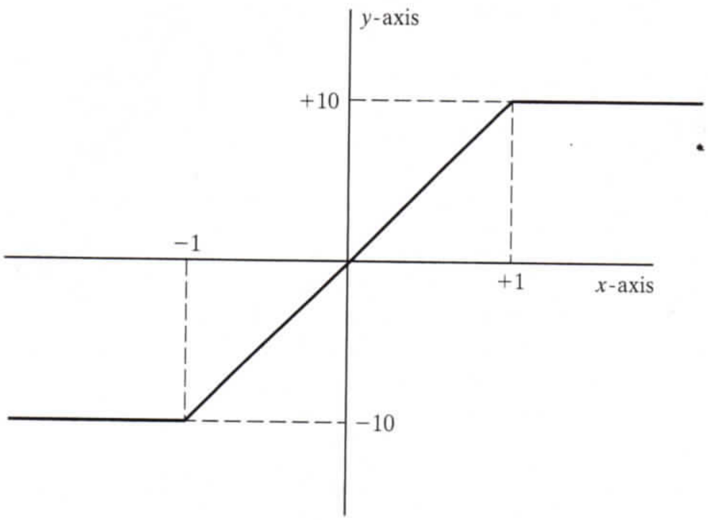

Figure 3.1.15 The amplifier has a gain of 10 over the output region from −10 to +10, but those are the limits. Thus all inputs greater in magnitude than 1 will drive the amplifier into the nonlinear region.

3.1.9 Mixed Random Variables

Chapter 2 dealt with discrete random variables, and Chapter 3 has dealt with continuous random variables. Random systems that include both are described by *mixed* random variables. For example, amplifiers have "rails" owing to power supply limits. The gain characteristic of an amplifier with rails is shown in Fig. 3.1.15. This characteristic can produce a mixed random variable, as we now explore.

The input to the amplifier, X, will be uniformly distributed between −1.2 and +1.3. Thus the PDF of the input is

$$f_X(x) = \frac{1}{2.5}, -1.2 < x \leq +1.3, \text{ zow} \quad (3.1.57)$$

We will use the transformation in Fig. 3.1.15 to find the PDF of the output, Y. First, we determine the CDF of Y. As shown in Eq. (3.1.41), we should first determine the critical regions. The definition of the CDF of Y is

$$F_Y(y) = P[Y \leq y] \quad (3.1.58)$$

The easy regions in this case are (1) $y < -10$, where the CDF is zero, since it is impossible for Y to be less than any y in this region, and (2) $y \geq +10$, where the CDF is 1, since all values of Y are +10 and below.

The discrete value y
to the negative saturation

$F_Y(-10)$

Thus the CDF jumps fro
mass at this value of y.
The next critical regi
end because $y = +10$ is
the event $\{Y \leq y\}$ is the s

$F_Y(y) = P[Y \leq y$

The next special case is y
be 1, as stated earlier. Thu
a jump of the CDF at $y =$

Thus the CDF jumps by
value of y. We may now p
that the output is a mixed

The PDF of *Y*. We r
PDF of Y. The discontinui
unless we resort to the use
a short review on unit step
Using the unit step not

$F_Y(y) = \frac{y/10 -}{}$

The first unit step function
it off. The third unit step fu
the derivative of Eq. (3.1.62
unit step functions). With th

$f_Y(y) = s(-10)\delta(y - (-10$

$= s(-10)\delta(y - (-10$

The discrete value $y = -10$ will constitute a region in this case, since this value corresponds to the negative saturation region. Here the result is

$$F_Y(-10) = P[Y \leq -10] = P[X \leq -1] = \frac{-1-(-1.2)}{2.5} = \frac{0.2}{2.5} \tag{3.1.59}$$

Thus the CDF jumps from 0 to 0.08 at exactly $y = -10$, which is indicative of a probability mass at this value of y.

The next critical region is $-10 < y < +10$. Here we do not put an equal sign on the upper end because $y = +10$ is the upper saturation region and is a special case. In this middle region the event $\{Y \leq y\}$ is the same as the event $\{X \leq \frac{y}{10}\}$. Thus the required probability calculation is

$$F_Y(y) = P[Y \leq y] = P[X \leq \frac{y}{10}] = \int_{x=-\infty}^{x=y/10} f_X(x)\,dx = \frac{y/10-(-1.2)}{2.5} \tag{3.1.60}$$

The next special case is $y = +10$. Because all values of Y fall at or below $+10$, the CDF must be 1, as stated earlier. Thus increasing y beyond $+10$ adds nothing. Note, however, that there is a jump of the CDF at $y = +10$. If we substitute $y = 10$ into Eq. (3.1.60), we find

$$F_Y(+10) = \frac{10/10+1.2}{2.5} = \frac{2.2}{2.5} \tag{3.1.61}$$

Thus the CDF jumps by $\frac{0.3}{2.5}$ at exactly $y = +10$, which indicates finite probability mass at this value of y. We may now plot the CDF of Y in Fig. 3.1.16. The finite jumps at $y = \pm 10$ indicate that the output is a mixed random variable, having both discrete and continuous components.

