CHAPTER 7

CONTINUOUS RANDOM VARIABLES

When we have a continuous random variable, we believe all values over some range are possible if our measurement device is sufficiently accurate. There is an uncountably infinite number of real numbers in an interval, so the probability of getting any particular value must be zero. This makes it impossible to find the probability function of a continuous random variable the same way we did for a discrete random variable. We will have to find a different way to determine its probability distribution. First we consider a thought experiment similar to those done in Chapter 5 for discrete random variables.

Thought Experiment 4 *We start taking a sequence of independent trials of the random variable. We sketch a graph with a spike at each value in the sample equal to the proportion in the sample having that value. After each draw we update the proportions in the accumulated sample that have each value, and update our graph. The updating of the graph at step n is made by scaling all the existing spikes down by the ratio $\frac{n-1}{n}$ and adding $\frac{1}{n}$ to the spike at the value observed at trial n. This keeps the sum of the spike heights equal to 1. Figure 7.1 shows this after 25 draws. Because there are infinitely many possible numbers, it is almost inevitable that we do not draw any of the previous values, so we get a new spike at each draw. After n draws we will* *have n spikes, each having height $\frac{1}{n}$. Figure 7.2 shows this after 100 draws. As the sample size, n, approaches infinity, the heights of the spikes shrink to zero. This means the probability of getting any particular value is zero. The output of this thought experiment is not the probability function, which gives the probability of each possible value. This is not like the output of the thought experiments in Chapter 6 where the random variable was discrete.*

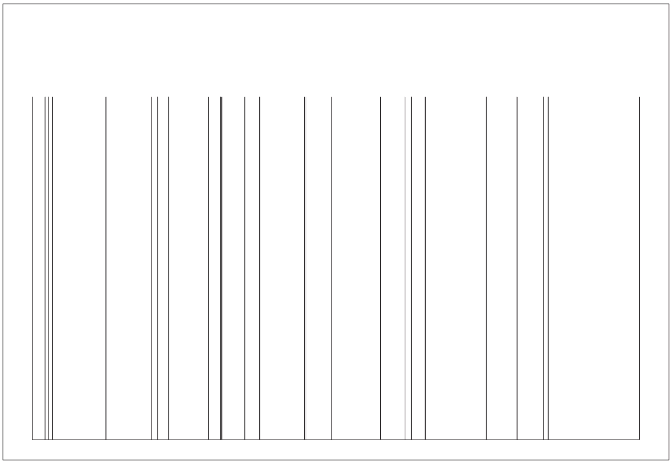

Figure 7.1 Sample probability function after 25 draws.

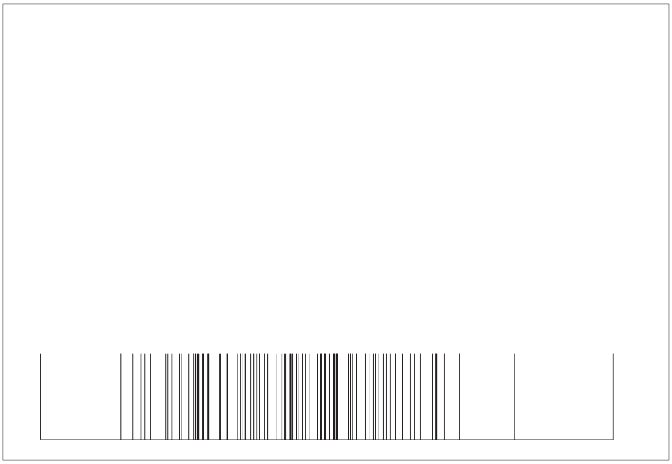

Figure 7.2 Sample probability function after 100 draws.

What we do notice is that there are some places with many spikes close by, and there are other places with very few spikes close by. In other words, the density of spikes varies. We can think of partitioning the interval into subintervals, and recording the number of observations that fall into each subinterval. We can form a density histogram by dividing the number in each subinterval by the width of the subinterval. This makes the area under the histogram equal to one. Figure 7.3 shows the density histogram for the first 100 observations. Now let n increase, and let the width of the subintervals

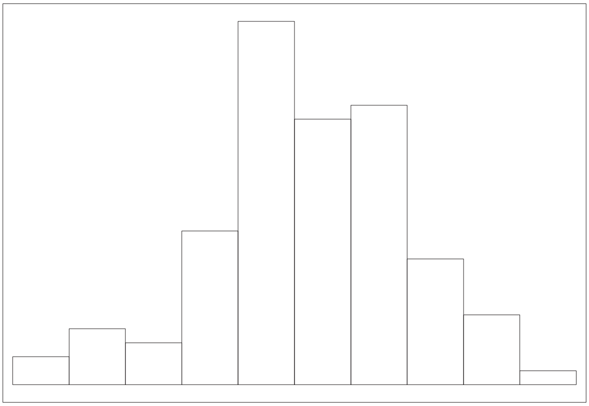

Figure 7.3 Density histogram after 100 draws.

decrease, but at a slower rate than n. Figures 7.4 and 7.5 show the density histogram for the first 1,000 and for the first 10,000 observations, respectively. The proportion of observations in a subinterval approaches the probability of being in the subinterval. As n increases, we get a larger number of shorter subintervals. The histograms get closer and closer to a smooth curve.

