hydrocarbon Emission Factors,” *J. Air Waste Manag. Assoc.*, 2002: 435–448). The corresponding density curve looks exactly like the one in Fig. 3.26 for $\alpha = 2, \beta = 1$ except that now the values 50 and 100 replace 5 and 10 on the horizontal axis (because β is a “scale parameter”). Then

$$P(X \leq 10) = F(10; 2, 10) = 1 - e^{-(10/10)^2} = 1 - e^{-1} = .632$$

Similarly, $P(X \leq 25) = .998$, so the distribution is almost entirely concentrated on values between 0 g/gal and 25 g/gal. The value c which separates the 5% of all engines having the largest amounts of NO_x emissions from the remaining 95%, satisfies

$$.95 = F(c; 2, 10) = 1 - e^{-(c/10)^2}$$

Isolating the exponential term on one side, taking logarithms, and solving the resulting equation gives $c \approx 17.3$ g/gal as the 95th percentile of the emission distribution. ■

Frequently, in practical situations, a Weibull model may be reasonable except that the smallest possible X value may be some value γ other than zero (Exercise 78 considered this for a gamma model). The quantity γ can then be regarded as a third parameter of the distribution, which is what Weibull did in his original work. For, say, $\gamma = 3$, all curves in Fig. 3.26 would be shifted 3 units to the right. This is equivalent to saying that $X - \gamma$ has the pdf Eq. (3.8), so that the cdf of X is obtained by replacing x in Eq. (3.9) by $x - \gamma$.

Example 3.30 An understanding of the volumetric properties of asphalt is important in designing mixtures that will result in high-durability pavement. The article “Is a Normal Distribution the Most Appropriate Statistical Distribution for Volumetric Properties in Asphalt Mixtures” (*J. of Testing and Evaluation*, Sept. 2009: 1–11) used the analysis of some sample data to recommend that for a particular mixture, $X =$ air void volume (%) be modeled with a three-parameter Weibull distribution. Suppose the values of the parameters are $\gamma = 4, \alpha = 1.3$, and $\beta = .8$, which are quite close to estimates given in the article.

For $x \geq 4$, the cumulative distribution function is

$$F(x; \alpha, \beta, \gamma) = F(x; 1.3, .8, 4) = 1 - e^{-[(x-4)/.8]^{1.3}}$$

The probability that the air void volume of a specimen is between 5% and 6% is

$$\begin{aligned} P(5 \leq X \leq 6) &= F(6; 1.3, .8, 4) - F(5; 1.3, .8, 4) = e^{-[(5-4)/.8]^{1.3}} - e^{-[(6-4)/.8]^{1.3}} \\ &= .263 - .037 = .226 \end{aligned}$$

Figure 3.27 shows a graph of the corresponding Weibull density function, in which the shaded area corresponds to the probability just calculated.

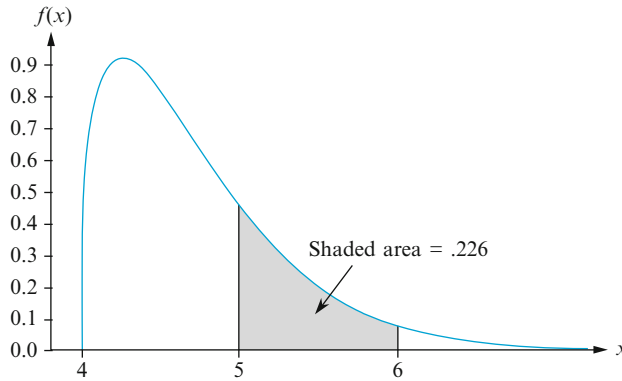


Fig. 3.27 Weibull density curve with threshold = 4, shape = 1.3, scale = .8

3.5.2 The Lognormal Distribution

Lognormal distributions have been used extensively in engineering, medicine, and more recently, finance.

DEFINITION

A nonnegative rv X is said to have a **lognormal distribution** if the rv $Y = \ln(X)$ has a normal distribution. The resulting pdf of a lognormal rv when $\ln(X)$ is normally distributed with parameters μ and σ is

$$f(x; \mu, \sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} e^{-[\ln(x)-\mu]^2/(2\sigma^2)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Be careful here: the parameters μ and σ are not the mean and standard deviation of X but of $\ln(X)$. The mean and variance of a lognormal random variable can be shown to be

$$E(X) = e^{\mu+\sigma^2/2} \quad \text{Var}(X) = e^{2\mu+2\sigma^2} \cdot (e^{\sigma^2} - 1)$$

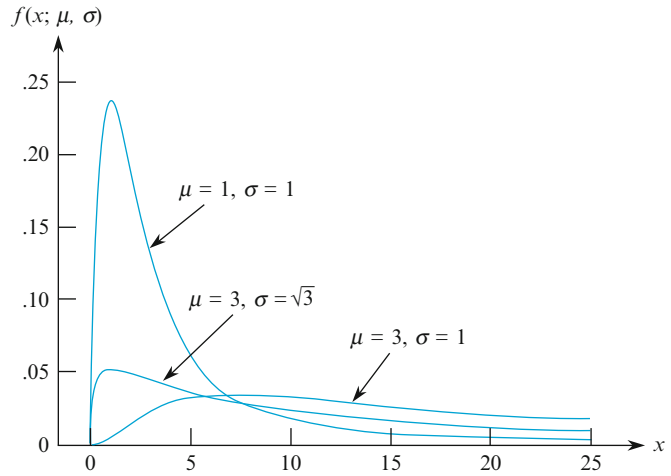
In Chap. 4, we will present a theoretical justification for this distribution in connection with the Central Limit Theorem, but as with other distributions, the lognormal can be used as a model even in the absence of such justification. Figure 3.28 illustrates graphs of the lognormal pdf; although a normal curve is symmetric, a lognormal curve has a positive skew.

Because $\ln(X)$ has a normal distribution, the cdf of X can be expressed in terms of the cdf $\Phi(z)$ of a standard normal rv Z . For $x \geq 0$,

$$\begin{aligned} F(x; \mu, \sigma) &= P(X \leq x) = P[\ln(X) \leq \ln(x)] = P\left[\frac{\ln(X) - \mu}{\sigma} \leq \frac{\ln(x) - \mu}{\sigma}\right] \\ &= P\left[Z \leq \frac{\ln(x) - \mu}{\sigma}\right] = \Phi\left[\frac{\ln(x) - \mu}{\sigma}\right] \end{aligned} \tag{3.10}$$

Differentiating $F(x; \mu, \sigma)$ with respect to x gives the pdf $f(x; \mu, \sigma)$ above.

Fig. 3.28 Lognormal density curves



Example 3.31 According to the article “Predictive Model for Pitting Corrosion in Buried Oil and Gas Pipelines” (*Corrosion*, 2009: 332–342), the lognormal distribution has been reported as the best option for describing the distribution of maximum pit depth data from cast iron pipes in soil. The authors suggest that a lognormal distribution with $\mu = .353$ and $\sigma = .754$ is appropriate for maximum pit depth (mm) of buried pipelines. For this distribution, the mean value and variance of pit depth are

$$E(X) = e^{.353 + (.754)^2/2} = e^{.6383} = 1.893$$

$$\text{Var}(X) = e^{2(.353) + (.754)^2} \cdot (e^{(.754)^2} - 1) = (3.57697)(.765645) = 2.7387$$

The probability that maximum pit depth is between 1 and 2 mm is

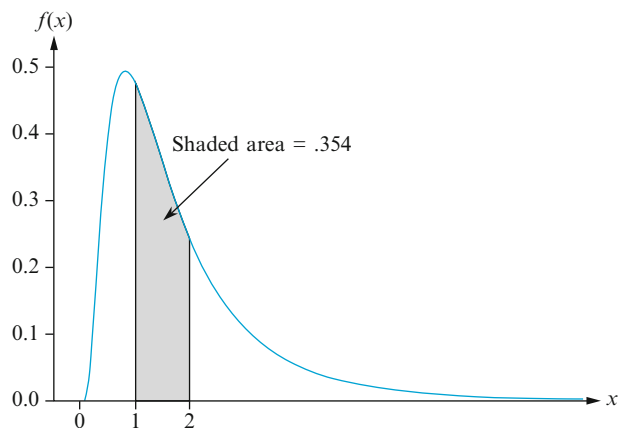
$$P(1 \leq X \leq 2) = P(\ln(1) \leq \ln(X) \leq \ln(2)) = P(0 \leq \ln(X) \leq .693)$$

$$= P\left(\frac{0 - .353}{.754}\right) \leq Z \leq \left(\frac{.693 - .353}{.754}\right) = \Phi(.45) - \Phi(-.47) = .354$$

This probability is illustrated in Fig. 3.29.

What value c is such that only 1% of all specimens have a maximum pit depth exceeding c ? The desired value satisfies

Fig. 3.29 Lognormal density curve with $\mu = .353$ and $\sigma = .754$



$$.99 = P(X \leq c) = \Phi\left(\frac{\ln(c) - .353}{.754}\right)$$

Appendix Table A.3 indicates that $z = 2.33$ is the 99th percentile of the standard normal distribution, which implies that

$$\frac{\ln(c) - .353}{.754} = 2.33$$

Solving for c gives $\ln(c) = 2.1098$ and $c = 8.247$. Thus 8.247 mm is the 99th percentile of the maximum pit depth distribution. ■

As with the Weibull distribution, a third parameter γ can be introduced so that the distribution has positive density for $x > \gamma$ rather than for $x > 0$.

3.5.3 The Beta Distribution

All families of continuous distributions discussed so far except for the uniform distribution have positive density over an infinite interval (although typically the density function decreases rapidly to zero beyond a few standard deviations from the mean). The beta distribution provides positive density only for X in an interval of finite length.

DEFINITION

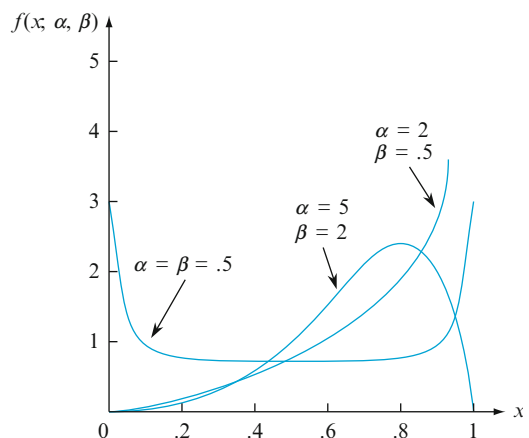
A random variable X is said to have a **beta distribution** with parameters α, β (both positive), A , and B if the pdf of X is

$$f(x; \alpha, \beta, A, B) = \begin{cases} \frac{1}{B-A} \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \left(\frac{x-A}{B-A}\right)^{\alpha-1} \left(\frac{B-x}{B-A}\right)^{\beta-1} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

The case $A = 0, B = 1$ gives the **standard beta distribution**.

Figure 3.30 illustrates several standard beta pdfs. Graphs of the general pdf are similar, except they are shifted and then stretched or compressed to fit over $[A, B]$. Unless α and β are integers, integration

Fig. 3.30 Standard beta density curves



of the pdf to calculate probabilities is difficult, so either a table of the incomplete beta function or software is generally used.

The standard beta distribution is commonly used to model variation in the proportion or percentage of a quantity occurring in different samples, such as the proportion of a 24-h day that an individual is asleep or the proportion of a certain element in a chemical compound.

The mean and variance of X are

$$\mu = A + (B - A) \cdot \frac{\alpha}{\alpha + \beta} \quad \sigma^2 = \frac{(B - A)^2 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

The moment generating function of the beta distribution is too complicated to be useful.

Example 3.32 Project managers often use a method labeled PERT—for program evaluation and review technique—to coordinate the various activities making up a large project. (One successful application was in the construction of the *Apollo* spacecraft.) A standard assumption in PERT analysis is that the time necessary to complete any particular activity once it has been started has a beta distribution with A = the optimistic time (if everything goes well) and B = the pessimistic time (if everything goes badly). Suppose that in constructing a single-family house, the time X (in days) necessary for laying the foundation has a beta distribution with $A = 2$, $B = 5$, $\alpha = 2$, and $\beta = 3$. Then $\alpha/(\alpha + \beta) = .4$, so $E(X) = 2 + (3)(.4) = 3.2$. For these values of α and β , the pdf of X is a simple polynomial function. The probability that it takes at most 3 days to lay the foundation is

$$\begin{aligned} P(X \leq 3) &= \int_2^3 \frac{1}{3} \cdot \frac{4!}{1! \cdot 2!} \left(\frac{x-2}{3}\right) \left(\frac{5-x}{3}\right)^2 dx \\ &= \frac{4}{27} \int_2^3 (x-2)(5-x)^2 dx = \frac{4}{27} \cdot \frac{11}{4} = \frac{11}{27} = .407 \quad \blacksquare \end{aligned}$$

Software, including Matlab and R, can be used to perform probability calculations for the Weibull, lognormal, and beta distributions. Interested readers should consult the help menus in those packages.

3.5.4 Exercises: Section 3.5 (84–100)

84. The lifetime X (in hundreds of hours) of a type of transistor has a Weibull distribution with parameters $\alpha = 2$ and $\beta = 3$. Compute the following:
- $E(X)$ and $\text{Var}(X)$
 - $P(X \leq 6)$
 - $P(1.5 \leq X \leq 6)$
- (This Weibull distribution is suggested as a model for time in service in “On the Assessment of Equipment Reliability: Trading Data Collection Costs for Precision,” *J. Engrg. Manuf.*, 1991: 105–109.)
85. The authors of the article “A Probabilistic Insulation Life Model for Combined Thermal-Electrical Stresses” (*IEEE Trans. Electr. Insul.*, 1985: 519–522) state that “the Weibull distribution is widely used in statistical problems relating to aging of solid insulating materials subjected to aging and stress.” They propose the use of the distribution as a model for time (in hours) to failure of solid insulating specimens subjected to ac voltage. The values of the parameters depend on the voltage and temperature; suppose $\alpha = 2.5$ and $\beta = 200$ (values suggested by data in the article).