The PDF of Y. We must take the derivative of $F_Y(y)$, pictured in Fig. 3.1.16, to obtain the PDF of Y. The discontinuities at ± 10 indicate that no derivative can be defined at these points unless we resort to the use of impulse functions, which is exactly what we will do. If you need a short review on unit step and impulse functions, see endnote 12.

Using the unit step notation, we can express the CDF of Y as

$$F_Y(y) = \frac{y/10-(-1.2)}{2.5}(u(y-(-10)) - u(y-10)) + u(y-10) \tag{3.1.62}$$

The first unit step function "turns on" the straight line at $y = -10$, and the second unit step turns it off. The third unit step function makes the CDF 1 for all values greater than $+10$. The PDF is the derivative of Eq. (3.1.62). We will call the straight-line function $s(y)$ (the part in front of the unit step functions). With this notation, the derivative is

$$\begin{aligned} f_Y(y) &= s(-10)\delta(y-(-10)) - s(+10)\delta(y-10) + \delta(y-10) + \frac{1}{25}(u(y-(-10)) - u(y-10)) \\ &= s(-10)\delta(y-(-10)) + (1-s(+10))\delta(y-10) + \frac{1}{25}(u(y-(-10)) - u(y-10)) \end{aligned} \tag{3.1.63}$$

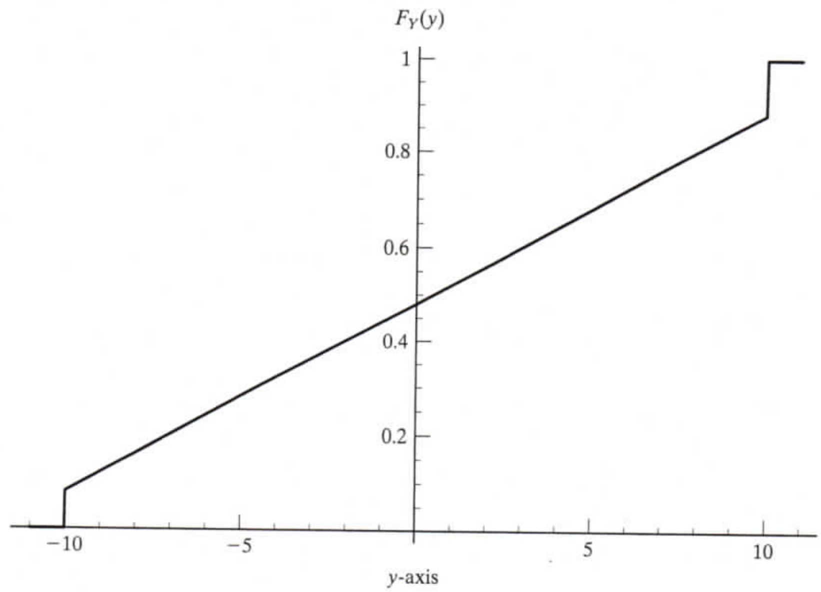

Figure 3.1.16 The CDF of the amplifier output, *Y*, shows jumps at ±10 corresponding to the saturation regions of the amplifier. The output will therefore be a mixed random variable, having discrete and continuous components.

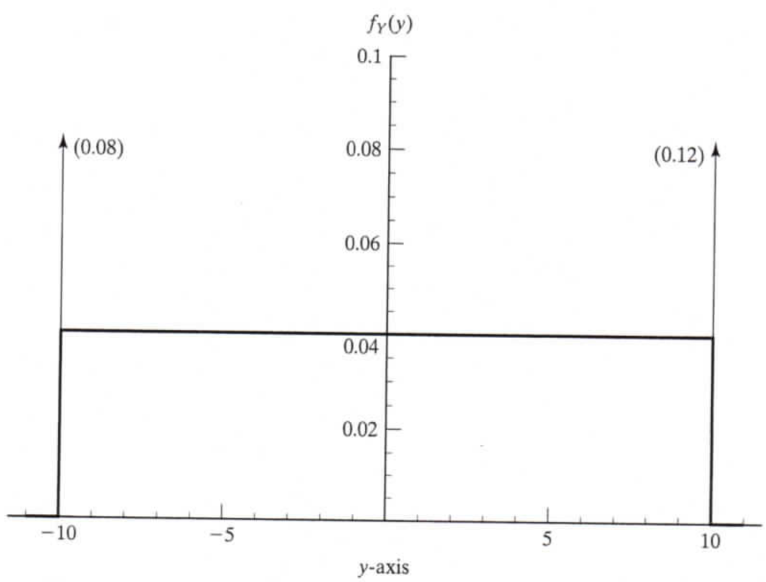

Figure 3.1.17 The PDF of *Y* has discrete and continuous aspects. The impulse functions indicate the presence of discrete components, and the continuous part, here a constant, indicates the continuous component.

Equation (3.1.63) loo
section. The first impu
discontinuity in Fig. 3.
$\frac{3}{25}$ at $y = +10$ and cc
Eq. (3.1.63) has a con
straight-line portion in
The PDF in Fig.
the CDF lead to the ir
features.

Summary. Sect
ables. The major point

- The distributio
probability den
and the CDF cc
- The CDF is a
PDF is a densit
- Conditional CD
calculation.