7.1 Probability Density Function

The smooth curve is called the probability density function (pdf). It is the limiting shape of the histograms as n goes to infinity, and the width of the bars goes to 0. Its height at a point is not the probability of that point. The thought experiment showed us that probability was equal to zero at every point. Instead, the height of the curve measures how *dense* is the probability at that point.

Since the areas under the histograms all equaled one, the total area under the probability density function must also equal 1:

$$\int_{-\infty}^{\infty} f(y)\, dy = 1\,. \tag{7.1}$$

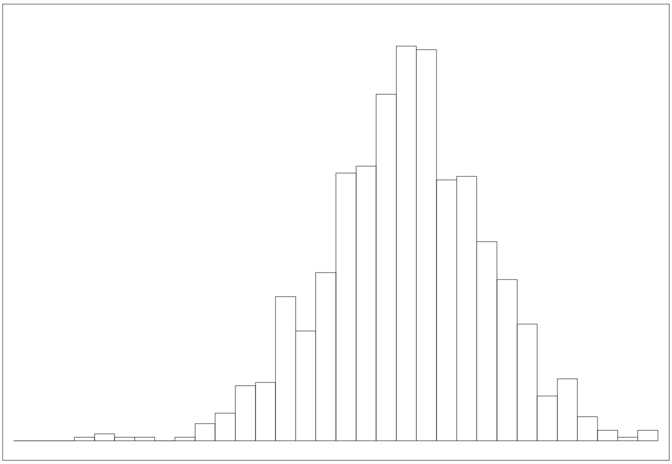

Figure 7.4 Density histogram after 1,000 draws.

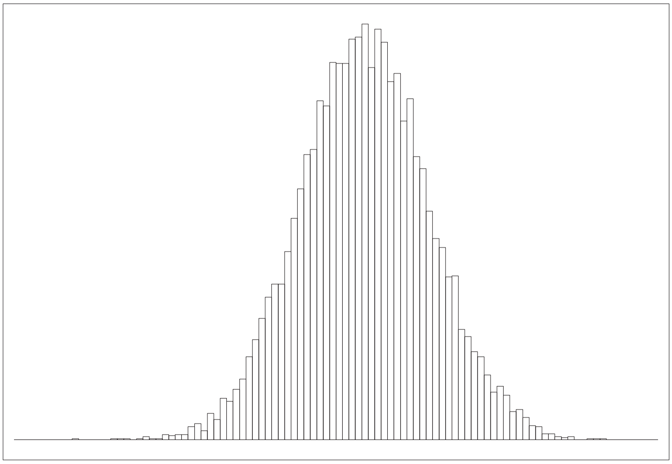

Figure 7.5 Density histogram after 10,000 draws.

The proportion of the observations that lie in an interval (a, b) is given by the area of the histogram bars that lie in the interval. In the limit as n increases to infinity, the histograms become the smooth curve, the probability density function. The area of the bars that lie in the interval becomes the area under the curve over that interval. The proportion of observations that lie in the interval becomes the probability that the random variable lies in the interval.

We know the area under a curve is found by integration, so we can find the probability that the random variable lies in the interval (a, b) by integrating the probability density function over that range:

$$P(a < Y < b) = \int_a^b f(y)\, dy\,. \tag{7.2}$$

Mean of a Continuous Random Variable

In Section 3.3 we defined the mean of the random sample of observations from the random variable to be

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}\,.$$

Suppose we put the observations in a density histogram where all groups have equal width. The grouped mean of the data is

$$\bar{y} = \sum_j m_j \frac{n_j}{n}\,,$$

where m_j is the midpoint of the j^{th} bar and $\frac{n_j}{n}$ is its relative frequency. Multiplying and dividing by the width of the bars, we get

$$\bar{y} = \sum_j m_j \times width \times \frac{n_j}{n \times width}\,,$$

where the relative frequency density $\frac{n_j}{n \times width}$ gives the height of bar j. Multiplying it by *width* gives the area of the bar. Thus the sample mean is the midpoint of each bar times the area of that bar summed over all bars.

Suppose we let n increase without bound, and let the number of bars increase, but at a slower rate. For example, as n increases by a factor of 4, we let the number of bars increase by a factor of 2 so the width of each bar is divided by 2. As n increases without bound, each observation in a group becomes quite close to the midpoint of the group, the number of bars increase without bound, and the width of each bar goes to zero. In the limit, the midpoint of the bar containing the point y approaches y, and the height of the bar containing point y (which is the relative frequency density) approaches $f(y)$. So, in the limit, the relative frequency density approaches the probability density and the sample mean reaches its limit

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y f(y)\, dy\,, \tag{7.3}$$

which is called the *expected value* of the random variable. The expected value is like the mean of all possible values of the random variable. Sometimes it is referred to as the mean of the random variable Y and denoted μ.