- (a) What is the probability that a specimen's lifetime is at most 250? Less than 250? More than 300?
- (b) What is the probability that a specimen's lifetime is between 100 and 250?
- (c) What value is such that exactly 50% of all specimens have lifetimes exceeding that value?
86. Let X = the time (in 10^{-1} weeks) from shipment of a defective product until the customer returns the product. Suppose that the minimum return time is $\gamma = 3.5$ and that the excess $X - 3.5$ over the minimum has a Weibull distribution with parameters $\alpha = 2$ and $\beta = 1.5$ (see the article "Practical Applications of the Weibull Distribution," *Indust. Qual. Control*, 1964: 71–78).
- (a) What is the cdf of X ?
- (b) What are the expected return time and variance of return time? [*Hint*: First obtain both $E(X - 3.5)$ and $\text{Var}(X - 3.5)$.]
- (c) Compute $P(X > 5)$.
- (d) Compute $P(5 \leq X \leq 8)$.
87. Let X have a Weibull distribution. Verify that $\mu = \beta\Gamma(1 + 1/\alpha)$. [*Hint*: In the integral for $E(X)$, make the change of variable $y = (x/\beta)^\alpha$, so that $x = \beta y^{1/\alpha}$.]
88. (a) In Exercise 84, what is the median lifetime of such tubes? [*Hint*: Use Expression (3.9).]
- (b) In Exercise 86, what is the median return time?
- (c) If X has a Weibull distribution with the cdf from Expression (3.9), obtain a general expression for the $(100p)$ th percentile of the distribution.
- (d) In Exercise 86, the company wants to refuse to accept returns after t weeks. For what value of t will only 10% of all returns be refused?
89. Let X denote the ultimate tensile strength (ksi) at -200° of a randomly selected steel specimen of a certain type that exhibits "cold brittleness" at low temperatures. Suppose that X has a Weibull distribution with $\alpha = 20$ and $\beta = 100$.
- (a) What is the probability that X is at most 105 ksi?
- (b) If specimen after specimen is selected, what is the long-run proportion having strength values between 100 and 105 ksi?
- (c) What is the median of the strength distribution?
90. The article "On Assessing the Accuracy of Offshore Wind Turbine Reliability-Based Design Loads from the Environmental Contour Method" (*Intl. J. of Offshore and Polar Engr.*, 2005: 132–140) proposes the Weibull distribution with $\alpha = 1.817$ and $\beta = .863$ as a model for 1-h significant wave height (m) at a certain site.
- (a) What is the probability that wave height is at most .5 m?
- (b) What is the probability that wave height exceeds its mean value by more than one standard deviation?
- (c) What is the median of the wave-height distribution?
- (d) For $0 < p < 1$, give a general expression for the $100p$ th percentile of the wave-height distribution.
91. Nonpoint source loads are chemical masses that travel to the main stem of a river and its tributaries in flows that are distributed over relatively long stream reaches, in contrast to those that enter at well-defined and regulated points. The article "Assessing Uncertainty in Mass Balance Calculation of River Nonpoint Source Loads" (*J. of Envir. Engr.*, 2008: 247–258) suggested that for a certain time period and location, nonpoint source load of total dissolved solids could be modeled with a lognormal distribution having mean value 10,281 kg/day/km and a coefficient of variation $\text{CV} = .40$ ($\text{CV} = \sigma_X/\mu_X$).
- (a) What are the mean value and standard deviation of $\ln(X)$?
- (b) What is the probability that X is at most 15,000 kg/day/km?

- (c) What is the probability that X exceeds its mean value, and why is this probability not .5?
(d) Is 17,000 the 95th percentile of the distribution?
92. The authors of the article “Study on the Life Distribution of Microdrills” (*J. of Engr. Manufacture*, 2002: 301-305) suggested that a reasonable probability model for drill lifetime was a lognormal distribution with $\mu = 4.5$ and $\sigma = .8$.
- (a) What are the mean value and standard deviation of lifetime?
(b) What is the probability that lifetime is at most 100?
(c) What is the probability that lifetime is at least 200? Greater than 200?
93. Use Equation (3.10) to write a formula for the median η of the lognormal distribution. What is the median for the load distribution of Exercise 91?
94. As in the case of the Weibull distribution, the lognormal distribution can be modified by the introduction of a third parameter γ such that the pdf is shifted to be positive only for $x > \gamma$. The article cited in Exercise 46 suggested that a shifted lognormal distribution with shift = 1.0, mean value = 2.16, and standard deviation = 1.03 would be an appropriate model for the rv $X = \text{maximum-to-average depth ratio}$ of a corrosion defect in pressurized steel.
- (a) What are the values of μ and σ for the proposed distribution?
(b) What is the probability that depth ratio exceeds 2?
(c) What is the median of the depth ratio distribution?
(d) What is the 99th percentile of the depth ratio distribution?
95. Sales delay is the elapsed time between the manufacture of a product and its sale. According to the article “Warranty Claims Data Analysis Considering Sales Delay” (*Quality and Reliability Engr. Intl.*, 2013: 113–123), it is quite common for investigators to model sales delay using a lognormal distribution. For a particular product, the cited article proposes this distribution with parameter values $\mu = 2.05$ and $\sigma^2 = .06$ (here the unit for delay is months).
- (a) What are the variance and standard deviation of delay time?
(b) What is the probability that delay time exceeds 12 months?
(c) What is the probability that delay time is within one standard deviation of its mean value?
(d) What is the median of the delay time distribution?
(e) What is the 99th percentile of the delay time distribution?
(f) Among 10 randomly selected such items, how many would you expect to have a delay time exceeding 8 months?
96. The article “The Statistics of Phytotoxic Air Pollutants” (*J. Roy. Statist Soc.*, 1989: 183–198) suggests the lognormal distribution as a model for SO_2 concentration above a forest. Suppose the parameter values are $\mu = 1.9$ and $\sigma = .9$.
- (a) What are the mean value and standard deviation of concentration?
(b) What is the probability that concentration is at most 10? Between 5 and 10?
97. What condition on α and β is necessary for the standard beta pdf to be symmetric?
98. Suppose the proportion X of surface area in a randomly selected quadrat that is covered by a certain plant has a standard beta distribution with $\alpha = 5$ and $\beta = 2$.
- (a) Compute $E(X)$ and $\text{Var}(X)$.
(b) Compute $P(X \leq .2)$.
(c) Compute $P(.2 \leq X \leq .4)$.
(d) What is the expected proportion of the sampling region not covered by the plant?
99. Let X have a standard beta density with parameters α and β .
- (a) Verify the formula for $E(X)$ given in the section.
(b) Compute $E[(1 - X)^m]$. If X represents the proportion of a substance consisting of a particular ingredient, what is the expected proportion that does not consist of this ingredient?

100. Stress is applied to a 20-in. steel bar that is clamped in a fixed position at each end. Let Y = the distance from the left end at which the bar snaps. Suppose $Y/20$ has a standard beta distribution with $E(Y) = 10$ and $\text{Var}(Y) = 100/7$.
- What are the parameters of the relevant standard beta distribution?
 - Compute $P(8 \leq Y \leq 12)$.
 - Compute the probability that the bar snaps more than 2 in. from where you expect it to snap.

3.6 Probability Plots

An investigator will often have obtained a numerical sample consisting of n observations and wish to know whether it is plausible that this sample came from a population distribution of some particular type (e.g., from a normal distribution). For one thing, many formal procedures from statistical inference (Chap. 5) are based on the assumption that the population distribution is of a specified type. The use of such a procedure is inappropriate if the actual underlying probability distribution differs greatly from the assumed type. Additionally, understanding the underlying distribution can sometimes give insight into the physical mechanisms involved in generating the data. An effective way to check a distributional assumption is to construct what is called a **probability plot**. The basis for our construction is a comparison between percentiles of the sample data and the corresponding percentiles of the assumed underlying distribution.

3.6.1 Sample Percentiles

The details involved in constructing probability plots differ a bit from source to source. Roughly speaking, sample percentiles are defined in the same way that percentiles of a population distribution are defined. The sample 50th percentile (i.e., the sample median) should separate the smallest 50% of the sample from the largest 50%, the sample 90th percentile should be such that 90% of the sample lies below that value and 10% lies above, and so on. Unfortunately, we run into problems when we actually try to compute the sample percentiles for a particular sample of n observations. If, for example, $n = 10$, then we can split off 20% or 30% of the data, but there is no value that will split off exactly 23% of these ten observations. To proceed further, we need an operational definition of sample percentiles (this is one place where different people and different software packages do slightly different things).

Statistical convention states that when n is odd, the sample median is the middle value in the ordered list of sample observations, for example, the sixth-largest value when $n = 11$. This amounts to regarding the middle observation as being half in the lower half of the data and half in the upper half. Similarly, suppose $n = 10$. Then if we call the third-smallest value the 25th percentile, we are regarding that value as being half in the lower group (consisting of the two smallest observations) and half in the upper group (the seven largest observations). This leads to the following general definition of sample percentiles.

DEFINITION

Order the n sample observations from smallest to largest. Then the i th-smallest observation in the list is taken to be the **sample $[100(i - .5)/n]$ th percentile**.

For example, if $n = 10$, the percentages corresponding to the ordered sample observations are $100(1 - .5)/10 = 5%$, $100(2 - .5)/10 = 15%$, $25%$, \dots , and $100(10 - .5)/10 = 95%$. That is, the smallest observation is designated the sample 5th percentile, the next-smallest value the sample 15th percentile, and so on. All other percentiles could then be determined by interpolation, e.g., the sample 10th percentile would then be halfway between the 5th percentile (smallest sample observation) and the 15th percentile (second smallest observation) of the $n = 10$ values. For the purposes of a probability plot, such interpolation will not be necessary, because a probability plot will be based only on the percentages $100(i - .5)/n$ corresponding to the n sample observations.

3.6.2 A Probability Plot

We now wish to determine whether our sample data could plausibly have come from some particular population distribution (e.g., a normal distribution with $\mu = 10$ and $\sigma = 3$). If the sample was actually selected from the specified distribution, the sample percentiles (ordered sample observations) should be reasonably close to the corresponding population distribution percentiles. That is, for $i = 1, 2, \dots, n$ there should be reasonable agreement between the i th-smallest sample observation and the theoretical $[100(i - .5)/n]$ th percentile for the specified distribution. Consider the (sample percentile, population percentile) pairs—that is, the pairs

$$\left(\begin{array}{l} i\text{th smallest sample} \\ \text{observation} \end{array}, \begin{array}{l} [100(i - .5)/n]\text{th percentile} \\ \text{of the population distribution} \end{array} \right)$$

for $i = 1, \dots, n$. Each such pair can be plotted as a point on a two-dimensional coordinate system. If the sample percentiles are close to the corresponding population distribution percentiles, the first number in each pair will be roughly equal to the second number, and the plotted points will then fall close to a 45° line. Substantial deviations of the plotted points from a 45° line suggest that the assumed distribution might be wrong.

Example 3.33 The value of a physical constant is known to an experimenter. The experimenter makes $n = 10$ independent measurements of this value using a measurement device and records the resulting measurement errors (error = observed value – true value). These observations appear in the accompanying table.

Percentage	5	15	25	35	45
Sample observation	–1.91	–1.25	–.75	–.53	.20
z percentile	–1.645	–1.037	–.675	–.385	–.126
Percentage	55	65	75	85	95
Sample observation	.35	.72	.87	1.40	1.56
z percentile	.126	.385	.675	1.037	1.645

Is it plausible that the random variable *measurement error* has a standard normal distribution? The needed standard normal (z) percentiles are also displayed in the table and were determined as follows:

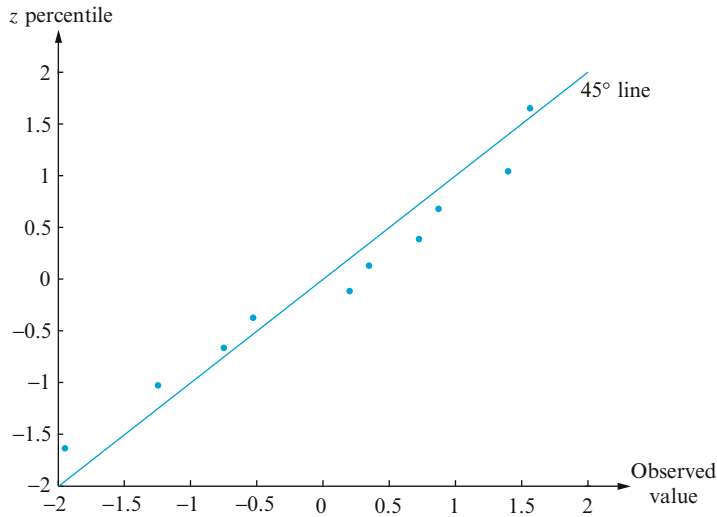


Fig. 3.31 Plots of pairs (observed value, z percentile) for the data of Example 3.33

the 5th percentile of the distribution under consideration, $N(0,1)$, is given by $\Phi(z) = .05$. From software or Appendix Table A.3, the solution is roughly $z = -1.645$. The other nine population (z) percentiles were found in a similar fashion.

Thus the points in the probability plot are $(-1.91, -1.645)$, $(-1.25, -1.037)$, \dots , and $(1.56, 1.645)$. Figure 3.31 shows the resulting plot. Although the points deviate a bit from the 45° line, the predominant impression is that this line fits the points reasonably well. The plot suggests that the standard normal distribution is a realistic probability model for measurement error.

An investigator is typically not interested in knowing whether a completely specified probability distribution, such as the normal distribution with $\mu = 0$ and $\sigma = 1$ or the exponential distribution with $\lambda = .1$, is a plausible model for the population distribution from which the sample was selected. Instead, the investigator will want to know whether *some* member of a family of probability distributions specifies a plausible model—the family of normal distributions, the family of exponential distributions, the family of Weibull distributions, and so on. The values of the parameters of a distribution are usually not specified at the outset. If the family of Weibull distributions is under consideration as a model for lifetime data, the issue is whether there are *any* values of the parameters α and β for which the corresponding Weibull distribution gives a good fit to the data. Fortunately, it is almost always the case that just one probability plot will suffice for assessing the plausibility of an entire family. If the plot deviates substantially from a straight line, but not necessarily the 45° line, no member of the family is plausible.

To see why, let's focus on a plot for checking normality. As mentioned earlier, such a plot can be very useful in applied work because many formal statistical procedures are appropriate (i.e., give accurate inferences) only when the population distribution is at least approximately normal. These procedures should generally not be used if a normal probability plot shows a very pronounced departure from linearity. The key to constructing an omnibus normal probability plot is the relationship between standard normal (z) percentiles and those for any other normal distribution, which was presented in Sect. 3.3:

$$\text{percentile for a } N(\mu, \sigma) \text{ distribution} = \mu + \sigma \cdot (\text{corresponding } z \text{ percentile})$$

If each sample observation were exactly equal to the corresponding $N(\mu, \sigma)$ percentile, then the pairs (observation, $\mu + \sigma \cdot [z \text{ percentile}]$) would fall on the 45° line, $y = x$. But since $\mu + \sigma z$ is itself a linear function, the pairs (observation, z percentile) would also fall on a straight line, just not the line with slope 1 and y -intercept 0. (The latter pairs would pass through the line $z = x/\sigma - \mu/\sigma$, but the equation itself isn't important.)

DEFINITION

A plot of the n pairs

$$(i\text{th-smallest observation, } [100(i - .5)/n]\text{th } z \text{ percentile})$$

on a two-dimensional coordinate system is called a **normal probability plot**. If the sample observations are in fact drawn from a normal distribution then the points should fall close to a straight line (although not necessarily a 45° line). Thus a plot for which the points fall close to *some* straight line suggests that the assumption of a normal population distribution is plausible.