In Sec. 3.2 we extend

3.2 BIVARIATE RAN

3.2.1 Bivariate Prob

Mapping two ra
with two continuous ra
sample space into the (
chance experiment ther
Many physical pro
location of an imperfect
a population, or the vel
The definitions and mo
routinely extended to th

Definition of biv
with appropriate accom
also called the *joint PD*

$f_{XY}(x$

and $f_{XY}(x, y)$ is thus a
that are continuous in b
dx and *dy* be positive f
PDF must be positive.

Equation (3.1.63) looks complicated but can be interpreted as two impulses and a continuous section. The first impulse has a magnitude of $s(-10) = \frac{2}{25}$ at $y = -10$, which corresponds to the discontinuity in Fig. 3.1.16 at that point. The second term is an impulse of magnitude $1-s(+10) = \frac{3}{25}$ at $y = +10$ and corresponds to the discontinuity in Fig. 3.1.16 at that point. The last term in Eq. (3.1.63) has a constant height of $\frac{1}{25}$ over the range $-10 < y < +10$ and corresponds to the straight-line portion in Fig. 3.1.16. The corresponding PDF is shown in Fig. 3.1.17.

The PDF in Fig. 3.1.17 is the derivative of the CDF in Fig. 3.1.16. The discontinuities in the CDF lead to the impulse functions in the PDF. Mixed random variables always lead to these features.

Summary. Section 3.1 presented the basic tools for describing continuous random variables. The major points follow:

- The distribution of probabilities of a continuous random variable may be described by a probability density function, PDF, or a cumulative distribution function, CDF. The PDF and the CDF contain the same information.
- The CDF is a probability and is useful for derivations and for analyzing problems. The PDF is a density of probability and is useful for calculating probabilities and expectations.
- Conditional CDFs and PDFs can be defined and provide useful tools for analysis and calculation.

In Sec. 3.2 we extend these concepts and definitions to the bivariate case, two random variables.

3.2 BIVARIATE RANDOM VARIABLES

3.2.1 Bivariate Probability Density Functions, PDFs

Mapping two random variables into the Cartesian plane. In this section we deal with two continuous random variables, call them X and Y. The two random variables map the sample space into the Cartesian plane, as shown in Fig. 3.2.1(a), that is, for each outcome of the chance experiment there is an associated point X, Y in the x, y plane.

Many physical problems introduce two or more such random variables. For examples, the location of an imperfection on a semiconductor wafer, the height and weight of an individual from a population, or the velocity of a molecule in a gas (three components = three random variables). The definitions and methodologies we introduce for two random variables in this section are routinely extended to three or more random variables.

Definition of bivariate PDF. Definitions are the same as for the discrete bivariate PMF, with appropriate accommodation for continuous space. For example, the bivariate PDF, $f_{XY}(x, y)$, also called the *joint PDF,* is defined as

$$f_{XY}(x, y)\,dx\,dy = P[(x < X \leq x + dx) \cap (y < Y \leq y + dy)] \qquad (3.2.1)$$

and $f_{XY}(x, y)$ is thus a probability per unit area.[13] This definition applies to random variables that are continuous in both X and Y and is illustrated in Fig. 3.2.1(b). The definition requires that dx and dy be positive for the probability to be defined, from which it follows that the bivariate PDF must be positive.

of expectation to calculate the mean and variance of the Pascal random variable, with the results

$$E[N] = \frac{k}{p} \text{ and } Var[N] = \frac{kq}{p^2} \tag{3.3.27}$$

where $N =$ the number of independent trials to the kth success, and $p = P[\text{success}]$ and $q = 1-p$. In the present discussion of the central limit theorem, we would expect the Pascal random variable for large k to be approximately Gaussian. Let us test this by comparing the free-throw shooter's distribution, Fig. 2.1.4, with a Gaussian of the same mean and variance. Figure 3.3.10 compares the two distributions.

Figure 3.3.10 shows that convergence is slower with nonsymmetric distributions. If you look at the geometric distribution (p. 93), you will see that it is very skewed. Adding many of these gives a result that is still skewed, as shown in Fig. 3.3.10. The Gaussian, being symmetric, has trouble matching the unsymmetric results, as you can see.

The practice of replacing a computationally difficult discrete distribution, such as the binomial or the Pascal, with the Gaussian is a holdover from the time when calculations were done by hand. Now that there is no lack of computational power, techniques that substitute the Gaussian for some other distribution are of less importance than in the past. We present these examples to illustrate one aspect of the CLT.

3.4 THE GAUSSIAN (NORMAL) RANDOM VARIABLE

3.4.1 The Normalized Gaussian Random Variable

Background. We showed in Sec. 3.3 that the sum of n IID random variables has a PDF that is the n-fold convolution of the PDF of the individual random variables. We asserted and demonstrated that such multiple convolutions tend to approach a stable shape, regardless of the original PDF. This shape is "the" bell-shaped curve, the Gaussian distribution, also called the *normal* distribution. The central limit theorem gives a mathematical foundation for believing that the Gaussian distribution will emerge in situations where complexity of a certain type underlies the randomness. Specifically, where many causes add together to create an effect, that effect will follow a normal distribution. That's why it is called *normal,* because it fits so many situations. Mathematical proof aside, the normal distribution is justified by its success in modeling real random systems.