Variance of a Continuous Random Variable

The expected value $\mathrm{E}[(Y - E[Y])^2]$ is called the variance of the random variable. We can look at the variance of a random sample of numbers and let the sample size increase.

$$\mathrm{Var}[y] = \frac{1}{n} \times \sum_{i=1}^{n} (y_i - \bar{y})^2 .$$

As we let n increase, we decrease the width of the bars. This makes each observation become closer to the midpoint of the bar it is in. Now, when we sum over all groups, the variance becomes

$$\mathrm{Var}[y] = \sum_j \frac{n_j}{n} (m_j - \bar{y})^2 .$$

We multiply and divide by the width of the bar to get

$$\mathrm{Var}[y] = \sum_j \frac{n_j}{n \times width} \times width \times (m_j - \bar{y})^2 .$$

This is the square of the midpoint minus the mean times the area of the bar summed over all bars. As n increases to ∞, the relative frequency density approaches the probability density, the midpoint of the bar containing the point y approaches y, and the sample mean $\bar{y}$ approaches the expected value $\mathrm{E}[Y]$, so in the limit the variance becomes

$$\mathrm{Var}[Y] = E[(Y - \mathrm{E}[Y])^2] = \int_{-\infty}^{\infty} (y - \mu)^2 f(y)\, dy . \tag{7.4}$$

The variance of the random variable is denoted σ^2. We can square the term in brackets,

$$\mathrm{Var}[Y] = \int_{-\infty}^{\infty} (y^2 - 2\mu y + \mu^2) f(y)\, dy ,$$

break the integral into three terms,

$$\mathrm{Var}[Y] = \int_{-\infty}^{\infty} y^2\, f(y)\, dy - 2\mu \int_{-\infty}^{\infty} y\, f(y)\, dy + \mu^2 \int_{-\infty}^{\infty} f(y)\, dy ,$$

and simplify to get an alternate form for the variance:

$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - [\mathrm{E}[Y]]^2 . \tag{7.5}$$

7.2 Some Continuous Distributions

Uniform Distribution

The random variable has the *uniform* $(0, 1)$ distribution if its probability density function is constant over the interval [0,1], and 0 everywhere else.

$$g(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1, \\ 0 & \text{for } x \notin [0, 1] \end{cases}$$

It is easily shown that the mean and variance of a uniform (0,1) random variable are $\frac{1}{2}$ and $\frac{1}{12}$, respectively.

Beta Family of Distributions

The *beta*(a, b) distribution is another commonly used distribution for a continuous random variable that can only take on values $0 \leq x \leq 1$. It has the probability density function

$$g(x; a, b) = \begin{cases} k \times x^{a-1}(1-x)^{b-1} & \text{for } 0 \leq x \leq 1, \\ 0 & \text{for } x \notin [0, 1] \end{cases}$$

The most important thing is that $x^{a-1}(1-x)^{b-1}$ determines the shape of the curve, and k is only the constant needed to make this a probability density function. Figure 7.6 shows the graphs of this for $a = 2$ and $b = 3$ for a number of values of k. We see that the curves all have the same basic shape but have different areas under the curves. The value of $k = 12$ gives area equal to 1, so that is the one that makes a density function. The distribution with shape given by $x^{a-1}(1-x)^{b-1}$ is called the *beta*(a, b) distribution. The constant needed to make the curve a density function is given by the formula

$$k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)},$$

where $\Gamma(c)$ is the Gamma function, which is a generalization of the factorial function.[1] The probability density function of the *beta*(a, b) distribution is given by

$$g(x; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}. \tag{7.6}$$

All we need remember is that $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ is the constant needed to make the curve with shape given by $x^{a-1}(1-x)^{b-1}$ a density. a equals one plus the power of x, and b equals one plus the power of $(1-x)$.

[1]When c is an integer, $\Gamma(c) = (c-1)!$. The Gamma function always satisfies the equation $\Gamma(c) = (c-1) \times \Gamma(c-1)$ whether or not c is an integer.

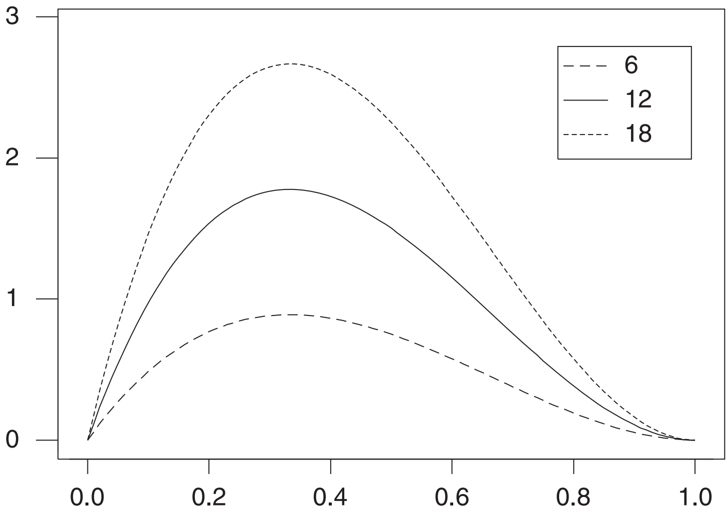

Figure 7.6 The curve $g(x) = kx^1(1-x)^2$ for several values of k.