Example 3.34 The accompanying sample consisting of $n = 20$ observations on dielectric breakdown voltage of a piece of epoxy resin appeared in the article “Maximum Likelihood Estimation in the 3-Parameter Weibull Distribution” (*IEEE Trans. Dielectrics Electr. Insul.*, 1996: 43–55). Values of $(i - .5)/n$ for which z percentiles are needed are $(1 - .5)/20 = .025$, $(2 - .5)/20 = .075$, . . . , and $.975$.

Observation	24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
z percentile	-1.96	-1.44	-1.15	-.93	-.76	-.60	-.45	-.32	-.19	-.06
Observation	27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88
z percentile	.06	.19	.32	.45	.60	.76	.93	1.15	1.44	1.96

Figure 3.32 shows the resulting normal probability plot. The pattern in the plot is quite straight, indicating it is plausible that the population distribution of dielectric breakdown voltage is normal.

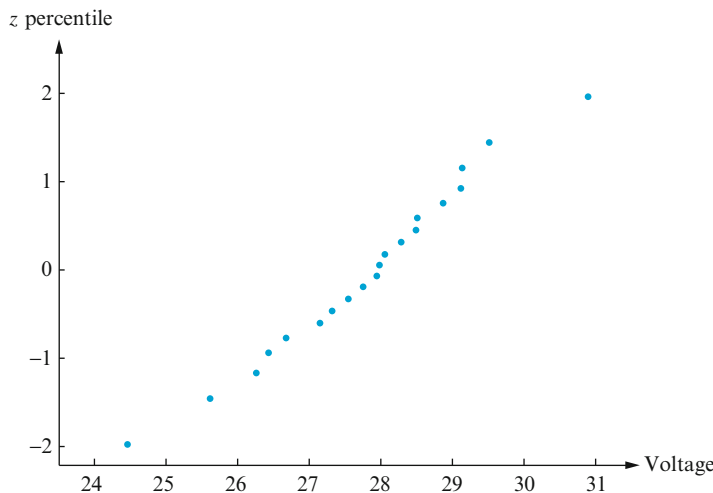
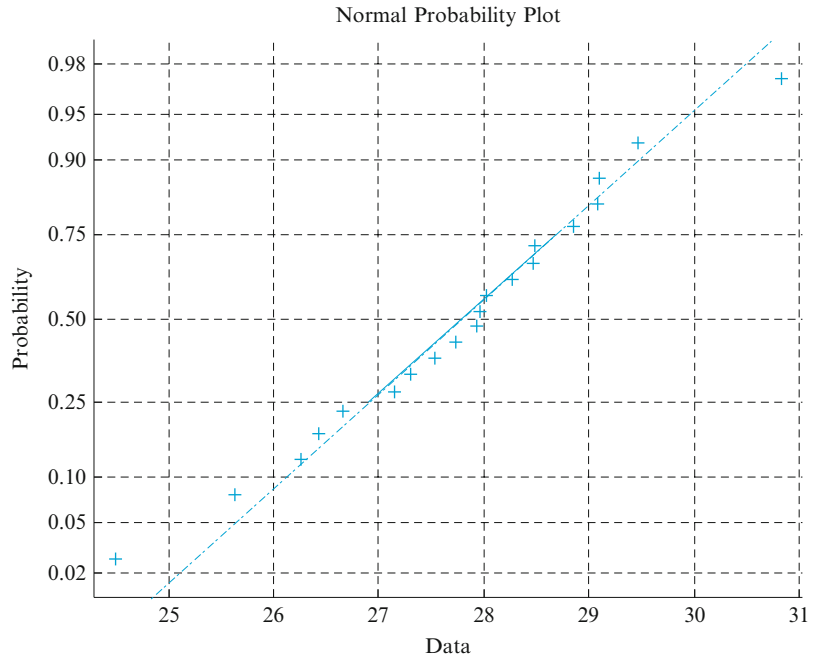


Fig. 3.32 Normal probability plot for the dielectric breakdown voltage sample



There is an alternative version of a normal probability plot in which the z percentile axis is replaced by a nonlinear probability axis. The scaling on this axis is constructed so that plotted points should again fall close to a line when the sampled distribution is normal. Figure 3.33 shows such a plot from Matlab, obtained using the `normplot` command, for the breakdown voltage data of Example 3.34. The plot remains essentially the same, and it is just the labeling of the axis that changes.

Fig. 3.33 Normal probability plot of the breakdown voltage data from Matlab



3.6.3 Departures from Normality

A nonnormal population distribution can often be placed in one of the following three categories:

1. It is symmetric and has “lighter tails” than does a normal distribution; that is, the density curve declines more rapidly out in the tails than does a normal curve.
2. It is symmetric and heavy-tailed compared to a normal distribution.
3. It is skewed; that is, the distribution is not symmetric, but rather tapers off more in one direction than the other.

A uniform distribution is light-tailed, since its density function drops to zero outside a finite interval. The density function $f(x) = 1/[\pi(1 + x^2)]$, for $-\infty < x < \infty$, is one example of a heavy-tailed distribution, since $1/(1 + x^2)$ declines much less rapidly than does $e^{-x^2/2}$. Lognormal and Weibull distributions are among those that are skewed. When the points in a normal probability plot do not adhere to a straight line, the pattern will frequently suggest that the population distribution is in a particular one of these three categories.

Figure 3.34 illustrates typical normal probability plots corresponding to three situations above. If the sample was selected from a light-tailed distribution, the largest and smallest observations are

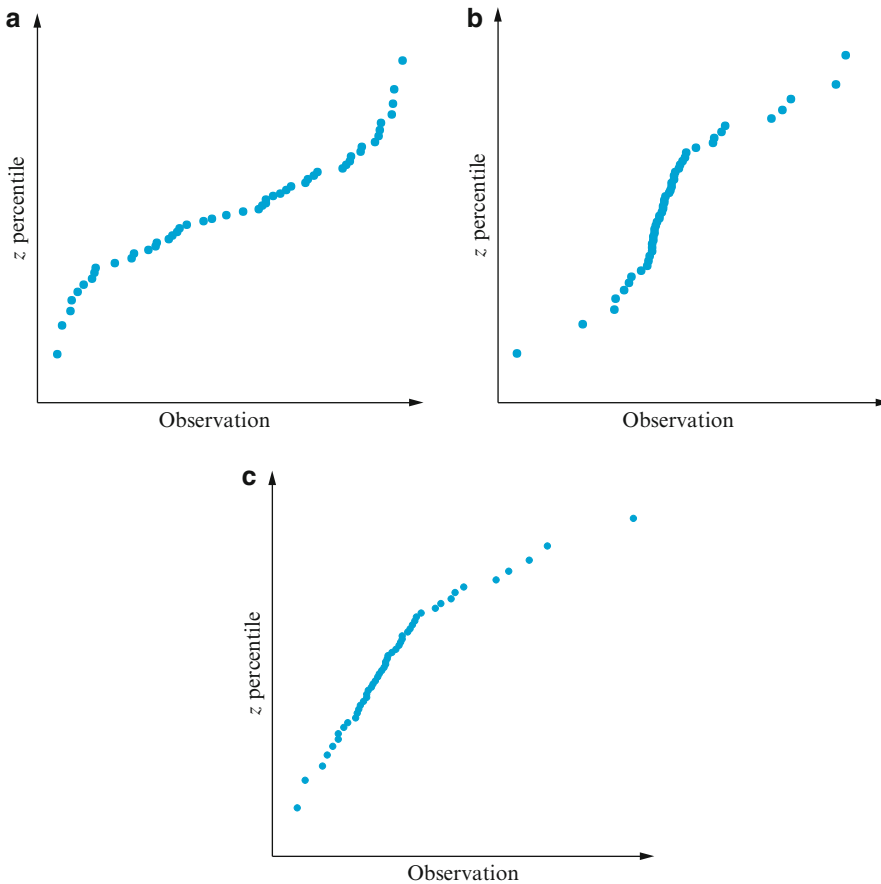


Fig. 3.34 Probability plots that suggest a non-normal distribution: (a) a plot consistent with a light-tailed distribution; (b) a plot consistent with a heavy-tailed distribution; (c) a plot consistent with a (positively) skewed distribution

usually not as extreme as would be expected from a normal random sample. Visualize a straight line drawn through the middle part of the plot; points on the far right tend to be above the line (z percentile $>$ observed value), whereas points on the left end of the plot tend to fall below the straight line (z percentile $<$ observed value). The result is an S-shaped pattern of the type pictured in Fig. 3.34a. For sample observations from a heavy-tailed distribution, the opposite effect will occur, and a normal probability plot will have an S shape with the opposite orientation, as in Fig. 3.34b. If the underlying distribution is positively skewed (a short left tail and a long right tail), the smallest sample observations will be larger than expected from a normal sample and so will the largest observations. In this case, points on both ends of the plot will fall below a straight line through the middle part, yielding a curved pattern, as illustrated in Fig. 3.34c. For example, a sample from a lognormal distribution will usually produce such a pattern; a plot of $(\ln(\text{observation}), z \text{ percentile})$ pairs should then resemble a straight line.

Even when the population distribution is normal, the sample percentiles will not coincide exactly with the theoretical percentiles because of sampling variability. How much can the points in the probability plot deviate from a straight-line pattern before the assumption of population normality is no longer plausible? This is not an easy question to answer. Generally speaking, a small sample from a normal distribution is more likely to yield a plot with a nonlinear pattern than is a large sample.

The book *Fitting Equations to Data* by Daniel Cuthbert and Fred Wood presents the results of a simulation study in which numerous samples of different sizes were selected from normal distributions. The authors concluded that there is typically greater variation in the appearance of the probability plot for sample sizes smaller than 30, and only for much larger sample sizes does a linear pattern generally predominate. When a plot is based on a small sample size, only a very substantial departure from linearity should be taken as conclusive evidence of nonnormality. A similar comment applies to probability plots for checking the plausibility of other types of distributions.

3.6.4 Beyond Normality

Consider a generic family of probability distributions involving two parameters, θ_1 and θ_2 , and let $F(x; \theta_1, \theta_2)$ denote the corresponding cdf. The family of normal distributions is one such family, with $\theta_1 = \mu$, $\theta_2 = \sigma$, and $F(x; \mu, \sigma) = \Phi[(x - \mu)/\sigma]$. Another example is the Weibull family, with $\theta_1 = \alpha$, $\theta_2 = \beta$, and

$$F(x; \alpha, \beta) = 1 - e^{-(x/\beta)^\alpha}$$

Still another family of this type is the gamma family, for which the cdf is an integral involving the incomplete gamma function that cannot be expressed in any simpler form.

The parameters θ_1 and θ_2 are said to be **location** and **scale parameters**, respectively, if $F(x; \theta_1, \theta_2)$ is a function of $(x - \theta_1)/\theta_2$. The parameters μ and σ of the normal family are location and scale parameters, respectively. Changing μ shifts the location of the bell-shaped density curve to the right or left, and changing σ amounts to stretching or compressing the measurement scale (the scale on the horizontal axis when the density function is graphed). Another example is given by the cdf

$$F(x; \theta_1, \theta_2) = 1 - e^{-e^{(x-\theta_1)/\theta_2}} \quad -\infty < x < \infty$$

A random variable with this cdf is said to have an *extreme value distribution*. It is used in applications involving component lifetime and material strength.

The parameter β of the Weibull distribution is a scale parameter. However, α is not a location parameter but instead is called a **shape parameter**. The same is true for the parameters α and β of the gamma distribution. In the usual form, the density function for any member of either the gamma or Weibull distribution is positive for $x > 0$ and zero otherwise. A location (or shift) parameter can be introduced as a third parameter γ (we did this for the Weibull distribution in Sect. 3.5) to shift the density function so that it is positive if $x > \gamma$ and zero otherwise.

When the family under consideration has only location and scale parameters, the issue of whether any member of the family is a plausible population distribution can be addressed by a single probability plot. This is exactly what we did to obtain an omnibus normal probability plot. One first obtains the percentiles of the *standardized distribution*, i.e. the one with $\theta_1 = 0$ and $\theta_2 = 1$, for percentages $100(i - .5)/n$ ($i = 1, \dots, n$). The n (observation, standardized percentile) pairs give the points in the plot.

Somewhat surprisingly, this methodology can be applied to yield an omnibus Weibull probability plot. The key result is that if X has a Weibull distribution with shape parameter α and scale parameter β , then the transformed variable $\ln(X)$ has an extreme value distribution with location parameter $\theta_1 = \ln(\beta)$ and scale parameter $\theta_2 = 1/\alpha$ (see Exercise 169). Thus a plot of the

($\ln(\text{observation})$, extreme value standardized percentile)

pairs that shows a strong linear pattern provides support for choosing the Weibull distribution as a population model.

Example 3.35 The accompanying observations are on lifetime (in hours) of power apparatus insulation when thermal and electrical stress acceleration were fixed at particular values (“On the Estimation of Life of Power Apparatus Insulation Under Combined Electrical and Thermal Stress,” *IEEE Trans. Electr. Insul.*, 1985: 70–78). A Weibull probability plot necessitates first computing the 5th, 15th, . . . , and 95th percentiles of the standard extreme value distribution. The $(100p)$ th percentile η_p satisfies

$$p = F(\eta_p; 0, 1) = 1 - e^{-e^{\eta_p}}$$

from which $\eta_p = \ln(-\ln(1 - p))$.

Observation	282	501	741	851	1072	1122	1202	1585	1905	2138
$\ln(\text{Obs.})$	5.64	6.22	6.61	6.75	6.98	7.02	7.09	7.37	7.55	7.67
Percentile	-2.97	-1.82	-1.25	-0.84	-0.51	-0.23	.05	.33	.64	1.10

The pairs $(5.64, -2.97)$, $(6.22, -1.82)$, . . . , $(7.67, 1.10)$ are plotted as points in Fig. 3.35. The straightness of the plot argues strongly for using the Weibull distribution as a model for insulation life, a conclusion also reached by the author of the cited article.

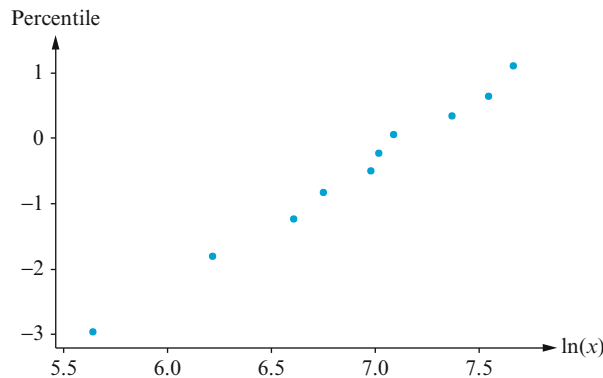


Fig. 3.35 A Weibull probability plot of the insulation lifetime data

The gamma distribution is an example of a family involving a shape parameter for which there is no transformation into a distribution that depends only on location and scale parameters. Construction of a probability plot necessitates first estimating the shape parameter from sample data (some general methods for doing this are described in Chap. 5).