Although *Gaussian* and *normal* are synonymous in this context, we favor *"Gaussian"* to honor Carl Frederick Gauss, considered by many the most productive mathematician of all time. Earlier workers derived and worked with this distribution, but Gauss was the first to state it clearly and recognize its significance.

Looking ahead. This section investigates the mathematical properties of the normal distribution and begins the exploration of its use as a model in random systems. We will become familiar with the Gaussian CDF and PDF, both in normalized and in general form. We will show how independent Gaussian random variables add, and give applications in error analysis. We will continue the practice of using the Gaussian to approximate discrete distributions, since this approximation has some use in practical analysis.

The normalized Gaussian PDF

Normalization. Here we assume a Gaussian random variable with zero mean and unit variance. We will use the following notation as a shorthand: $Z = N(0, 1)$ means "The random variable Z is normal (Gaussian) with zero mean and unity variance." The general form for a Gaussian distribution is

$$f_Z(z) = \alpha e^{-\beta(z-\gamma)^2} \text{ for all } z \tag{3.4.1}$$

where $\alpha, \beta,$ and γ are constants.

We now normalize the distribution for unit area, zero mean, and unity variance.[30] We may achieve zero mean by setting $\gamma = 0$, since that gives the distribution even symmetry about the origin. If a PDF has even symmetry about any point, that point is clearly the balance point and thus the mean. Using standard integral tables, we find for unit area

$$\int_{-\infty}^{+\infty} f_Z(z)\,dz = \int_{-\infty}^{+\infty} \alpha e^{-\beta z^2}\,dz = \alpha\sqrt{\frac{\pi}{\beta}} = 1 \tag{3.4.2}$$

and for unit variance

$$\int_{-\infty}^{+\infty} (z-0)^2 f_Z(z)\,dz = \int_{-\infty}^{+\infty} \alpha z^2 e^{-\beta z^2}\,dz = \alpha\sqrt{\frac{\pi}{4\beta^3}} = 1 \tag{3.4.3}$$

Simultaneous solution of Eqs. (3.4.2) and (3.4.3) yields

$$\alpha = \frac{1}{\sqrt{2\pi}} \text{ and } \beta = \frac{1}{2} \tag{3.4.4}$$

Thus the normalized Gaussian PDF is

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \text{ for all } z \tag{3.4.5}$$

Equation (3.4.5) is another form of what we mean when we write $Z = N(0, 1)$. Figure 3.4.1 shows a plot of the normalized Gaussian PDF.

The square in the exponent causes this function to be fairly flat on top and drop rapidly as z increases. The tails of the distribution for large z in the plus or minus directions get very small. For example, for $z = \pm 7$, they are

```
PDF[NormalDistribution[0, 1], 7.0]
9.13472 × 10^-12
```

The cumulative distribution function (CDF) for the normalized Gaussian random variable

Definition. By definition, the CDF of the normal distribution is

$$F_Z(z) = P[Z \le z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy \tag{3.4.6}$$

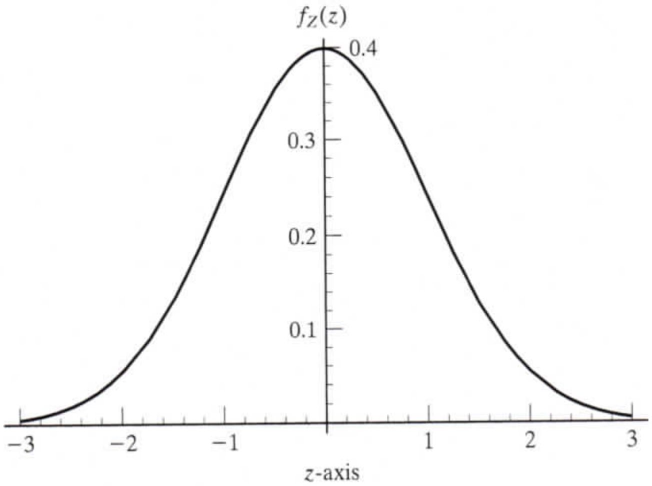

Figure 3.4.1 The PDF for the normalized Gaussian random variable, meaning zero mean and unit variance. This is "the" bell-shaped curve.

where we have used y for the dummy variable of integration, since z is the upper limit on the integral. There is no function whose derivative is the integrand, so this integral must be calculated numerically and tabulated. We have such a table,[31] but in the electronic book we mostly use Mathematica's library of functions. Figure 3.4.2 is a plot of the normalized Gaussian CDF.

Rather than look up the values of this function in a table, we call the CDF function. For example, for $z = +7$, the CDF has the value

```
CDF[NormalDistribution[0, 1], +7];
N[%, 20]
0.99999999999872018746
```

Some practical matters. Because we cannot count on using Mathematica at all times and places, we introduce the use of tables to determine the CDF of the Gaussian random variable. Because the Gaussian random variable is so important, a special notation is used for its CDF: $\Phi(z)$ is used to designate the CDF of $N(0, 1)$. Thus the following equation defines $\Phi(z)$:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \tag{3.4.7}$$