This curve can have different shapes depending on the values a and b, so the *beta*(a, b) is actually a family of distributions. The *uniform(0,1)* distribution is a special case of the *beta*(a, b) distribution, where $a = 1$ and $b = 1$.

Mean of a beta distribution. The expected value of a continuous random variable x is found by integrating the variable times the density function over the whole range of possible values. (Since the *beta*(a, b) density equals 0 for x outside the interval $[0, 1]$, the integration only has to go from 0 to 1, not $-\infty$ to ∞.) For a random variable having the *beta*(a, b) distribution,

$$\mathrm{E}[X] = \int_0^1 x \times g(x; a, b) dx = \int_0^1 x \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1}\, dx\,.$$

However, by using our understanding of the *beta* distribution, we can evaluate this integral without having to do the integration. First move the constant out in front of the integral, then combine the x terms by adding exponents:

$$\mathrm{E}[X] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x \times x^{a-1}(1-x)^{b-1} dx = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^a (1-x)^{b-1}\, dx\,.$$

We recognize the part under the integral sign as a curve that has the *beta*$(a + 1, b)$ shape. So we must multiply inside the integral by the appropriate constant to make it integrate to 1, and multiply by the reciprocal of the constant outside of the integral to keep the balance:

$$\mathrm{E}[X] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \int_0^1 \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} x^a(1-x)^{b-1}\, dx\,.$$

The integral equals 1, and when we use the fact that $\Gamma(c) = (c-1) \times \Gamma(c-1)$ and do some cancellation, we get the simple formula

$$\mathrm{E}[X] = \frac{a}{a+b} \tag{7.7}$$

for the mean of a $beta(a, b)$ random variable.

Variance of a beta distribution. The expected value of a *function* of a continuous random variable is found by integrating the *function* times the density function over the whole range of possible values. For a random variable having the $beta(a, b)$ distribution,

$$\mathrm{E}[X^2] = \int_0^1 x^2 \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1}\, dx\,.$$

When we evaluate this integral using the properties of the $beta(a, b)$ distribution, we get

$$\mathrm{E}[X^2] = \frac{a(a+1)}{(a+b+1)(a+b)}\,.$$

When we substitute this formula and the formula for the mean of the $beta(a, b)$ into Equation 7.5 and simplify, we find the variance of the random variable having the $beta(a, b)$ distribution is given by

$$\mathrm{Var}[X] = \frac{ab}{(a+b)^2(a+b+1)}\,. \tag{7.8}$$

Finding beta probabilities. When X has the $beta(a, b)$ distribution, we often want to calculate probabilities such as

$$P(X \le x_0) = \int_0^{x_0} g(x; a, b)\, dx\,.$$

[Minitab:] This can easily be done in Minitab. Pull down the *Calc* menu to *Probability Distributions* command, over to *Beta...* subcommand, and fill out the dialog box.

Gamma Family of Distributions

The $gamma(r, v)$ distribution is used for continuous random variables that can take on nonnegative values $0 \le x < \infty$. Its probability density function is given by

$$g(x; r, v) = k \times x^{r-1} e^{-vx} \text{ for } 0 \le x < \infty\,.$$

The shape of the curve is determined by $x^{r-1}e^{-vx}$, while k is only the constant needed to make this a probability density. Figure 7.7 shows the graphs of this for the case where $r = 4$ and $v = 4$ for several values of k. Clearly the curves

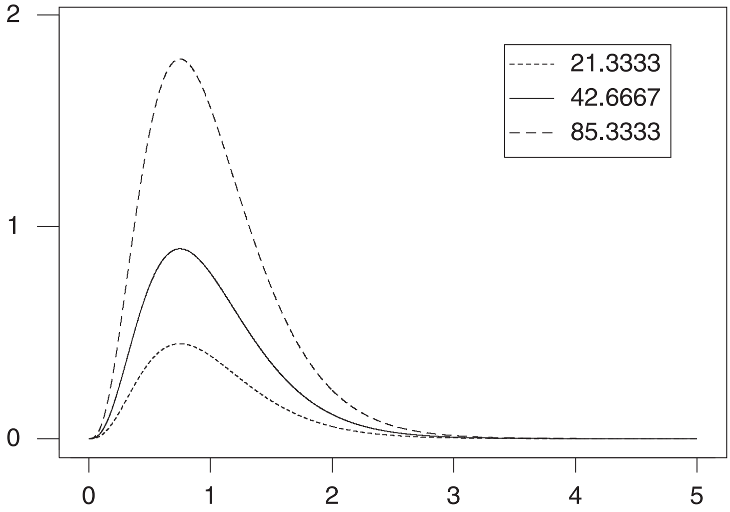

Figure 7.7 The curve $g(x) = kx^3 e^{-4x}$ for several values of k.

have the same basic shape, but have different areas under the curve. The curve with $k = 42.6667$ will have area equal to 1, so it is the exact density.

The distribution having shape given by $x^{r-1}e^{-vx}$ is called the $gamma(r, v)$ distribution. The constant needed to make this a probability density function is given by

$$k = \frac{v^r}{\Gamma(r)},$$

where $\Gamma(r)$ is the Gamma function. The probability density of the $gamma(r, v)$ distribution is given by

$$g(x; r, v) = \frac{v^r x^{r-1} e^{-vx}}{\Gamma(r)} \tag{7.9}$$

for $0 \leq x < \infty$.