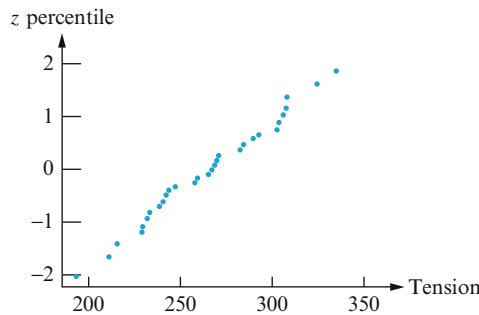
Sometimes an investigator wishes to know whether the transformed variable X^θ has a normal distribution for some value of θ (by convention, $\theta = 0$ is identified with the logarithmic transformation, in which case X has a lognormal distribution). The book *Graphical Methods for Data Analysis* by John Chambers et al. discusses this type of problem as well as other refinements of probability plotting.

3.6.5 Probability Plots in Matlab and R

Matlab, along with many statistical software packages (including R), have built-in probability plotting commands that vitiate the need for manual calculation of percentiles from the assumed population distribution. In Matlab, the `normplot(x)` command will produce a graph like the one seen in Fig. 3.33, assuming the vector `x` contains the observed data. The R command `qqnorm(x)` creates a similar graph, except that the axes are transposed (ordered observations on the vertical axis, theoretical quantiles on the horizontal). Both Matlab and R have a package called `probplot` that, with appropriate specifications of the inputs, can create probability plots for distributions besides normal (e.g., Weibull, exponential, extreme value). Refer to the help documentation in those languages for more information.

3.6.6 Exercises: Section 3.6 (101–111)

101. The accompanying normal probability plot was constructed from a sample of 30 readings on tension for mesh screens behind the surface of video display tubes. Does it appear plausible that the tension distribution is normal?



102. A sample of 15 female collegiate golfers was selected and the clubhead velocity (km/h) while swinging a driver was determined for each one, resulting in the following data (“Hip Rotational Velocities during the Full Golf Swing,” *J. of Sports Science and Medicine*, 2009: 296-299):

69.0	69.7	72.7	80.3	81.0
85.0	86.0	86.3	86.7	87.7
89.3	90.7	91.0	92.5	93.0

The corresponding *z* percentiles are

-1.83	-1.28	-0.97	-0.73	-0.52
-0.34	-0.17	0.0	0.17	0.34
0.52	0.73	0.97	1.28	1.83

Construct a normal probability plot. Is it plausible that the population distribution is normal?

103. Construct a normal probability plot for the following sample of observations on coating thickness for low-viscosity paint (“Achieving a Target Value for a Manufacturing Process: A Case Study,” *J. Qual. Tech.*, 1992: 22–26). Would you feel comfortable estimating population mean thickness using a method that assumed a normal population distribution?

.83	.88	.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

104. The article “A Probabilistic Model of Fracture in Concrete and Size Effects on Fracture Toughness” (*Mag. Concrete Res.*, 1996: 311–320) gives arguments for why fracture toughness in concrete specimens should have a Weibull distribution and presents several histograms of data that appear well fit by superimposed Weibull curves. Consider the following sample of size $n = 18$ observations on toughness for high-strength concrete (consistent with one of the histograms); values of $p_i = (i - .5)/18$ are also given.

Observation	.47	.58	.65	.69	.72	.74
p_i	.0278	.0833	.1389	.1944	.2500	.3056
Observation	.77	.79	.80	.81	.82	.84
p_i	.3611	.4167	.4722	.5278	.5833	.6389
Observation	.86	.89	.91	.95	1.01	1.04
p_i	.6944	.7500	.8056	.8611	.9167	.9722

Construct a Weibull probability plot and comment.

105. The propagation of fatigue cracks in various aircraft parts has been the subject of extensive study. The accompanying data consists of propagation lives (flight hours/10⁴) to reach a given crack size in fastener holes for use in military aircraft (“Statistical Crack Propagation in Fastener Holes Under Spectrum Loading,” *J. Aircraft*, 1983: 1028-1032):

.736	.863	.865	.913	.915	.937	.983	1.007
1.011	1.064	1.109	1.132	1.140	1.153	1.253	1.394

Construct a normal probability plot for this data. Does it appear plausible that propagation life has a normal distribution? Explain.

106. The article “The Load-Life Relationship for M50 Bearings with Silicon Nitride Ceramic Balls” (*Lubricat. Engrg.*, 1984: 153–159) reports the accompanying data on bearing load life (million revs.) for bearings tested at a 6.45 kN load.

47.1	68.1	68.1	90.8	103.6	106.0	115.0
126.0	146.6	229.0	240.0	240.0	278.0	278.0
289.0	289.0	367.0	385.9	392.0	505.0	

- (a) Construct a normal probability plot. Is normality plausible?
- (b) Construct a Weibull probability plot. Is the Weibull distribution family plausible?

107. The accompanying data on rainfall (acre-feet) from 26 seed clouds is taken from the article “A Bayesian Analysis of a Multiplicative Treatment Effect in Weather Modification” (*Technometrics*, 1975: 161-166). Construct a probability plot that will allow you to assess the plausibility of the lognormal distribution as a model for the rainfall data, and comment on what you find.

4.1	7.7	17.5	31.4	32.7	40.6	92.4
115.3	118.3	119.0	129.6	198.6	200.7	242.5
255.0	274.7	274.7	302.8	334.1	430.0	489.1
703.4	978.0	1656.0	1697.8	2745.6		

108. The accompanying observations are precipitation values during March over a 30-year period in Minneapolis–St. Paul.

.77	1.20	3.00	1.62	2.81	2.48
1.74	.47	3.09	1.31	1.87	.96
.81	1.43	1.51	.32	1.18	1.89
1.20	3.37	2.10	.59	1.35	.90
1.95	2.20	.52	.81	4.75	2.05

- (a) Construct and interpret a normal probability plot for this data set.
- (b) Calculate the square root of each value and then construct a normal probability plot based on this transformed data. Does it seem plausible that the square root of precipitation is normally distributed?
- (c) Repeat part (b) after transforming by cube roots.

109. Allowable mechanical properties for structural design of metallic aerospace vehicles requires an approval method for statistically analyzing empirical test data. The article “Establishing Mechanical Property Allowables for Metals” (*J. of Testing and Evaluation*, 1998: 293-299) used the accompanying data on tensile ultimate strength (ksi) as a basis for addressing the difficulties in developing such a method.

122.2	124.2	124.3	125.6	126.3	126.5	126.5	127.2	127.3
127.5	127.9	128.6	128.8	129.0	129.2	129.4	129.6	130.2
130.4	130.8	131.3	131.4	131.4	131.5	131.6	131.6	131.8
131.8	132.3	132.4	132.4	132.5	132.5	132.5	132.5	132.6
132.7	132.9	133.0	133.1	133.1	133.1	133.1	133.2	133.2
133.2	133.3	133.3	133.5	133.5	133.5	133.8	133.9	134.0
134.0	134.0	134.0	134.1	134.2	134.3	134.4	134.4	134.6
134.7	134.7	134.7	134.8	134.8	134.8	134.9	134.9	135.2
135.2	135.2	135.3	135.3	135.4	135.5	135.5	135.6	135.6
135.7	135.8	135.8	135.8	135.8	135.8	135.9	135.9	135.9
135.9	136.0	136.0	136.1	136.2	136.2	136.3	136.4	136.4
136.6	136.8	136.9	136.9	137.0	137.1	137.2	137.6	137.6
137.8	137.8	137.8	137.9	137.9	138.2	138.2	138.3	138.3
138.4	138.4	138.4	138.5	138.5	138.6	138.7	138.7	139.0
139.1	139.5	139.6	139.8	139.8	140.0	140.0	140.7	140.7
140.9	140.9	141.2	141.4	141.5	141.6	142.9	143.4	143.5
143.6	143.8	143.8	143.9	144.1	144.5	144.5	147.7	147.7

Use software to construct a normal probability plot of this data, and comment.

110. Let the ordered sample observations be denoted by y_1, y_2, \dots, y_n (y_1 being the smallest and y_n the largest). Our suggested check for normality is to plot the $(y_i, \Phi^{-1}[(i - .5)/n])$ pairs. Suppose we believe that the observations come from a distribution with mean 0, and let w_1, \dots, w_n be the ordered *absolute values* of the observed data. A **half-normal plot** is a probability plot of the w_i s. More specifically, since $P(|Z| \leq w) = P(-w \leq Z \leq w) = 2\Phi(w) - 1$, a half-normal plot is a plot of the $(w_i, \Phi^{-1}[(p_i + 1)/2])$ pairs, where $p_i = (i - .5)/n$. The virtue of this plot is that small or large outliers in the original sample will now appear only at the upper end of the plot rather than at both ends. Construct a half-normal plot for the following sample of measurement errors, and comment:

−3.78, −1.27, 1.44, −.39, 12.38, −43.40, 1.15, −3.96, −2.34, 30.84.

111. The following failure time observations (1000s of hours) resulted from accelerated life testing of 16 integrated circuit chips of a certain type:

82.8	11.6	359.5	502.5	307.8	179.7
242.0	26.5	244.8	304.3	379.1	212.6
229.9	558.9	366.7	203.6		

Use the corresponding percentiles of the exponential distribution with $\lambda = 1$ to construct a probability plot. Then explain why the plot assesses the plausibility of the sample having been generated from *any* exponential distribution.

3.7 Transformations of a Random Variable

Often we need to deal with a transformation $Y = g(X)$ of the random variable X . Here $g(X)$ could be a simple change of time scale. If X is the time to complete a task in minutes, then $Y = 60X$ is the completion time expressed in seconds. How can we get the pdf of Y from the pdf of X ? Consider first a simple example.

Example 3.36 The interval X in minutes between calls to a 911 center is exponentially distributed with mean 2 min, so its pdf is $f_X(x) = .5e^{-.5x}$ for $x > 0$. In order to get the pdf of $Y = 60X$, we first obtain its cdf:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(60X \leq y) = P(X \leq y/60) = F_X(y/60) \\ &= \int_0^{y/60} .5e^{-.5x} dx = 1 - e^{-y/120} \end{aligned}$$

Differentiating this with respect to y gives $f_Y(y) = (1/120)e^{-y/120}$ for $y > 0$. We see that the distribution of Y is exponential with mean 120 s (2 min).

There is nothing special here about the mean 2 and the multiplier 60. It should be clear that if we multiply an exponential random variable with mean μ by a positive constant c we get another exponential random variable with mean $c\mu$. ■

Sometimes it isn't possible to evaluate the cdf in closed form. Could we still find the pdf of Y without evaluating the integral? Yes, thanks to the following theorem.

TRANSFORMATION THEOREM

Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is monotonic (either strictly increasing or strictly decreasing) on the set of all possible values of X , so it has an inverse function $X = h(Y)$. Assume that h has a derivative $h'(y)$. Then

$$f_Y(y) = f_X(h(y)) \cdot |h'(y)| \quad (3.11)$$

Proof Here is the proof assuming that g is monotonically increasing. The proof for g monotonically decreasing is similar. First find the cdf of Y :

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq h(y)) = F_X(h(y))$$

The third equality above, wherein $g(X) \leq y$ is true iff $X \leq g^{-1}(y) = h(y)$, relies on g being a monotonically increasing function. Now differentiate the cdf with respect to y , using the Chain Rule:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(h(y)) = F'_X(h(y)) \cdot h'(y) = f_X(h(y)) \cdot h'(y)$$

The absolute value on the derivative in Eq. (3.11) is needed only in the other case where g is decreasing. The set of possible values for Y is obtained by applying g to the set of possible values for X . ■

Example 3.37 Let's apply the Transformation Theorem to the situation introduced in Example 3.36. There $Y = g(X) = 60X$ and $X = h(Y) = Y/60$.

$$f_Y(y) = f_X(h(y))|h'(y)| = .5e^{-.5x} \left| \frac{1}{60} \right| = \frac{1}{120} e^{-y/120} \quad y > 0$$

This matches the pdf of Y derived through the cdf in Example 3.36. ■

Example 3.38 Let $X \sim \text{Unif}[0, 1]$, so $f_X(x) = 1$ for $0 \leq x \leq 1$, and define a new variable $Y = 2\sqrt{X}$. The function $g(x) = 2\sqrt{x}$ is monotone on $[0, 1]$, with inverse $x = h(y) = y^2/4$. Apply the Transformation Theorem:

$$f_Y(y) = f_X(h(y))|h'(y)| = (1) \left| \frac{2y}{4} \right| = \frac{y}{2} \quad 0 \leq y \leq 2$$

The range $0 \leq y \leq 2$ comes from the fact that $y = 2\sqrt{x}$ maps $[0, 1]$ to $[0, 2]$. A graphical representation may help in understanding why the transformation $Y = 2\sqrt{X}$ yields $f_Y(y) = y/2$ if $X \sim \text{Unif}[0, 1]$. Figure 3.36a shows the uniform distribution with $[0, 1]$ partitioned into ten subintervals. In Fig. 3.36b the endpoints of these intervals are shown after transforming according to $y = 2\sqrt{x}$. The heights of the rectangles are arranged so each rectangle still has area .1, and therefore the probability in each interval is preserved. Notice the close fit of the dashed line, which has the equation $f_Y(y) = y/2$.