Normally, values are tabulated only for positive z. By symmetry, values for negative z are found as

$$\Phi(-z) = 1 - \Phi(z) \tag{3.4.8}$$

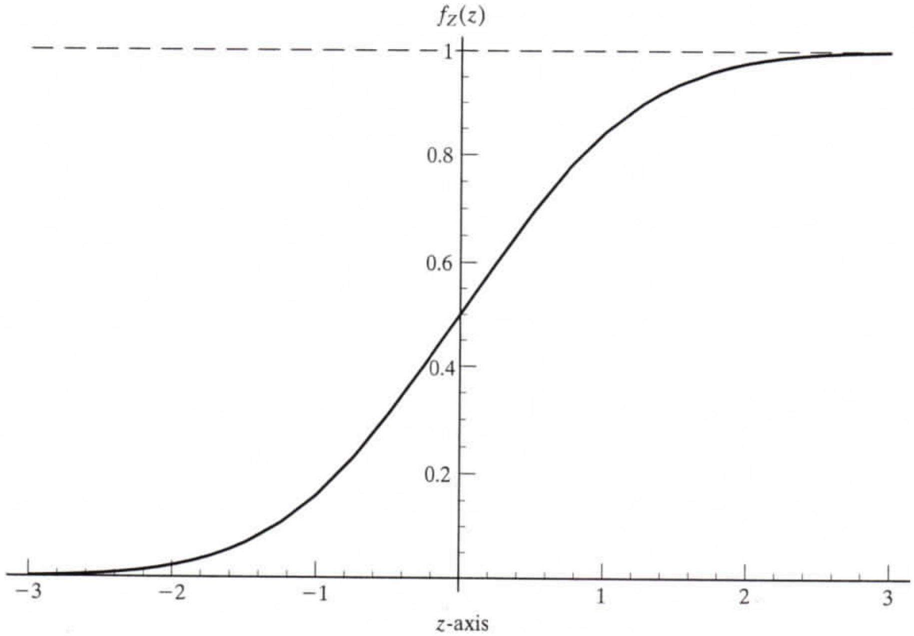

Figure 3.4.2 The CDF for the normalized Gaussian random variable.

The probability that Z is less than z in magnitude is

$$P[-z < Z < +z] = \Phi(z) - \Phi(-z) = 2\Phi(z) - 1 \tag{3.4.9}$$

Also, we note that the CDF approaches 1 as z gets large. To determine the tail probabilities, we need to use the $Q(z)$ function, which is defined as

$$Q(z) = \int_{z}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy \tag{3.4.10}$$

If you do not have a table for $Q(z)$ handy, you can use the asymptotic form:

$$Q(z) \to \frac{1}{z\sqrt{2\pi}} e^{-z^2/2} \text{ as } z \to +\infty \tag{3.4.11}$$

which works with less than 0.00002 error for $z > 3.5$. Also, you can use error function tables, since $Q(z) = \frac{1}{2}\text{erfc}\left(\frac{z}{\sqrt{2}}\right)$.

You do it. What is $\Phi(-1.5)$? Look it up in the table,[31] enter your answer in the cell box, and click Evaluate for a response.


```
myanswer = ;
```

```
Evaluate
```

For the answer, see en

3.4.2 The General (

Changing varia change of variables. W Z, by the following lin of expectation. The me

since $E[Z] = 0$. Simila

since $Var[Z] = 1$. No Eq. (2.3.43). Therefore a general Gaussian, $X =$

The PDF for X. must formally change r for X is

In Fig. 3.4.3, we show t by Eq. (3.4.15), in term

Because the event X of the former from the F

$$F_X(x) = P[X \le x]$$

For purposes of derivin actually is a very useful specific values for $\mu_X =$ reaches 50% at the mean

```
myanswer = ;
```

```
Evaluate
```

For the answer, see endnote 32.

3.4.2 The General (Unnormalized) Gaussian Random Variable

Changing variables. We may generalize to arbitrary mean and variance by making a change of variables. We define a random variable, X, in terms of the normalized random variable, Z, by the following linear transformation $X = aZ + b$. We can identify a and b from the algebra of expectation. The mean of X is

$$E[X] = E[aZ + b] = aE[Z] + b = \mu_X \Rightarrow b = \mu_X \tag{3.4.12}$$

since $E[Z] = 0$. Similarly, the variance of X is

$$Var[X] = Var[aZ + b] = a^2\, Var[Z] = \sigma_X^2 \Rightarrow a = \sigma_X \tag{3.4.13}$$

since $Var[Z] = 1$. Note that in Eq. (3.4.13) we drop the additive constant, b, according to Eq. (2.3.43). Therefore the required transformation from a normalized Gaussian, $Z = N(0, 1)$, to a general Gaussian, $X = N(\mu_X, \sigma_X^2)$, is

$$X = \sigma_X Z + \mu_X \tag{3.4.14}$$

The PDF for *X*. We now show that X has a Gaussian PDF. To find the PDF for X, we must formally change random variables. We begin with the CDF for X. By definition, the CDF for X is

$$F_X(x) = P[X \le x] \tag{3.4.15}$$

In Fig. 3.4.3, we show the plot of Z and X. We want to calculate a probability in X, as required by Eq. (3.4.15), in terms of an event in Z.