Mean of Gamma distribution. The expected value of a $gamma(r, v)$ random variable x is found by integrating the variable x times its density function over the whole range of possible values. It will be

$$\begin{aligned} \mathrm{E}[X] &= \int_0^\infty x g(x; r, v)\, dx \\ &= \int_0^\infty x \frac{v^r x^{r-1} e^{-vx}}{\Gamma(r)}\, dx \\ &= \frac{v^r}{\Gamma(r)} \int_0^\infty x^r e^{-vx}\, dx\,. \end{aligned}$$

We recognize the part under the integral to be a curve that has the shape of a $gamma(r+1, v)$ distribution. We multiply inside the integral by the appropriate constant to make it integrate to 1, and outside the integral we multiply by its reciprocal to keep the balance.

$$\mathrm{E}[X] = \frac{v^r}{\Gamma(r)} \times \frac{\Gamma(r+1)}{v^{r+1}} \int_0^\infty \frac{v^{r+1}}{\Gamma(r+1)} x^r e^{-vx}\, dx\,.$$

This simplifies to give

$$\mathrm{E}[X] = \frac{r}{v}\,. \tag{7.10}$$

Variance of a gamma distribution. First we find

$$\begin{aligned}\mathrm{E}[X^2] &= \int_0^\infty x^2 g(x; r, v)\, dx \\ &= \frac{v^r}{\Gamma(r)} \int_0^\infty x^{r+1} e^{-vx}\, dx\,.\end{aligned}$$

We recognize the shape of a $gamma(r+2, v)$ under the curve, so this simplifies to

$$\mathrm{E}[X^2] = \frac{(r+1)r}{v^2}\,.$$

When we substitute this, and the formula for the mean of the $gamma(r, v)$ into Equation 7.5 and simplify we find the variance of the $gamma(r, v)$ distribution to be

$$\mathrm{Var}[X] = \frac{r}{v^2}\,. \tag{7.11}$$

Finding gamma probabilities. When X has the $gamma(r, v)$ distribution we often want to calculate probabilities such as

$$P(X \le x_0) = \int_0^{x_0} g(x; r, v)\, dx\,.$$

This can easily be done in Minitab. Pull down the *Calc* menu to *Probability Distributions* command, over to *Gamma...* subcommand, and fill out the dialog box. Note: In Minitab the shape parameter is r and the scale parameter is $\frac{1}{v}$.

Normal Distribution

Very often data appear to have a symmetric bell-shaped distribution. In the early years of statistics, this shape seemed to occur so frequently that it was thought to be normal. The family of distributions with this shape has become known as the *normal* distribution family. It is also known as the *Gaussian* distribution after the mathematician Gauss, who studied its properties. It is the most widely used distribution in statistics. We will see that there is a good reason for its frequent occurrence. However, we must remain aware that the term *normal distribution* is only a name, and distributions with other shapes are not abnormal.

The $normal(\mu, \sigma^2)$ distribution is the member of the family having mean μ and variance σ^2. The probability density function of a $normal(\mu, \sigma^2)$ distribution is given by

$$g(x|\mu, \sigma^2) = ke^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

for $-\infty < x < \infty$, where k is the constant value needed to make this a probability density. The shape of the curve is determined by $e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. Figure 7.8 shows the curve $ke^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ for several values of k. Changing the value of k only changes the area under the curve, not its basic shape. To be a probability density function, the area under the curve must equal 1. The value of k that makes the curve a probability density is $k = \frac{1}{\sqrt{2\pi}\sigma}$.

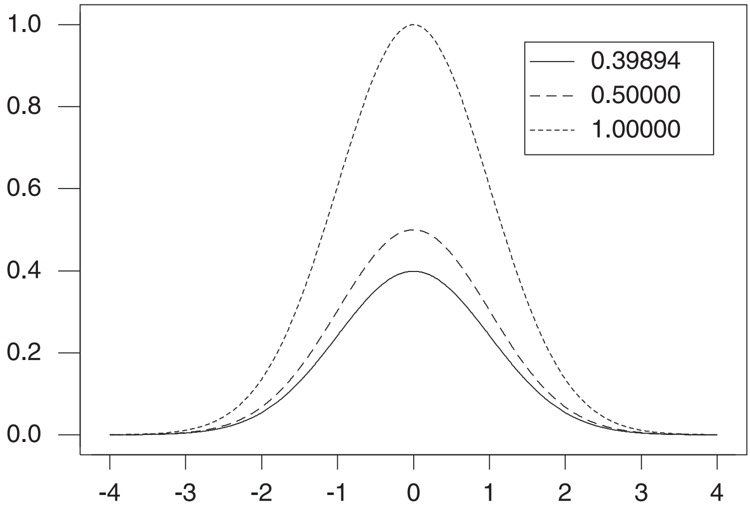

Figure 7.8 The curve $g(x) = ke^{-\frac{1}{2}(x-0)^2}$ for several values of k.