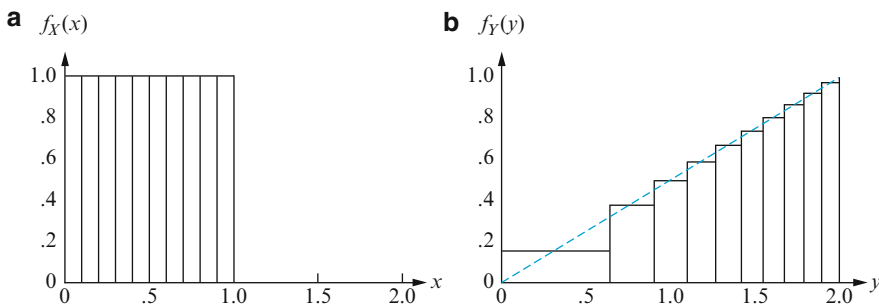


Fig. 3.36 The effect on the pdf if X is uniform on $[0, 1]$ and $Y = 2\sqrt{X}$ ■

Example 3.39 The variation in a certain electrical current source X (in milliamps) can be modeled by the pdf

$$f_X(x) = \begin{cases} 1.25 - .25x & 2 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

If this current passes through a 220- Ω resistor, the resulting power Y (in microwatts) is given by the expression $Y = 220X^2$. The function $y = g(x) = 220x^2$ is monotonically increasing on the range of X , the interval $[2, 4]$, and has inverse function $x = h(y) = g^{-1}(y) = \sqrt{y/220}$. (Notice that $g(x)$ is a parabola and thus not monotone on the entire real number line, but for the purposes of the theorem $g(x)$ only needs to be monotone on the range of the rv X .) Apply Eq. (3.11):

$$\begin{aligned} f_Y(y) &= f_X(h(y)) \cdot |h'(y)| \\ &= f_X\left(\sqrt{y/220}\right) \cdot \left|\frac{d}{dy} \sqrt{y/220}\right| \\ &= \left(1.25 - .25\sqrt{y/220}\right) \cdot \frac{1}{2\sqrt{220y}} = \frac{5}{8\sqrt{220y}} - \frac{1}{1760} \end{aligned}$$

The set of possible Y -values is determined by substituting $x = 2$ and $x = 4$ into $g(x) = 220x^2$; the resulting range for Y is $[880, 3520]$. Therefore, the pdf of $Y = 220X^2$ is

$$f_Y(y) = \begin{cases} \frac{5}{8\sqrt{220y}} - \frac{1}{1760} & 880 \leq y \leq 3520 \\ 0 & \text{otherwise} \end{cases}$$

The pdfs of X and Y appear in Fig. 3.37.

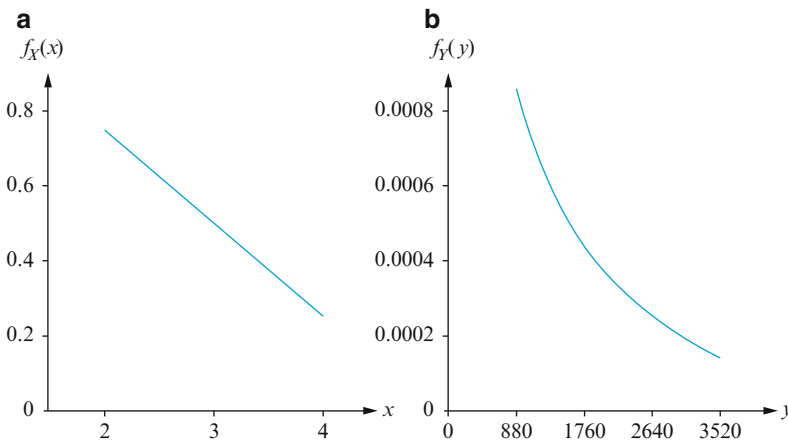


Fig. 3.37 pdfs from Example 3.39: (a) pdf of X ; (b) pdf of Y ■

The Transformation Theorem requires a monotonic transformation, but there are important applications in which the transformation is not monotone. Nevertheless, it may be possible to use the theorem anyway with a little trickery.

Example 3.40 In this example, we start with a standard normal random variable Z , and we transform to $Y = Z^2$. Of course, this is not monotonic over the interval for Z , $(-\infty, \infty)$. However, consider the transformation $U = |Z|$. Because Z has a symmetric distribution, the pdf of U is $f_U(u) = f_Z(u) + f_Z(-u) = 2f_Z(u)$. (Don't despair if this is not intuitively clear, because we'll verify it shortly. For the time being, assume it to be true.) Then $Y = Z^2 = |Z|^2 = U^2$, and the transformation in terms of U is monotonic because its set of possible values is $[0, \infty)$. Thus we can use the Transformation Theorem with $h(y) = y^{1/2}$:

$$\begin{aligned} f_Y(y) &= f_U[h(y)] |h'(y)| = 2f_Z[h(y)] |h'(y)| \\ &= \frac{2}{\sqrt{2\pi}} e^{-.5(y^{1/2})^2} \left| \frac{1}{2}y^{-1/2} \right| = \frac{1}{\sqrt{2\pi y}} e^{-y/2} \quad y > 0 \end{aligned}$$

This distribution is known as the *chi-squared distribution with one degree of freedom*. Chi-squared distributions arise frequently in statistical inference procedures, such as those in Chap. 5.

You were asked to believe intuitively that $f_U(u) = 2f_Z(u)$. Here is a little derivation that works as long as the distribution of Z is symmetric about 0. If $u > 0$,

$$\begin{aligned} F_U(u) &= P(U \leq u) = P(|Z| \leq u) = P(-u \leq Z \leq u) = 2P(0 \leq Z \leq u) \\ &= 2[F_Z(u) - F_Z(0)]. \end{aligned}$$

Differentiating this with respect to u gives $f_U(u) = 2f_Z(u)$. ■

Example 3.41 Sometimes the Transformation Theorem cannot be used at all, and you need to use the cdf. Let $f_X(x) = (x + 1)/8$, $-1 \leq x \leq 3$, and $Y = X^2$. The transformation is not monotonic on $[-1, 3]$; and, since $f_X(x)$ is not an even function, we can't employ the symmetry trick of the previous example. Possible values of Y are $\{y: 0 \leq y \leq 9\}$. Considering first $0 \leq y \leq 1$,

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{u + 1}{8} du = \frac{\sqrt{y}}{4}$$

Then, on the other subinterval, $1 < y \leq 9$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = P(-1 \leq X \leq \sqrt{y}) \\ &= \int_{-1}^{\sqrt{y}} \frac{u + 1}{8} du = (1 + y + 2\sqrt{y})/16 \end{aligned}$$

Differentiating, we get

$$f_Y(y) = \begin{cases} \frac{1}{8\sqrt{y}} & 0 < y \leq 1 \\ \frac{y + \sqrt{y}}{16y} & 1 < y \leq 9 \\ 0 & \text{otherwise} \end{cases}$$

Figure 3.38 shows the pdfs of both X and Y .

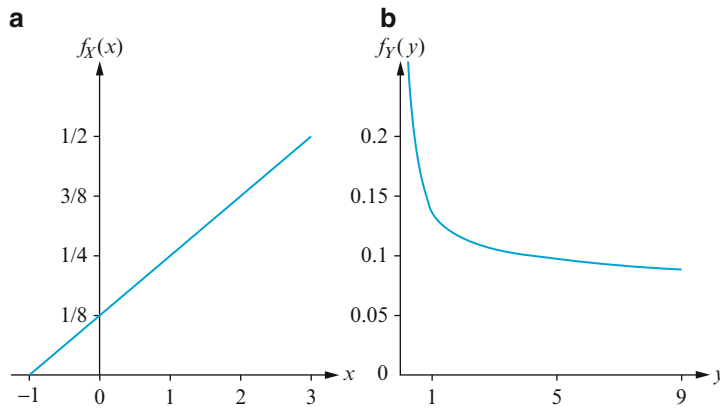


Fig. 3.38 pdfs from Example 3.41: (a) pdf of X ; (b) pdf of Y ■

3.7.1 Exercises: Section 3.7 (112–128)

112. Relative to the winning time, the time X of another runner in a ten kilometer race has pdf $f_X(x) = 2/x^3$, $x > 1$. The reciprocal $Y = 1/X$ represents the ratio of the time for the winner divided by the time of the other runner. Find the pdf of Y . Explain why Y also represents the speed of the other runner relative to the winner.
113. Let X be the fuel efficiency in miles per gallon of an extremely inefficient vehicle (a military tank, perhaps?), and suppose X has the pdf $f_X(x) = 2x$, $0 < x < 1$. Determine the pdf of $Y = 1/X$, which is fuel efficiency in gallons per mile. [Note: The distribution of Y is a special case of the *Pareto distribution* (see Exercise 10).]
114. Let X have the pdf $f_X(x) = 2/x^3$, $x > 1$. Find the pdf of $Y = \sqrt{X}$.
115. Let X have an exponential distribution with mean 2, so $f_X(x) = \frac{1}{2}e^{-x/2}$, $x > 0$. Find the pdf of $Y = \sqrt{X}$. [Note: Suppose you choose a point in two dimensions randomly, with the horizontal and vertical coordinates chosen independently from the standard normal distribution. Then X has the distribution of the squared distance from the origin and Y has the distribution of the distance from the origin. Y has a *Rayleigh distribution* (see Exercise 4).]
116. If X is distributed as $N(\mu, \sigma)$, find the pdf of $Y = e^X$. Verify that the distribution of Y matches the lognormal pdf provided in Sect. 3.5.
117. If the side of a square X is random with the pdf $f_X(x) = x/8$, $0 < x < 4$, and Y is the area of the square, find the pdf of Y .
118. Let $X \sim \text{Unif}[0, 1]$. Find the pdf of $Y = -\ln(X)$.
119. Let $X \sim \text{Unif}[0, 1]$. Find the pdf of $Y = \tan[\pi(X - .5)]$. [Note: The random variable Y has the *Cauchy distribution*, named after the famous mathematician.]
120. If $X \sim \text{Unif}[0, 1]$, find a linear transformation $Y = cX + d$ such that Y is uniformly distributed on $[A, B]$, where A and B are any two numbers such that $A < B$. Is there any other solution? Explain.
121. If X has the pdf $f_X(x) = x/8$, $0 < x < 4$, find a transformation $Y = g(X)$ such that $Y \sim \text{Unif}[0, 1]$. [Hint: The target is to achieve $f_Y(y) = 1$ for $0 \leq y \leq 1$. The Transformation Theorem will allow you to find $h(y)$, from which $g(x)$ can be obtained.]
122. If a measurement error X is uniformly distributed on $[-1, 1]$, find the pdf of $Y = |X|$, which is the magnitude of the measurement error.
123. If $X \sim \text{Unif}[-1, 1]$, find the pdf of $Y = X^2$.

124. Ann is expected at 7:00 pm after an all-day drive. She may be as much as 1 h early or as much as 3 h late. Assuming that her arrival time X is uniformly distributed over that interval, find the pdf of $|X - 7|$, the unsigned difference between her actual and predicted arrival times.
125. If $X \sim \text{Unif}[-1, 3]$, find the pdf of $Y = X^2$.
126. If a measurement error X is distributed as $N(0, 1)$, find the pdf of $|X|$, which is the magnitude of the measurement error.
127. A circular target has radius 1 foot. Assume that you hit the target (we shall ignore misses) and that the probability of hitting any region of the target is proportional to the region's area. If you hit the target at a distance Y from the center, then let $X = \pi Y^2$ be the corresponding circular area. Show that
- X is uniformly distributed on $[0, \pi]$. [Hint: Show that $F_X(x) = P(X \leq x) = x/\pi$.]
 - Y has pdf $f_Y(y) = 2y, 0 < y < 1$.
128. In Exercise 127 suppose instead that Y is uniformly distributed on $[0, 1]$. Find the pdf of $X = \pi Y^2$. Geometrically speaking, why should X have a pdf that is unbounded near 0?

3.8 Simulation of Continuous Random Variables

In Sects. 1.6 and 2.8, we discussed the need for simulation of random events and discrete random variables in situations where an “analytic” solution is very difficult or simply not possible. This section presents methods for simulating continuous random variables, including some of the built-in simulation tools of Matlab and R.

3.8.1 The Inverse CDF Method

Section 2.8 introduced the inverse cdf method for simulating discrete random variables. The basic idea was this: generate a $\text{Unif}[0, 1)$ random number and align it with the cdf of the random variable X we want to simulate. Then, determine which X value corresponds to that cdf value. We now extend this methodology to the simulation of values from a continuous distribution; the heart of the algorithm relies on the following theorem, often called the *probability integral transform*.

THEOREM

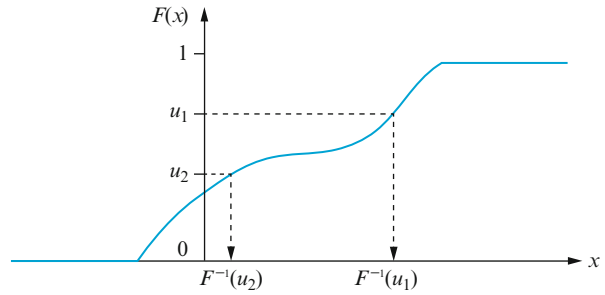
Consider any continuous distribution with pdf f and cdf F . Let $U \sim \text{Unif}[0, 1)$, and define a random variable X by

$$X = F^{-1}(U) \tag{3.12}$$

Then the pdf of X is f .

Before proving this theorem, let's consider its practical usage: Suppose we want to simulate a continuous rv whose pdf is $f(x)$, i.e., obtain successive values of X having pdf $f(x)$. If we can compute the corresponding cdf $F(x)$ and apply its inverse F^{-1} to standard uniform variates u_1, \dots, u_n , the theorem states that the resulting values $x_1 = F^{-1}(u_1), \dots, x_n = F^{-1}(u_n)$ will follow the desired distribution f . (We'll discuss the practical difficulties of implementing this method a little later.) A graphical description of the algorithm appears in Fig. 3.39.

Fig. 3.39 The inverse cdf method, illustrated



Proof Apply the Transformation Theorem (Sect. 3.7) with $f_U(u) = 1$ for $0 \leq u < 1$, $X = g(U) = F^{-1}(U)$, and thus $U = h(X) = g^{-1}(X) = F(X)$. The pdf of the transformed variable X is

$$f_X(x) = f_U(h(x)) \cdot |h'(x)| = f_U(F(x)) \cdot |F'(x)| = 1 \cdot |f(x)| = f(x)$$

In the last step, the absolute values may be removed because a pdf is always nonnegative. ■

The following box explains the implementation of the inverse cdf method justified by the preceding theorem.

INVERSE CDF METHOD

It is desired to simulate n values from a distribution with pdf $f(x)$. Let $F(x)$ be the corresponding cdf. Repeat n times:

1. Use a random-number generator (RNG) to produce a value, u , from $[0, 1)$.
2. Assign $x = F^{-1}(u)$.

The resulting values x_1, \dots, x_n form a simulation of a random variable with the original pdf, $f(x)$.