Because the event $X \le x$ corresponds to the event $Z \le \frac{x-\mu_X}{\sigma_X}$, we can calculate the probability of the former from the PDF of Z, as follows:

$$F_X(x) = P[X \le x] = P\left[Z \le \frac{x - \mu_X}{\sigma_X}\right] = \int_{-\infty}^{\frac{x-\mu_X}{\sigma_X}} f_Z(z)\, dz = \Phi\left(\frac{x - \mu_X}{\sigma_X}\right) \tag{3.4.16}$$

For purposes of deriving the PDF of X, Eq. (3.4.16) is merely an intermediate point, but it actually is a very useful result, and we will return to it later. For now, we give a plot, using specific values for $\mu_X = +10$, and $\sigma_X = 4$, in Fig. 3.4.4. This plot looks like $\Phi(z)$, except it reaches 50% at the mean of 10 and is expanded in width by a factor of 4.

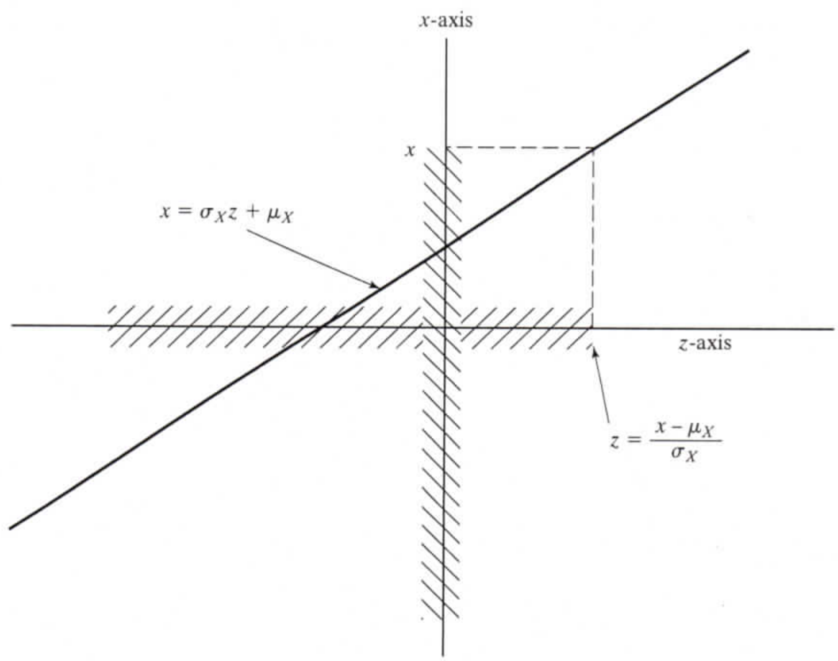

Figure 3.4.3 The mapping of Z into X. We have shaded in the event $X \leq x$ with the vertical crosshatched region and shown how it corresponds to the event $Z \leq \frac{x-\mu_X}{\sigma_X}$ in the horizontal crosshatched region. This event mapping allows us to calculate the CDF of X.

The PDF of X. The PDF of the generalized Gaussian random variable is the derivative of Eq. (3.4.16) with respect to x. To perform the derivative, we use Leibnitz's rule (see endnote 27):

$$\begin{aligned} f_X(x) &= \frac{d}{dx}\int_{-\infty}^{\frac{x-\mu_X}{\sigma_X}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \times \frac{d}{dx}\left(\frac{x-\mu_X}{\sigma_X}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \quad \text{for all } x \end{aligned} \tag{3.4.17}$$

The normalization term in front is sometimes expressed as $\frac{1}{\sigma_X\sqrt{2\pi}}$. This PDF is represented by the shorthand $X = N(\mu_X, \sigma_X^2)$ and is read, "X is a Gaussian (or normal) random variable with a mean of μ_X and a variance of σ_X^2." We can show easily that 68.3% of the probability falls within one standard deviation of the mean, and 95.4% falls within two standard deviations of the mean. We show this PDF for a mean of 10 and a variance of 16 ($\sigma_X = 4$) in Fig. 3.4.5.

The general Gaussian PDF in Eq. (3.4.17) is important and will be used frequently from now on, but when we calculate probabilities, we will have to return to the CDF in Eq. (3.4.16). For example, consider the random variable $X = N(10, 16)$, whose PDF is shown in Fig. 3.4.5. Let us say we wish to calculate the probability that X falls in the region between 5 and 20. Of course,

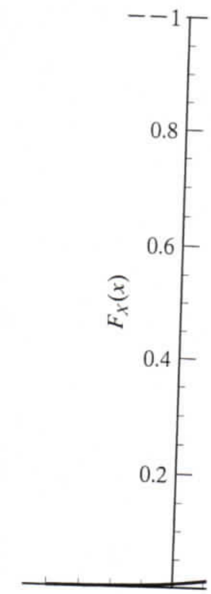

Figure 3.4.4 The CDF
Gaussian random varia
scale is expanded by a

Figure 3.4.5 The PDF fc
centered on the mean a
factor of 4.