Central limit theorem. The central limit theorem says that if you take a random sample $y_1, \ldots, y_n$ from any shape distribution having mean μ and variance σ^2, then the limiting distribution of $\frac{\bar{y}-\mu}{\sigma/\sqrt{n}}$ is $normal(0, 1)$. The shape of the limiting distribution is *normal* despite the original distribution not necessarily being normal. A linear transformation of a normal distribution is also normal, so the shape of $\bar{y}$ and $\sum y$ are also normal. Amazingly, n does not have to be particularly large for the shape to be approximately normal, $n \geq 25$ is sufficient.

The key factor of the central limit distribution is that when we are averaging a large number of independent effects, each of which is small in relation to the

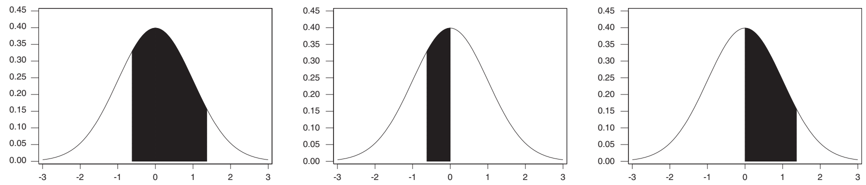

Figure 7.9 The area between $-.62$ and 1.37 split into two parts.

sum, the distribution of the sum approaches the *normal* shape regardless of the shapes of the individual distributions. Thus any random variable that arises as the sum of a large number of independent effects will be approximately normal! This explains why the normal distribution is encountered so frequently.

Finding probabilities using standard normal table. The standard normal density has mean $\mu = 0$ and variance $\sigma^2 = 1$. Its probability density function is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} .$$

We note that this curve is symmetric about $z = 0$. Unfortunately, Equation 7.2, the general form for finding the probability $P(a \leq z \leq b)$, is not of any practical use here. There is no closed form for integrating the standard normal probability density function. Instead, the area between 0 and z for values of z between 0 and 3.99 has been numerically calculated and tabulated in Table B.2 in Appendix B. We use this table to calculate the probability we need.

EXAMPLE 7.1

Suppose we want to find $P(-.62 \leq Z \leq 1.37)$. In Figure 7.9 we see that the shaded area between $-.62$ and 1.37 is the sum of the two areas between $-.62$ and 0 and between 0 and 1.37, respectively. The area between $-.62$ and 0 is the same as the area between 0 and $+.62$ because the standard normal density is symmetric about 0. In Table B.2 we find this area equals .2324. The area between 0 and 1.37 equals .4147 from the table. So

$$\begin{aligned} P(-.62 \leq Z \leq 1.37) &= .2324 + .4147 \\ &= .6471 . \end{aligned}$$

∎

Any normal distribution can be transformed into a standard normal by subtracting the mean and then dividing by the standard deviation. This lets us find any normal probability using the areas under the standard normal density found in Table B.2.

EXAMPLE 7.2

Suppose we know Y is normal with mean $\mu = 10.8$ and standard deviation $\sigma = 2.1$, and suppose we want to find the probability $P(Y \geq 9.9)$.

$$\begin{aligned} P(Y \geq 9.9) &= P(Y - 10.8 \geq 9.9 - 10.8) \\ &= P\left(\frac{Y - 10.8}{2.1} \geq \frac{9.9 - 10.8}{2.1}\right). \end{aligned}$$

The left side is a standard normal. The right side is a number. We find this probability from the standard normal:

$$\begin{aligned} P(Y \geq 9.9) &= P(Z \geq -.429) \\ &= .1659 + .5000 \\ &= .6659\,. \end{aligned}$$

■

Finding beta probabilities using normal approximation. We can approximate a *beta*(a, b) distribution by the *normal* distribution having the same mean and variance. This approximation is very effective when both a and b are greater than or equal to ten.

EXAMPLE 7.3

Suppose Y has the *beta*$(12, 25)$ distribution and we wish to find $P(Y > .4)$. The mean and variance of Y are

$$\mathrm{E}[Y] = \frac{12}{37} = .3243 \quad \text{and} \quad \mathrm{Var}[Y] = \frac{12 \times 25}{37^2 \times 38} = .005767\,,$$

respectively. We approximate the *beta*$(12, 25)$ distribution with a *normal*$(.3243, .005767)$ distribution. The approximate probability is

$$\begin{aligned} P(Y > .4) &= P\left(\frac{Y - .3243}{\sqrt{.005767}} > \frac{.4 - .3243}{\sqrt{.005767}}\right) \\ &= P(Z > .997) \\ &= .1594\,. \end{aligned}$$

■

Finding gamma probabilities using normal approximation is not recommended. As r approaches infinity the *gamma*(r, v) distribution approaches the *normal*(m, s^2) distribution where $m = \frac{r}{v}$ and $s^2 = \frac{r}{v^2}$. However, the approach is very slow, and the gamma probabilities calculated using the normal approximation will not be very accurate unless r is quite large (Johnson et al., 1970). Johnson et al. (1970) recommend that the normal approximation to the gamma not be used for this reason, and they give other approximations that are more accurate.