Example 3.42 Consider the electrical current distribution model of Example 3.11, where the pdf of X is given by $f(x) = 1.25 - .25x$ for $2 \leq x \leq 4$. Suppose a simulation of X is required as part of some larger system analysis. To implement the above method, the inverse of the cdf of X is required. First, compute the cdf:

$$\begin{aligned} F(x) &= P(X \leq x) = \int_2^x f(y) dy \\ &= \int_2^x (1.25 - .25y) dy = -0.125x^2 + 1.25x - 2, \quad 2 \leq x \leq 4 \end{aligned}$$

To find the probability integral transform Eq. (3.12), set $u = F(x)$ and solve for x :

$$u = F(x) = -0.125x^2 + 1.25x - 2 \Rightarrow x = F^{-1}(u) = 5 - \sqrt{9 - 8u}$$

The equation above has been solved using the quadratic formula; care must be taken to select the solution whose values lie in the interval $[2, 4]$ (the other solution, $x = 5 + \sqrt{9 - 8u}$, does not have that feature). Beginning with the usual $\text{Unif}[0, 1)$ RNG, the algorithm for simulating X is the following: given a value u from the RNG, assign $x = 5 - \sqrt{9 - 8u}$. Repeating this algorithm n times gives n simulated values of X . Programs in Matlab and R that implement this algorithm appear in Fig. 3.40; both return a vector, \mathbf{x} , containing $n = 10,000$ simulated values of the specified distribution.

```

a
x=zeros(10000,1);
for i=1:10000
    u=rand;
    x(i)=5-sqrt(9-8*u);
end

b
x <- NULL
for (i in 1:10000){
    u<-runif(1)
    x[i]<-5-sqrt(9-8*u)
}

```

Fig. 3.40 Simulation code for Example 3.42: (a) Matlab; (b) R

As discussed in Chap. 1, both of these programs can be accelerated by “vectorizing” the operations rather than using a for loop. In fact, a single line of code in either language can produce the desired result:

in Matlab: `x=5-sqrt(9-8*rand(10000,1))`
in R: `x<-5-sqrt(9-8*runif(10000))`

The pdf of the rv X and a histogram of simulation results from R appear in Fig. 3.41.

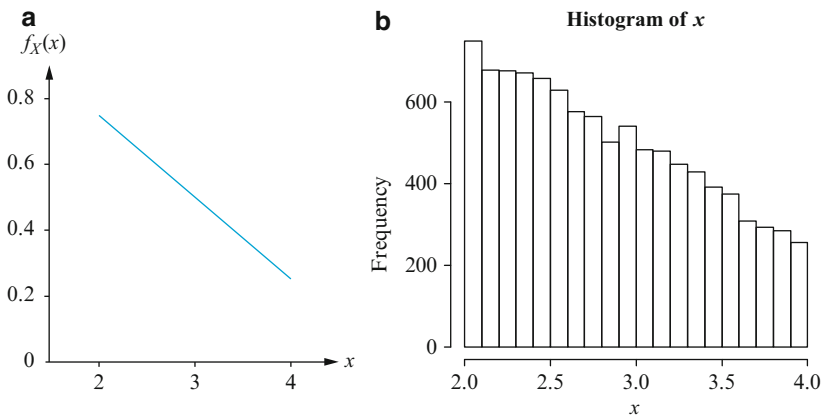


Fig. 3.41 (a) Theoretical pdf and (b) R simulation results for Example 3.42

Example 3.43 The lifetime of a certain type of drill bit has an exponential distribution with mean 100 h. An analysis of a large manufacturing process that uses these drill bits requires the simulation of this lifetime distribution, which can be achieved through the inverse cdf method. From Sect. 3.4, the cdf of this distribution is $F(x) = 1 - e^{-0.01x}$, and so the inverse cdf is $x = F^{-1}(u) = -100\ln(1 - u)$. Applying this function to $\text{Unif}[0, 1)$ random numbers will generate the desired simulation. (Don’t let the negative sign at the front worry you: since $0 \leq u < 1$, $1 - u$ lies between 0 and 1, and so its logarithm is negative and the resulting value of x is actually positive.)

As a check, the code `x=-100*log(1-rand(10000,1))` was submitted to Matlab and the resulting sample mean and sd were obtained using `mean(x)` and `std(x)`. Exponentially distributed rvs have standard deviation equal to the mean, so the theoretical answers are $\mu = 100$ and $\sigma = 100$. The Matlab simulation yielded $\bar{x} = 99.3724$ and $s = 100.8908$, both of which are reasonably close to 100 and validate the inverse cdf formula.

In general, an exponential distribution with mean μ (equivalently, parameter $\lambda = 1/\mu$) can be simulated using the transform $x = -\mu\ln(1 - u)$.

The preceding two examples illustrated the inverse cdf method for fairly simple density functions: a linear polynomial and an exponential function. In practice, the algebraic complexity of $f(x)$ can

often be a barrier to implementing this simulation technique. After all, the algorithm requires that we can (1) obtain the cdf $F(x)$ in closed form and (2) find the inverse function of F in closed form. Consider, for example, attempting to simulate values from the $N(0, 1)$ distribution: its cdf is the function denoted $\Phi(z)$ and given by the integral expression $(1/\sqrt{2\pi}) \int_{-\infty}^z e^{-u^2/2} du$. There is no closed-form expression for this integral, let alone a method to solve $u = \Phi(z)$ for z and implement Eq. (3.12). (As a reminder, the lack of a closed-form expression for $\Phi(z)$ is the reason that software or tables are always required for calculations involving normal probabilities.) Thankfully, most software packages, including Matlab and R, have built-in tools to simulate normally distributed variates (using a very clever algorithm called the *Box-Muller method*; see Sect. 4.6). We'll discuss built-in simulation tools at the end of this section.

As the next example illustrates, even when $F(x)$ can be determined in closed form we cannot necessarily implement the inverse cdf method, because $F(x)$ cannot always be inverted. This difficulty surfaces in practice when attempting to simulate values from a gamma distribution, for instance.

Example 3.44 The measurement error X (in mV) of a particular volt-meter has the following distribution: $f(x) = (4 - x^2)/9$ for $-1 \leq x \leq 2$ (and $f(x) = 0$ otherwise). To use the inverse cdf method to simulate X , begin by calculating its cdf:

$$F(x) = \int_{-1}^x \frac{4 - y^2}{9} dy = \frac{-x^3 + 12x + 11}{27}$$

To implement step 2 of the inverse cdf method requires solving $F(x) = u$ for x ; since $F(x)$ is a cubic polynomial, this is not a simple task. Advanced computer algebra systems can solve this equation, though the general solution is unwieldy (and such a solution doesn't exist at all for 5th-degree and higher polynomials). Readers familiar with numerical analysis methods may recognize that, for any specified numerical value of u , a root-finding algorithm (such as Newton–Raphson) can be implemented to *approximate* the solution x . This latter method, however, is computationally intensive, especially if it's desirable to generate 10,000 or more simulated values of x . ■

The preceding example suggests that the inverse cdf method is insufficient for simulating all continuous distributions in practice. We next consider an alternative algorithm that, while less efficient, has a broader scope.

3.8.2 The Accept–Reject Method

When the inverse cdf method of simulation cannot be implemented, the *accept–reject method* provides an alternative. The downside of the accept–reject method, as will be explained below, is that only some of the random numbers generated by software will be used (“accepted”), while others will be “rejected.” As a result, one needs to create more—sometimes, many more—random variates than the desired number of simulated values.

Suppose we wish to simulate a random variable X , whose pdf is $f(x)$. The key to the accept–reject method is to begin with a *different* pdf, call it $g(x)$, that satisfies two properties: (1) we can already simulate values from $g(x)$, so g is either algebraically simple or else built into our software package; (2) the set of possible x -values for the distribution specified by $g(x)$ equals (or exceeds) that of $f(x)$. For example, to simulate the distribution in Example 3.44, whose range of x -values is $[-1, 2]$, one might select for $g(x)$ the uniform distribution on $[-1, 2]$, i.e., $g(x) = 1/3$ for $-1 \leq x \leq 2$. If X takes on values across $[0, \infty)$, then an exponential pdf would be a logical choice for $g(x)$.

ACCEPT-REJECT METHOD

It is desired to simulate n values from a distribution with pdf $f(x)$. Let $g(x)$ be some other pdf such that the ratio f/g is bounded, i.e., there exists a constant c such that $f(x)/g(x) \leq c$ for all x . (The constant c is sometimes called the *majorization constant*.) Proceed as follows:

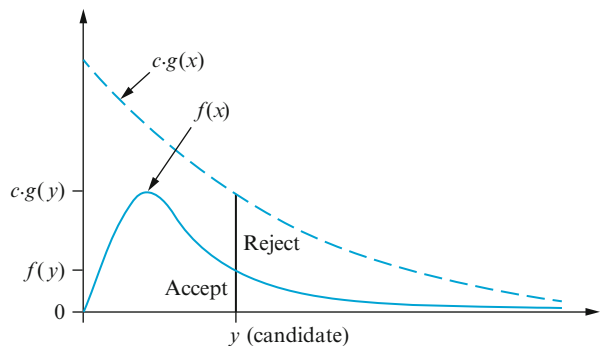
1. Generate a variate, y , from the distribution g . This value y is called a *candidate*.
2. Generate a standard uniform variate, u .
3. If $u \cdot c \cdot g(y) \leq f(y)$, then assign $x = y$ (i.e., “accept” the candidate). Otherwise, discard (“reject”) y and return to step 1.

These steps are repeated until n candidate values have been accepted. The resulting accepted values x_1, \dots, x_n form a simulation of a random variable with the original pdf, $f(x)$.

A proof that the method works—i.e., that the resulting values really do simulate the target distribution $f(x)$ —requires material from Chap. 4 (see Exercise 22 at the end of Sect. 4.1).

Figure 3.42 illustrates the key step in this algorithm. A candidate y has been generated on the common interval of the pdfs f and g . Given y , the left-hand side of the inequality in step 3, $U \cdot c \cdot g(y)$, is uniformly distributed on the interval from 0 to $c \cdot g(y)$ (since U itself is standard uniform). If it happens that $u \cdot c \cdot g(y)$ falls between 0 and $f(y)$, i.e., lies underneath the target pdf f , then that y -value is accepted as coming from f ; otherwise, y is rejected.

Fig. 3.42 The accept–reject method



As a corollary to proving the validity of the accept–reject method, it can also be shown that the probability any particular candidate y is accepted equals $1/c$. (The value of c must always exceed 1; can you see why?) Since successive candidates are independent, it follows that the number of candidates required to generate a single acceptable value has a geometric distribution, and the expected number of candidates to generate one x from $f(x)$ is $1/(1/c) = c$. By extension, the expected number of candidates required to generate our simulation sample of size n is cn . Consequently, the majorization constant c should always be made as small as possible, i.e., we should find the *smallest* value c such that $f(x)/g(x) \leq c$ for all x under consideration.

Example 3.45 (Example 3.44 continued) In order to simulate 10,000 values from $f(x) = (4 - x^2)/9$, $-1 \leq x \leq 2$, we will rely on our ability to generate variates from $g(x) = 1/3$ on $-1 \leq x \leq 2$, the uniform pdf. To implement the accept–reject method, we must determine the majorization constant, c , by looking at the ratio f/g :

$$\frac{f(x)}{g(x)} = \frac{(4-x^2)/9}{1/3} = \frac{4-x^2}{3} \leq \frac{4-0^2}{3} = \frac{4}{3} \quad \text{for } -1 \leq x \leq 2$$

The expression $4 - x^2$ represents a downward-facing parabola with vertex at $x = 0$, so it is clearly maximized at 0. We conclude that $c = 4/3$ is the smallest possible majorization constant, and that is what we shall use. To create the desired simulation, the following steps are repeated until 10,000 values are accepted in step 3.

1. Generate y from the uniform distribution on $[-1, 2]$.
2. Generate u from the standard uniform RNG.
3. If $u \cdot \frac{4}{3} \cdot \frac{1}{3} \leq \frac{4-y^2}{9}$, assign $x = y$; otherwise, discard y and return to step 1.

Figure 3.43 shows the preceding algorithm implemented in Matlab and R. Both programs result in a vector of 10,000 simulated values from the pdf $f(x)$. Figure 3.44 shows $f(x)$ alongside the simulated values from Matlab. Since $c = 4/3$, it's expected to require $4/3(10,000) = 13,333$ iterations of the while loop to create the desired simulation size; by adding a counter to the program, one run of the Matlab code was found to use 13,303 candidates.

<p>a</p> <pre> x=zeros(10000,1); i=0; while i<10000 y=-1+3*rand; u=rand; if u*4/3*1/3<=(4-y^2)/9 i=i+1; x(i)=y; end end end </pre>	<p>b</p> <pre> x <- NULL i <- 0 while (i<10000){ y <- -1+3*runif(1) u <- runif(1) if (u*4/3*1/3<=(4-y^2)/9){ i <- i+1 x[i] <- y } } </pre>
---	---

Fig. 3.43 Simulation code for Example 3.45: (a) Matlab; (b) R

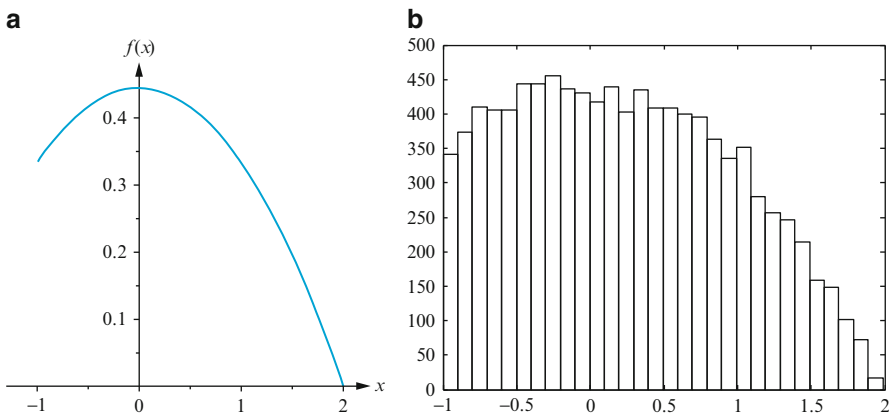


Fig. 3.44 pdf and histogram of simulated values for Example 3.45

You may have noticed that step 3 may be simplified: the inequality $u \leq (4 - y^2)/4$ would be equivalent to the one presented. In fact, it is very common to see this final step of the accept–reject algorithm written as “accept y iff $u \leq f(y)/[c \cdot g(y)]$.” ■

For more information on the accept–reject method and selection of a sensible “candidate” distribution $g(x)$ consult the text *Simulation* by Ross listed in the references.

3.8.3 Built-In Simulation Packages for Matlab and R

As was true for the most common discrete distributions, many software packages have built-in tools for simulating values from the continuous models named in this chapter. Table 3.3 summarizes the relevant functions in Matlab and R for the uniform, normal, gamma, and exponential distributions; the variable n refers to the desired number of simulated values of the distribution. Both packages include similar commands for the Weibull, lognormal, and beta distributions.