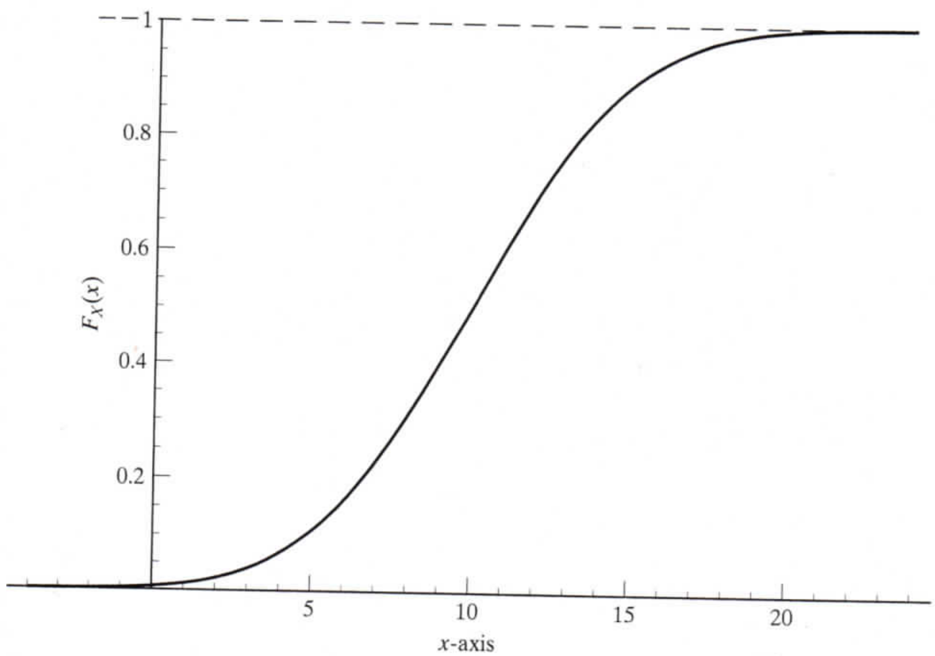

Figure 3.4.4 The CDF for $X = N(10,16)$. This CDF is the same as the CDF for the normalized Gaussian random variable in Fig. 3.4.2, except the 50% point is at $x = +10$, the mean, and the scale is expanded by a factor of 4, which is the ratio of the standard deviations.

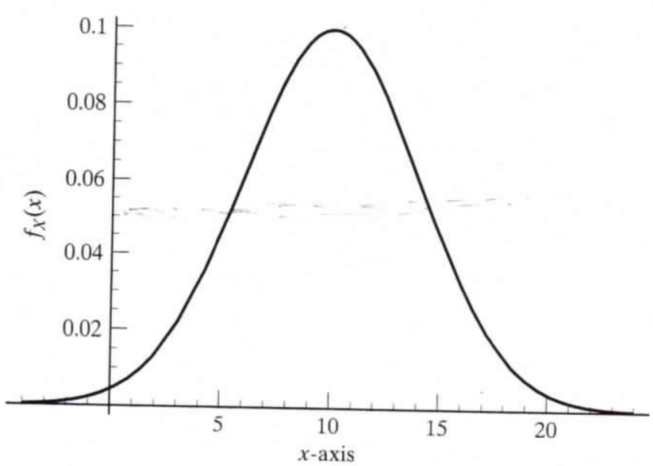

Figure 3.4.5 The PDF for $X = N(\mu_X, \sigma_X^2)$ with $\mu_X = +10$ and $\sigma_X = 4$. Note that the PDF is centered on the mean and is broader and shorter than the normalized PDF in Fig. 3.4.1 by a factor of 4.

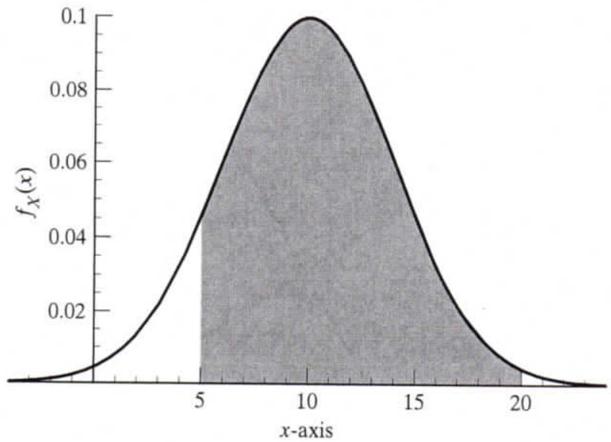

Figure 3.4.6 The probability $P[5 < X \leq 20]$ is the shaded area shown.

we can express this as an integral:

$$P[5 < X \leq 20] = \int_5^{20} f_X(x)\,dx \tag{3.4.18}$$

where $f_X(x)$ is given in Eq. (3.4.17); however, the antiderivative does not exist in closed form, so we resort directly to the tabulated CDF, $\Phi(z)$. Here is how we can determine this probability:

$$P[5 < X \leq 20] = F_X(20) - F_X(5) = \Phi\left(\frac{20-10}{4}\right) - \Phi\left(\frac{5-10}{4}\right) = \Phi(2.5) - \Phi(-1.25) \tag{3.4.19}$$

When we look up $\Phi(-1.25)$ in a table, we do not find it there and have to use the relationship in Eq. (3.4.8) to obtain

$$P[5 < X \leq 20] = \Phi(2.5) - \Phi(-1.25) = \Phi(2.5) + \Phi(+1.25) - 1 \tag{3.4.20}$$

Of course, having Mathematica handy, we can use it for the calculation in Eq. (3.4.19) directly:

```
CDF [NormalDistribution[10, 4], 20.0] -
 CDF [NormalDistribution[10, 4], 5.0]
0.888141
```

The value calculated is represented by the shaded area in Fig. 3.4.6.