7.3 Joint Continuous Random Variables

We consider two (or more) random variables distributed together. If both X and Y are continuous random variables, they have joint density $f(x, y)$, which measures the probability density at the point (x, y). This would be found by dividing the plane into rectangular regions by partitioning both the x axis and y axis. We look at the proportion of the sample that lie in a region. We increase n, the sample size of the joint random variables without bound, and at the same time decrease the width of the regions (in both dimensions) at a slower rate. In the limit, the proportion of the sample lying in the region centered at (x, y) approaches the joint density $f(x, y)$. Figure 7.10 shows a joint density function.

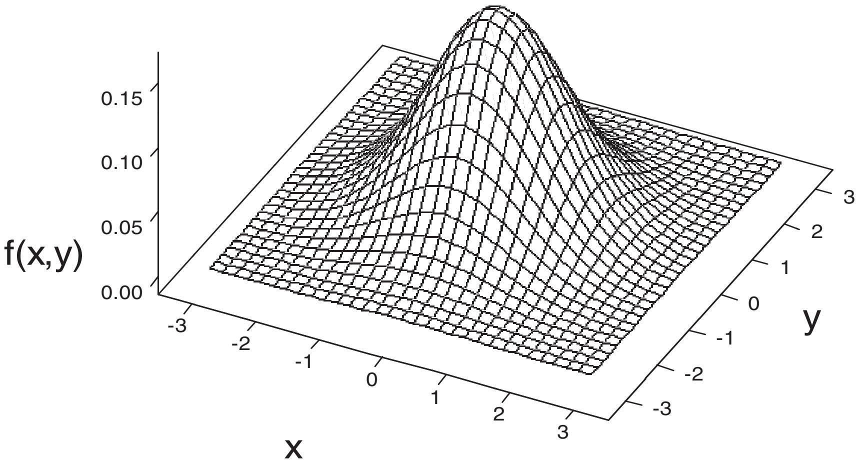

Figure 7.10 A joint density.

We might be interested in determining the density of one of the joint random variables by itself, its *marginal* density. When X and Y are joint random variables that are both continuous, the marginal density of Y is found by integrating the joint density over the whole range of X:

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

and vice versa. (Finding the marginal density by integrating the joint density over the whole range of one variable is analogous to finding the marginal probability distribution by summing the joint probability distribution over all possible values of one variable for jointly distributed discrete random variables.)

Conditional Probability Density

The conditional density of X given $Y = y$ is given by

$$f(x|y) = \frac{f(x,y)}{f(y)}\,.$$

We see that the conditional density of X given $Y = y$ is proportional to the joint density where $Y = y$ is held fixed. Dividing by the marginal density $f(y)$ makes the integral of the conditional density over the whole range of x equal 1. This makes it a proper density function.

7.4 Joint Continuous and Discrete Random Variables

It may be that one of the variables is continuous, and the other is discrete. For instance, let X be continuous, and let Y be discrete. In that case, $f(x, y_j)$ is a joint probability–probability density function. In the x direction it is continuous, and in the y direction it is discrete. This is shown in Figure 7.11.

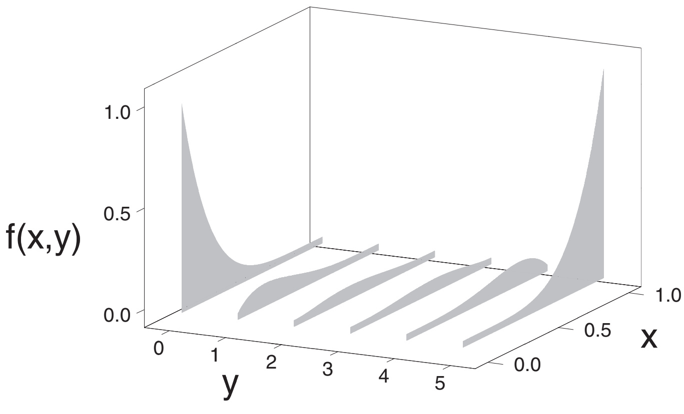

Figure 7.11 A joint continuous and discrete distribution.

In this case, the marginal density of the continuous random variable X is found by

$$f(x) = \sum_j f(x, y_j)\,,$$

and the marginal probability function of the discrete random variable Y is found by

$$f(y_j) = \int f(x, y_j)\,dx\,.$$

The conditional density of X given $Y = y_j$ is given by

$$f(x|y_j) = \frac{f(x, y_j)}{f(y_j)} = \frac{f(x, y_j)}{\int f(x, y_j)\, dx}\,.$$

We see that this is proportional to the joint probability–probability density function $f(x, y_j)$ where x is allowed to vary over its whole range. Dividing by the marginal probability $f(y_j)$ just scales it to be a proper density function (integrates to 1).

Similarly, the conditional distribution of $Y = y_j$ given x is found by

$$f(y_j|x) = \frac{f(x, y_j)}{f(x)} = \frac{f(x, y_j)}{\sum_j f(x, y_j)}\,.$$

This is also proportional to the joint probability–probability density function $f(x, y_j)$ where x is fixed, and Y is allowed to take on all the possible values $y_1, \ldots, y_J$.

Main Points

- The probability that a continuous random variable equals any particular value is zero!