Table 3.3 Functions to simulate major continuous distributions in Matlab and R

Distribution	Matlab code	R code
Unif[A, B]	<code>unifrnd(A, B, n, 1)</code>	<code>runif(n, A, B)</code>
$N(\mu, \sigma)$	<code>normrnd(μ, σ, n, 1)</code>	<code>rnorm(n, μ, σ)</code>
Gamma(α, β)	<code>gamrnd(α, β, n, 1)</code>	<code>rgamma(n, α, $1/\beta$)</code>
Exponential(λ)	<code>exprrnd($1/\lambda$, n, 1)</code>	<code>rexp(n, λ)</code>

As was the case with the cdf commands discussed in Sect. 3.4, Matlab and R parameterize the gamma and exponential distributions differently: Matlab always requires the “scale” parameter $\beta = 1/\lambda$, while R takes in the “rate” parameter $\lambda = 1/\beta$. (In the gamma simulation command, this can be overridden by naming the final argument `scale`, as in `rgamma(n, α , scale = β)`.) In R, the command `rnorm(n)` will generate standard normal variates (i.e., with $\mu = 0$ and $\sigma = 1$), but the μ and σ arguments are required in Matlab. Similarly, R will generate standard uniform variates ($A = 0$ and $B = 1$), the basis for many of our simulation methods, with the command `runif(n)`. Matlab’s corresponding syntax is `rand(n, 1)`; if you type `rand(100)` instead of `rand(100, 1)`, you will receive a 100-by-100 *matrix* of Unif[0, 1) values.

3.8.4 Precision of Simulation Results

Sect. 2.8 discusses in detail the precision of estimates associated with simulating discrete random variables. The same results apply in the continuous case. In particular, the estimated standard error in using a sample proportion \hat{p} to estimate the true probability of an event is still $\sqrt{\hat{p}(1 - \hat{p})/n}$, where n is the simulation size. Also, the estimated standard error in using a sample mean, \bar{x} , to estimate the true expected value μ of a (continuous) rv X is s/\sqrt{n} , where s is the sample standard deviation of the simulated values of X . Refer back to Sect. 2.8 for more details.

3.8.5 Exercises: Section 3.8 (129–139)

129. The amount of time (hours) required to complete an unusually short statistics homework assignment is modeled by the pdf $f(x) = x/2$ for $0 < x < 2$ (and $= 0$ otherwise).
- Obtain the cdf and then its inverse.
 - Write a program to simulate 10,000 values from this distribution.
 - Compare the sample mean and standard deviation of your 10,000 simulated values to the theoretical mean and sd of this distribution (which you can determine by calculating the appropriate integrals).
130. The Weibull distribution was introduced in Sect. 3.5.
- Find the inverse cdf for the Weibull distribution.
 - Write a program to simulate n values from a Weibull distribution. Your program should have three inputs: the desired number of simulated values n and the two parameters α and β . It should have a single output: an $n \times 1$ vector of simulated values.
 - Use your program from part (b) to simulate 10,000 values from a Weibull(4, 6) distribution and estimate the mean of this distribution. The correct value of the mean is $6\Gamma(5/4) \approx 5.438$; how close is your sample mean?
131. Consider the pdf for the rv $X =$ magnitude (in newtons) of a dynamic load on a bridge, given in Example 3.7:

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Write a program to simulate values from this distribution using the inverse cdf method.

132. In distributed computing, any given task is split into smaller sub-tasks which are handled by separate processors (which are then recombined by a multiplexer). Consider a distributed computing system with 4 processors, and suppose for one particular purpose that pdf of completion time for a particular sub-task (microseconds) on any one of the processors is given by

$$f(x) = \begin{cases} \frac{20}{3x^2} & 4 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

That is, the sub-task completion times X_1, X_2, X_3, X_4 of the four processors each have the above pdf.

- Write a program to simulate the above pdf using the inverse cdf method.
 - The overall time to complete any task is the largest of the four sub-task completion times: if we call this variable Y , then $Y = \max(X_1, X_2, X_3, X_4)$. (We assume that the multiplexing time is negligible). Use your program in part (a) to simulate 10,000 values of the rv Y . Create a histogram of the simulated values of Y , and also use your simulation to estimate both $E(Y)$ and $SD(Y)$.
133. Exercise 16 in Sect. 3.1 introduced the following model for wait times at street crossings:

$$f(x; \theta, \tau) = \begin{cases} \frac{\theta}{\tau} (1 - x/\tau)^{\theta-1} & 0 \leq x < \tau \\ 0 & \text{otherwise} \end{cases}$$

where $\theta > 0$ and $\tau > 0$ are the parameters of the model.

- (a) Write a function to simulate values from this distribution, implementing the inverse cdf method. Your function should have three inputs: the desired number of simulated values n and values for the two parameters for θ and τ .
- (b) Use your function in part (a) to simulate 10,000 values from this wait time distribution with $\theta = 4$ and $\tau = 80$. Estimate $E(X)$ under these parameter settings. How close is your estimate to the correct value of 16?
134. Explain why the transformation $x = -\mu \ln(u)$ may be used to simulate values from an exponential distribution with mean μ . (This expression is slightly simpler than the one established in this section.)
135. Recall the rv X = amount of gravel (in tons) sold by a construction supply company in a given week from Example 3.9, whose pdf is

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Consider simulating values from this distribution using the accept–reject method with a Unif[0, 1] “candidate” distribution, i.e., $g(x) = 1$ for $0 \leq x \leq 1$.

- (a) Find the smallest majorization constant c so that $f(x)/g(x) \leq c$ for all x in $[0, 1]$.
- (b) Write a program to simulate values from this distribution.
- (c) On the average, how many candidate values must your program generate in order to create 10,000 “accepted” values?
- (d) Simulate 10,000 values from this distribution, and use these to estimate the mean μ of this distribution. How close is your sample mean to the true value of μ (which you can determine using the appropriate integral)?
- (e) The supply company’s management looks at quarterly data for X , i.e., values X_1, \dots, X_{13} for 13 weeks (one quarter of a year). Of particular interest is the variable $M = \min(X_1, \dots, X_{13})$, the least amount of gravel sold in one week during a quarter. Use your program in (b) to simulate the rv M , and use the results of at least 10,000 simulated values of M to estimate $P(M < .1)$, the chance that the worst sales week in a quarter saw less than .1 tons of gravel sold. [*Hint*: Simulate each X_i 10,000 times for $i = 1, \dots, 13$, and then compute the minimum of each set of 13 values to create a value for M .]
136. The time required to complete a 3-h final exam is modeled by the following pdf:

$$f(x) = \begin{cases} \frac{4}{27}x^2(3 - x) & 0 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Consider simulating values from this distribution using the accept–reject method with a uniform “candidate” distribution on the interval $[0, 3]$.

- (a) Find the smallest majorization constant c so that $f(x)/g(x) \leq c$ for all x in $[0, 3]$. [*Hint*: What is the pdf of the uniform distribution on $[0, 3]$?]
- (b) Write a program to simulate values from this distribution.
- (c) On the average, how many candidate values must your program generate in order to create 10,000 “accepted” values?
- (d) A professor has 20 students taking her class (lucky professor!). Assume her 20 students’ completion times on the final exam can be modeled as 20 independent observations from the above pdf. The professor must stay at the final exam until all 20 students are finished (i.e., until the last student leaves). Use your program in (b) to simulate the rv L = time, in

hours, that the professor sits proctoring her final exam to 20 students. Use your simulation to estimate $P(L \geq 35/12)$, the probability she will have to stay into the last 5 min of the final exam period.

137. The *half-normal* distribution has the following pdf:

$$f(x) = \begin{cases} \sqrt{\frac{2}{\pi}} \cdot e^{-x^2/2} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

This is the distribution of $|Z|$, where $Z \sim N(0, 1)$; equivalently, it's the pdf that arises by "folding" the standard normal distribution in half along its line of symmetry. Consider simulating values from this distribution using the accept–reject method with a candidate distribution $g(x) = e^{-x}$ for $x \geq 0$ (i.e., an exponential pdf with $\lambda = 1$).

- Find the inverse cdf corresponding to $g(x)$. (This will allow us to simulate values from the candidate distribution.)
 - Find the smallest majorization constant c so that $f(x)/g(x) \leq c$ for all $x \geq 0$. [*Hint*: Use calculus to determine where the ratio $f(x)/g(x)$ is maximized.]
 - On the average, how many candidate values will be required to generate 10,000 "accepted" values?
 - Write a program to construct 10,000 values from a half-normal distribution.
138. As discussed previously, the normal distribution cannot be simulated using the inverse cdf method. One possibility for simulating from a standard normal distribution is to employ the accept–reject method with candidate distribution

$$g(x) = \frac{1}{\pi(1+x^2)} \quad -\infty < x < \infty$$

(This is the *Cauchy* distribution.)

- Find the cdf and inverse cdf corresponding to $g(x)$. (This will allow us to simulate values from the candidate distribution.)
 - Find the smallest majorization constant c so that $f(x)/g(x) \leq c$ for all x , where $f(x)$ is the standard normal pdf. [*Hint*: Use calculus to determine where the ratio $f(x)/g(x)$ is maximized.]
 - On the average, how many candidate values will be required to generate 10,000 "accepted" values?
 - Write a program to construct 10,000 values from a standard normal distribution.
 - Suppose that you now wish to simulate from a $N(\mu, \sigma)$ distribution. How would you modify your program in part (d)?
139. Explain why the majorization constant c in the accept–reject algorithm must be ≥ 1 . [*Hint*: If $c < 1$, then $f(x) < g(x)$ for all x . Why is this bad?]

3.9 Supplementary Exercises (140–172)

140. An insurance company issues a policy covering losses up to 5 (in thousands of dollars). The loss, X , follows a distribution with density function:

$$f(x) = \begin{cases} \frac{3}{x^4} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

What is the expected value of the amount paid under the policy?

141. Let X = the time it takes a read/write head to locate a desired record on a computer disk memory device once the head has been positioned over the correct track. If the disks rotate once every 25 msec, a reasonable assumption is that X is uniformly distributed on the interval $[0, 25]$.
- Compute $P(10 \leq X \leq 20)$.
 - Compute $P(X \geq 10)$.
 - Obtain the cdf $F(x)$.
 - Compute $E(X)$ and $SD(X)$.
142. A 12-in. bar clamped at both ends is subjected to an increasing amount of stress until it snaps. Let Y = the distance from the left end at which the break occurs. Suppose Y has pdf

$$f(y) = \begin{cases} \frac{y}{24} \left(1 - \frac{y}{12}\right) & 0 \leq y \leq 12 \\ 0 & \text{otherwise} \end{cases}$$

Compute the following:

- The cdf of Y , and graph it.
 - $P(Y \leq 4)$, $P(Y > 6)$, and $P(4 \leq Y \leq 6)$.
 - $E(Y)$, $E(Y^2)$, and $SD(Y)$.
 - The probability that the break point occurs more than 2 in. from the expected break point.
 - The expected length of the shorter segment when the break occurs.
143. Let X denote the time to failure (in years) of a hydraulic component. Suppose the pdf of X is $f(x) = 32/(x+4)^3$ for $x > 0$.
- Verify that $f(x)$ is a legitimate pdf.
 - Determine the cdf.
 - Use the result of part (b) to calculate the probability that time to failure is between 2 and 5 years.
 - What is the expected time to failure?
 - If the component has a salvage value equal to $100/(4+x)$ when its time to failure is x , what is the expected salvage value?
144. The completion time X for a task has cdf $F(x)$ given by

$$\begin{cases} 0 & x < 0 \\ \frac{x^3}{3} & 0 \leq x < 1 \\ 1 - \frac{1}{2} \left(\frac{7}{3} - x\right) \left(\frac{7}{4} - \frac{3}{4}x\right) & 1 \leq x \leq \frac{7}{3} \\ 1 & x \geq \frac{7}{3} \end{cases}$$

- Obtain the pdf $f(x)$ and sketch its graph.
 - Compute $P(.5 \leq X \leq 2)$.
 - Compute $E(X)$.
145. The breakdown voltage of a randomly chosen diode of a certain type is known to be normally distributed with mean value 40 V and standard deviation 1.5 V.

- (a) What is the probability that the voltage of a single diode is between 39 and 42?
- (b) What value is such that only 15% of all diodes have voltages exceeding that value?
- (c) If four diodes are independently selected, what is the probability that at least one has a voltage exceeding 42?
146. The article “Computer Assisted Net Weight Control” (*Qual. Prog.*, 1983: 22–25) suggests a normal distribution with mean 137.2 oz and standard deviation 1.6 oz, for the actual contents of jars of a certain type. The stated contents was 135 oz.
- (a) What is the probability that a single jar contains more than the stated contents?
- (b) Among ten randomly selected jars, what is the probability that at least eight contain more than the stated contents?
- (c) Assuming that the mean remains at 137.2, to what value would the standard deviation have to be changed so that 95% of all jars contain more than the stated contents?
147. When circuit boards used in the manufacture of compact disk players are tested, the long-run percentage of defectives is 5%. Suppose that a batch of 250 boards has been received and that the condition of any particular board is independent of that of any other board.
- (a) What is the approximate probability that at least 10% of the boards in the batch are defective?
- (b) What is the approximate probability that there are exactly 10 defectives in the batch?
148. The article “Reliability of Domestic-Waste Biofilm Reactors” (*J. Envir. Engr.*, 1995: 785–790) suggests that substrate concentration (mg/cm^3) of influent to a reactor is normally distributed with $\mu = .30$ and $\sigma = .06$.
- (a) What is the probability that the concentration exceeds .25?
- (b) What is the probability that the concentration is at most .10?
- (c) How would you characterize the largest 5% of all concentration values?
149. Let X = the hourly median power (in decibels) of received radio signals transmitted between two cities. The authors of the article “Families of Distributions for Hourly Median Power and Instantaneous Power of Received Radio Signals” (*J. Res. Nat. Bureau Standards*, vol. 67D, 1963: 753–762) argue that the lognormal distribution provides a reasonable probability model for X . If the parameter values are $\mu = 3.5$ and $\sigma = 1.2$, calculate the following:
- (a) The mean value and standard deviation of received power.
- (b) The probability that received power is between 50 and 250 dB.
- (c) The probability that X is less than its mean value. Why is this probability not .5?
150. Let X be a nonnegative continuous random variable with cdf $F(x)$ and mean $E(X)$.
- (a) The definition of expected value is $E(X) = \int_0^{\infty} xf(x)dx$. Replace the first x inside the integral with $\int_0^x 1 dy$ to create a double integral expression for $E(X)$. [The “order of integration” should be $dy dx$.]
- (b) Rearrange the order of integration, keeping track of the revised limits of integration, to show that

$$E(X) = \int_0^{\infty} \int_y^{\infty} f(x)dx dy$$

- (c) Evaluate the dx integral in (b) to show that $E(X) = \int_0^{\infty} [1 - F(y)]dy$. (This provides an alternate derivation of the formula established in Exercise 38.)
- (d) Use the result of (c) to verify that the expected value of an exponentially distributed rv with parameter λ is $1/\lambda$.