- The probability density function of a continuous random variable is a smooth curve that measures the *density* of probability at each value. It is found as the limit of density histograms of random samples of the random variable, where the sample size increases to infinity and the width of the bars goes to zero.

- The probability a continuous random variable lies between two values a and b is given by the area under the probability density function between the two values. This is found by the integral

$$P(a < X < b) = \int_a^b f(x)\, dx\,.$$

- The expected value of a continuous random variable X is found by integrating x times the density function $f(x)$ over the whole range.

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

- A *beta*(a, b) random variable has probability density

$$f(x|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, x^{a-1}(1-x)^{b-1} \quad \text{for} \quad 0 \le x \le 1\,.$$

- The mean and variance of a $beta(a, b)$ random variable are given by

$$E[X] = \frac{a}{a+b} \qquad \text{and} \qquad Var[X] = \frac{a \times b}{(a+b)^2 \times (a+b+1)} .$$

- A $gamma(r, v)$ random variable has probability density

$$g(x; r, v) = \frac{v^r x^{r-1} e^{-vx}}{\Gamma(r)} \text{ for } 0 \le x < \infty .$$

- The mean and variance of a $gamma(r, v)$ random variable are given by

$$E[X] = \frac{r}{v} \qquad \text{and} \qquad Var[X] = \frac{r}{v^2} .$$

- A $normal(\mu, \sigma^2)$ random variable has probability density

$$g(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} ,$$

where μ is the mean, and σ^2 is the variance.

- The central limit theorem says that for a random sample $y_1, \ldots y_n$ from any distribution $f(y)$ having mean μ and variance σ^2, the distribution of

$$\frac{\bar{y} - \mu}{\sigma / \sqrt{n}}$$

is approximately $normal(0, 1)$ for $n > 25$. This is regardless of the shape of the original density $f(y)$.

- By reasoning similar to that of the central limit theorem, any random variable that is the sum of a large number of independent random variables will be approximately normal. This is the reason why the normal distribution occurs so frequently.

- The marginal distribution of y is found by integrating the joint distribution $f(x, y)$ with respect to x over its whole range.

- The conditional distribution of x given y is proportional to the joint distribution $f(x, y)$ where y fixed and x is allowed to vary over its whole range.

$$f(x|y) = \frac{f(x, y)}{f(y)} .$$

Dividing by the marginal distribution of $f(y)$ scales it properly so that $f(y|x)$ integrates to 1 and is a probability density function.

Exercises

7.1. Let X have a $beta(3,5)$ distribution.

(a) Find $\text{E}[X]$.

(b) Find $\text{Var}[X]$.

7.2. Let X have a $beta(12,4)$ distribution.

(a) Find $\text{E}[X]$.

(b) Find $\text{Var}[X]$.

7.3. Let X have the *uniform* distribution.

(a) Find $\text{E}[X]$.

(b) Find $\text{Var}[X]$.

(c) Find $P(X \leq .25)$.

(d) Find $P(.33 < X < .75)$.

7.4. Let X be a random variable having probability density function

$$f(x) = 2x \quad \text{for} \quad 0 \leq x \leq 1\,.$$

(a) Find $P(X \geq .75)$.

(b) Find $P(.25 \leq X \leq .6)$.

7.5. Let Z have the standard normal distribution.

(a) Find $P(0 \leq Z \leq .65)$.

(b) Find $P(Z \geq .54)$.

(c) Find $P(-.35 \leq Z \leq 1.34)$.

7.6. Let Z have the standard normal distribution.

(a) Find $P(0 \leq Z \leq 1.52)$.

(b) Find $P(Z \geq 2.11)$.

(c) Find $P(-1.45 \leq Z \leq 1.74)$.

7.7. Let Y be normally distributed with mean $\mu = 120$ and variance $\sigma^2 = 64$.

(a) Find $P(Y \leq 130)$.

(b) Find $P(Y \geq 135)$.

(c) Find $P(114 \leq Y \leq 127)$.

7.8. Let Y be normally distributed with mean $\mu = 860$ and variance $\sigma^2 = 576$.

(a) Find $P(Y \leq 900)$.

(b) Find $P(Y \geq 825)$.

(c) Find $P(840 \leq Y \leq 890)$.

7.9. Let Y be distributed according to the $beta(10, 12)$ distribution.

(a) Find $\mathrm{E}[Y]$.

(b) Find $\mathrm{Var}[Y]$.

(c) Find $P(Y > .5)$ using the normal approximation.

7.10. let Y be distributed according to the $beta(15, 10)$ distribution.

(a) Find $\mathrm{E}[Y]$.

(b) Find $\mathrm{Var}[Y]$.

(c) Find $P(Y < .5)$ using the normal approximation.

7.11. Let Y be distributed according to the $gamma(12, 4)$ distribution.

(a) Find $\mathrm{E}[Y]$.

(b) Find $\mathrm{Var}[Y]$.

(c) Find $P(Y \leq 4)$

7.12. Let Y be distributed according to the $gamma(26, 5)$ distribution.

(a) Find $\mathrm{E}[Y]$.

(b) Find $\mathrm{Var}[Y]$.

(c) Find $P(Y > 5)$