151. The reaction time (in seconds) to a stimulus is a continuous random variable with pdf

$$f(x) = \begin{cases} \frac{3}{2x^2} & 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Obtain the cdf.
 - (b) What is the probability that reaction time is at most 2.5 s? Between 1.5 and 2.5 s?
 - (c) Compute the expected reaction time.
 - (d) Compute the standard deviation of reaction time.
 - (e) If an individual takes more than 1.5 s to react, a light comes on and stays on either until one further second has elapsed or until the person reacts (whichever happens first). Determine the expected amount of time that the light remains lit. [Hint: Let $h(X)$ = the time that the light is on as a function of reaction time X .]
152. The article “Characterization of Room Temperature Damping in Aluminum-Indium Alloys” (*Metallurgical Trans.*, 1993: 1611-1619) suggests that aluminum matrix grain size (μm) for an alloy consisting of 2% indium could be modeled with a normal distribution with mean 96 and standard deviation 14.
- (a) What is the probability that grain size exceeds 100 μm ?
 - (b) What is the probability that grain size is between 50 and 80 μm ?
 - (c) What interval (a, b) includes the central 90% of all grain sizes (so that 5% are below a and 5% are above b)?
153. The article “Determination of the MTF of Positive Photoresists Using the Monte Carlo Method” (*Photographic Sci. Engrg.*, 1983: 254–260) proposes the exponential distribution with parameter $\lambda = .93$ as a model for the distribution of a photon’s free path length (mm) under certain circumstances. Suppose this is the correct model.
- (a) What is the expected path length, and what is the standard deviation of path length?
 - (b) What is the probability that path length exceeds 3.0? What is the probability that path length is between 1.0 and 3.0?
 - (c) What value is exceeded by only 10% of all path lengths?
154. The article “The Prediction of Corrosion by Statistical Analysis of Corrosion Profiles” (*Corrosion Sci.*, 1985: 305–315) suggests the following cdf for the depth X of the deepest pit in an experiment involving the exposure of carbon manganese steel to acidified seawater:

$$F(x; \theta_1, \theta_2) = e^{-e^{-(x-\theta_1)/\theta_2}} \quad -\infty < x < \infty$$

(This is called the *largest extreme value distribution* or *Gumbel distribution*.) The investigators proposed the values $\theta_1 = 150$ and $\theta_2 = 90$. Assume this to be the correct model.

- (a) What is the probability that the depth of the deepest pit is at most 150? At most 300? Between 150 and 300?
- (b) Below what value will the depth of the maximum pit be observed in 90% of all such experiments?
- (c) What is the density function of X ?
- (d) The density function can be shown to be unimodal (a single peak). Above what value on the measurement axis does this peak occur? (This value is the mode.)
- (e) It can be shown that $E(X) \approx .5772\theta_2 + \theta_1$. What is the mean for the given values of θ_1 and θ_2 , and how does it compare to the median and mode? Sketch the graph of the density function.

155. Let t = the amount of sales tax a retailer owes the government for a certain period. The article “Statistical Sampling in Tax Audits” (*Statistics and the Law*, 2008: 320–343) proposes modeling the uncertainty in t by regarding it as a normally distributed random variable with mean value μ and standard deviation σ (in the article, these two parameters are estimated from the results of a tax audit involving n sampled transactions). If a represents the amount the retailer is assessed, then an underassessment results if $t > a$ and an overassessment if $a > t$. We can express this in terms of a *loss function*, a function that shows zero loss if $t = a$ but increases as the gap between t and a increases. The proposed loss function is $L(a, t) = t - a$ if $t > a$ and $= k(a - t)$ if $t \leq a$ ($k > 1$ is suggested to incorporate the idea that over-assessment is more serious than under-assessment).
- Show that $a^* = \mu + \sigma\Phi^{-1}(1/(k + 1))$ is the value of a that minimizes the expected loss, where Φ^{-1} is the inverse function of the standard normal cdf.
 - If $k = 2$ (suggested in the article), $\mu = \$100,000$, and $\sigma = \$10,000$, what is the optimal value of a , and what is the resulting probability of over-assessment?
156. A *mode* of a continuous distribution is a value x^* that maximizes $f(x)$.
- What is the mode of a normal distribution with parameters μ and σ ?
 - Does the uniform distribution with parameters A and B have a single mode? Why or why not?
 - What is the mode of an exponential distribution with parameter λ ? (Draw a picture.)
 - If X has a gamma distribution with parameters α and β , and $\alpha > 1$, determine the mode. [Hint: $\ln[f(x)]$ will be maximized if and only if $f(x)$ is, and it may be simpler to take the derivative of $\ln[f(x)]$.]
157. The article “Error Distribution in Navigation” (*J. Institut. Navigation*, 1971: 429–442) suggests that the frequency distribution of positive errors (magnitudes of errors) is well approximated by an exponential distribution. Let X = the lateral position error (nautical miles), which can be either negative or positive. Suppose the pdf of X is

$$f(x) = .1e^{-.2|x|} \quad -\infty < x < \infty$$

- Sketch a graph of $f(x)$ and verify that $f(x)$ is a legitimate pdf (show that it integrates to 1).
 - Obtain the cdf of X and sketch it.
 - Compute $P(X \leq 0)$, $P(X \leq 2)$, $P(-1 \leq X \leq 2)$, and the probability that an error of more than 2 miles is made.
158. The article “Statistical Behavior Modeling for Driver-Adaptive Precrash Systems” (*IEEE Trans. on Intelligent Transp. Systems*, 2013: 1-9) proposed the following distribution for modeling the behavior of what the authors called “the criticality level of a situation” X .

$$f(x; \lambda_1, \lambda_2, p) = \begin{cases} p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

This is often called the *hyperexponential* or *mixed exponential distribution*.

- What is the cdf $F(x; \lambda_1, \lambda_2, p)$?
- If $p = .5$, $\lambda_1 = 40$, $\lambda_2 = 200$ (values of the λ s suggested in the cited article), calculate $P(X > .01)$.
- If X has $f(x; \lambda_1, \lambda_2, p)$ as its pdf, what is $E(X)$?
- Using the fact that $E(X^2) = 2/\lambda^2$ when X has an exponential distribution with parameter λ , compute $E(X^2)$ when X has pdf $f(x; \lambda_1, \lambda_2, p)$. Then compute $\text{Var}(X)$.

- (e) The *coefficient of variation* of a random variable (or distribution) is $CV = \sigma/\mu$. What is the CV for an exponential rv? What can you say about the value of CV when X has a hyperexponential distribution?
- (f) What is the CV for an Erlang distribution with parameters λ and n as defined in Sect. 3.4? [Note: In applied work, the sample CV is used to decide which of the three distributions might be appropriate.]
- (g) For the parameter values given in (b), calculate the probability that X is within one standard deviation of its mean value. Does this probability depend upon the values of the λ s (it does not depend on λ when X has an exponential distribution)?

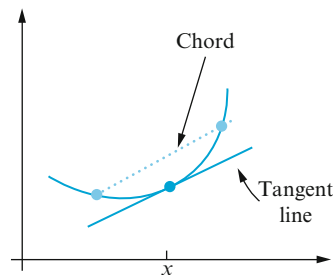
159. Suppose a state allows individuals filing tax returns to itemize deductions only if the total of all itemized deductions is at least \$5,000. Let X (in 1000s of dollars) be the total of itemized deductions on a randomly chosen form. Assume that X has the pdf

$$f(x; \alpha) = \begin{cases} k/x^\alpha & x \geq 5 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Find the value of k . What restriction on α is necessary?
 - (b) What is the cdf of X ?
 - (c) What is the expected total deduction on a randomly chosen form? What restriction on α is necessary for $E(X)$ to be finite?
 - (d) Show that $\ln(X/5)$ has an exponential distribution with parameter $\alpha - 1$.
160. Let I_i be the input current to a transistor and I_o be the output current. Then the current gain is proportional to $\ln(I_o/I_i)$. Suppose the constant of proportionality is 1 (which amounts to choosing a particular unit of measurement), so that current gain = $X = \ln(I_o/I_i)$. Assume X is normally distributed with $\mu = 1$ and $\sigma = .05$.
- (a) What type of distribution does the ratio I_o/I_i have?
 - (b) What is the probability that the output current is more than twice the input current?
 - (c) What are the expected value and variance of the ratio of output to input current?
161. The article “Response of SiC_f/Si₃N₄ Composites Under Static and Cyclic Loading—An Experimental and Statistical Analysis” (*J. Engr. Materials Tech.*, 1997: 186–193) suggests that tensile strength (MPa) of composites under specified conditions can be modeled by a Weibull distribution with $\alpha = 9$ and $\beta = 180$.
- (a) Sketch a graph of the density function.
 - (b) What is the probability that the strength of a randomly selected specimen will exceed 175? Will be between 150 and 175?
 - (c) If two randomly selected specimens are chosen and their strengths are independent of each other, what is the probability that at least one has strength between 150 and 175?
 - (d) What strength value separates the weakest 10% of all specimens from the remaining 90%?
162. (a) Suppose the lifetime X of a component, when measured in hours, has a gamma distribution with parameters α and β . Let $Y =$ lifetime measured in minutes. Derive the pdf of Y .
- (b) If X has a gamma distribution with parameters α and β , what is the probability distribution of $Y = cX$?
163. Based on data from a dart-throwing experiment, the article “Shooting Darts” (*Chance*, Summer 1997: 16–19) proposed that the horizontal and vertical errors from aiming at a point target should be independent of each other, each with a normal distribution having mean 0 and standard deviation σ . It can then be shown that the pdf of the distance V from the target to the landing point is

$$f(v) = \frac{v}{\sigma^2} \cdot e^{-v^2/(2\sigma^2)} \quad v > 0$$

- (a) This pdf is a member of what family introduced in this chapter?
- (b) If $\sigma = 20$ mm (close to the value suggested in the paper), what is the probability that a dart will land within 25 mm (roughly 1 in.) of the target?
164. The article “Three Sisters Give Birth on the Same Day” (*Chance*, Spring 2001: 23–25) used the fact that three Utah sisters had all given birth on March 11, 1998, as a basis for posing some interesting questions regarding birth coincidences.
- (a) Disregarding leap year and assuming that the other 365 days are equally likely, what is the probability that three randomly selected births all occur on March 11? Be sure to indicate what, if any, extra assumptions you are making.
- (b) With the assumptions used in part (a), what is the probability that three randomly selected births all occur on the same day?
- (c) The author suggested that, based on extensive data, the length of gestation (time between conception and birth) could be modeled as having a normal distribution with mean value 280 days and standard deviation 19.88 days. The due dates for the three Utah sisters were March 15, April 1, and April 4, respectively. Assuming that all three due dates are at the mean of the distribution, what is the probability that all births occurred on March 11? [Hint: The deviation of birth date from due date is normally distributed with mean 0.]
- (d) Explain how you would use the information in part (c) to calculate the probability of a common birth date.
165. Exercise 49 introduced two machines that produce wine corks, the first one having a normal diameter distribution with mean value 3 cm and standard deviation .1 cm and the second having a normal diameter distribution with mean value 3.04 cm and standard deviation .02 cm. Acceptable corks have diameters between 2.9 and 3.1 cm. If 60% of all corks used come from the first machine and a randomly selected cork is found to be acceptable, what is the probability that it was produced by the first machine?
166. A function $g(x)$ is *convex* if the chord connecting any two points on the function’s graph lies above the graph. When $g(x)$ is differentiable, an equivalent condition is that for every x , the tangent line at x lies entirely on or below the graph. (See the accompanying figure.) How does $g(\mu) = g[E(X)]$ compare to the expected value $E[g(X)]$? [Hint: The equation of the tangent line at $x = \mu$ is $y = g(\mu) + g'(\mu) \cdot (x - \mu)$. Use the condition of convexity, substitute X for x , and take expected values. Note: Unless $g(x)$ is linear, the resulting inequality (usually called Jensen’s inequality) is strict ($<$ rather than \leq); it is valid for both continuous and discrete rvs.]



167. Let X have a Weibull distribution with parameters $\alpha = 2$ and β . Show that $Y = 2X^2/\beta^2$ has an exponential distribution with $\lambda = 1/2$.

168. Let X have the pdf $f(x) = 1/[\pi(1 + x^2)]$ for $-\infty < x < \infty$ (a central Cauchy distribution), and show that $Y = 1/X$ has the same distribution. [Hint: Consider $P(|Y| \leq y)$, the cdf of $|Y|$, then obtain its pdf and show it is identical to the pdf of $|X|$.]
169. Let X have a Weibull distribution with shape parameter α and scale parameter β . Show that the transformed variable $Y = \ln(X)$ has an extreme value distribution as defined in Section 3.6, with $\theta_1 = \ln(\beta)$ and $\theta_2 = 1/\alpha$.
170. A store will order q gallons of a liquid product to meet demand during a particular time period. This product can be dispensed to customers in any amount desired, so demand during the period is a continuous random variable X with cdf $F(x)$. There is a fixed cost c_0 for ordering the product plus a cost of c_1 per gallon purchased. The per-gallon sale price of the product is d . Liquid left unsold at the end of the time period has a salvage value of e per gallon. Finally, if demand exceeds q , there will be a shortage cost for loss of goodwill and future business; this cost is f per gallon of unfulfilled demand. Show that the value of q that maximizes expected profit, denoted by q^* , satisfies

$$P(\text{satisfying demand}) = F(q^*) = \frac{d - c_1 + f}{d - e + f}$$

Then determine the value of $F(q^*)$ if $d = \$35$, $c_0 = \$25$, $c_1 = \$15$, $e = \$5$, and $f = \$25$. [Hint: Let x denote a particular value of X . Develop an expression for profit when $x \leq q$ and another expression for profit when $x > q$. Now write an integral expression for expected profit (as a function of q) and differentiate.]

171. An individual's credit score is a number calculated based on that person's credit history that helps a lender determine how much s/he should be loaned or what credit limit should be established for a credit card. An article in the *Los Angeles Times* gave data which suggested that a beta distribution with parameters $A = 150$, $B = 850$, $\alpha = 8$, $\beta = 2$ would provide a reasonable approximation to the distribution of American credit scores. [Note: credit scores are integer-valued.]
- Let X represent a randomly selected American credit score. What are the mean and standard deviation of this random variable? What is the probability that X is within 1 standard deviation of its mean?
 - What is the approximate probability that a randomly selected score will exceed 750 (which lenders consider a very good score)?
172. Let V denote rainfall volume and W denote runoff volume (both in mm). According to the article "Runoff Quality Analysis of Urban Catchments with Analytical Probability Models" (*J. of Water Resource Planning and Management*, 2006: 4–14), the runoff volume will be 0 if $V \leq v_d$ and will be $k(V - v_d)$ if $V > v_d$. Here v_d is the volume of depression storage (a constant), and k (also a constant) is the runoff coefficient. The cited article proposes an exponential distribution with parameter λ for V .
- Obtain an expression for the cdf of W . [Note: W is neither purely continuous nor purely discrete; instead it has a "mixed" distribution with a discrete component at 0 and is continuous for values $w > 0$.]
 - What is the pdf of W for $w > 0$? Use this to obtain an expression for the expected value of runoff volume.