Example 3.4.1: A bolt

A hole is exactly 0.500 in. in diameter. A bolt is manufactured with an OD of $0.498 \pm 0.003(1\sigma)$ in. Assume that the errors are normal, and let D be the random variable representing the bolt diameter. Thus $D = N(0.498, (0.003)^2)$. Find the probability that the bolt will not fit into the hole.

Solution The event of interest is $D > 0.500$ if we assume no force is applied to force the bolt into the hole. Thus we calculate the probability

$$P[D > 0.500] = 1 - P[D \leq 0.500] = 1 - \Phi\left(\frac{0.500 - 0.498}{0.003}\right) = 1 - \Phi(0.667) \tag{3.4.21}$$

Using Mathematica, we calculate

```
1 - CDF[NormalDistribution[0, 1], 0.667]
0.252386
```

You do it. Let us say there is also a problem if the bolt is too loose, but that we allow for forcing the bolt. Let us assume that a bolt is accepted if its diameter falls in the range $0.495 < D \leq 0.502$. Calculate the probability that a bolt selected at random will be accepted. You can use Mathematica if you are running the full system, or you can use the table in endnote 31. Enter your answer in the cell box, and click Evaluate for a response.

```
myanswer = ;
```

```
Evaluate
```

For the answer, see endnote 33.

Example 3.4.2: A resistor

Assume resistors as manufactured have a normal distribution. If 9% of a production run are not within ±10% of the nominal value and thus cannot be sold as ±10% resistors, what fraction is within ±5% of the nominal value?

Solution Assume R is a random variable describing the resistance value chosen from the population of resistors. The mean μ_R is the nominal value, but the standard deviation, σ_R is unknown. Our first task is to find σ_R, which we can determine from the information that 9% of the resistors are outside the ±10% boundary. Thus we must have

$$P[0.9\,\mu_R < R \leq 1.1\,\mu_R] = 0.91 \tag{3.4.22}$$

Using the normalized Gaussian CDF in Eq. (3.4.16) to evaluate this probability, we have

$$P[0.9\,\mu_R < R \leq 1.1\,\mu_R] = \Phi\left(\frac{0.1\,\mu_R}{\sigma_R}\right) - \Phi\left(\frac{-0.1\,\mu_R}{\sigma_R}\right) = 2\,\Phi\left(\frac{0.1\,\mu_R}{\sigma_R}\right) - 1 = 0.91 \tag{3.4.23}$$

where we used Eq. (3.4.8) for the negative argument. We solve Eq. (3.4.23) to get the requirement

$$\Phi\left(\frac{0.1\,\mu_R}{\sigma_R}\right) = 0.955 \tag{3.4.24}$$

which we can determine by interpolation in the Φ table in endnote 31. But it is even easier to use an inverse CDF table, $z = \Phi^{-1}(p)$, which we also give.[34] You can see from the table that $\Phi^{-1}(0.955) = 1.6950$. Thus

$$\frac{0.1\mu_R}{\sigma_R} = 1.695 \Rightarrow \frac{\mu_R}{\sigma_R} = 16.95 \tag{3.4.25}$$

Thus from the fact that 9% of the resistors do not meet the $\pm 10\%$ specification we derive the ratio of the nominal resistance to the standard deviation of the distribution. The question asks us to determine the probability that a resistor falls within $\pm 5\%$ of the nominal value. This question can be interpreted two ways: (1) What is the probability that a resistor chosen at random from the original batch falls within $\pm 5\%$ of the nominal? or (2) What is the probability that a resistor chosen at random from the resistors that meet the $\pm 10\%$ specification also falls within $\pm 5\%$ of the nominal value? The first is unconditioned, and the second is conditioned. We address the first interpretation first. We may write this as

$$P[0.95\,\mu_R < R \le 1.05\,\mu_R] = \Phi\left(\frac{0.05\,\mu_R}{\sigma_R}\right) - \Phi\left(\frac{-0.05\,\mu_R}{\sigma_R}\right) = 2\Phi\left(\frac{0.05\,\mu_R}{\sigma_R}\right) - 1 \tag{3.4.26}$$

Using the result derived in Eq. (3.4.25), we have

$$P[0.95\,\mu_R < R \le 1.05\,\mu_R] = 2\Phi(0.05 \times 16.95) - 1 = 2\Phi(0.8475) - 1 = 0.603 \tag{3.4.27}$$

Thus 60.3% of the resistors meet the 5% specification. If we take the second interpretation, we merely increase this value to $\frac{0.603}{0.91} = 0.663$.

You do it. Let us say the machine that makes resistors is misadjusted one day, and the mean for the resistors that are actually manufactured is $1.03\,\mu_R$; that is, on average, the resistances are 3% too high. What now is the yield of resistors in the $\pm 10\%$ range? Calculate the yield (before, it was 91%) to at least three places, and enter your answer in the cell box. Use percentage, but do not enter a % sign. Click Evaluate for a response.

```
myanswer = ;
```

```
Evaluate
```

For the answer, see endnote 35.

Example 3.4.3: Gas mileage

We will offer one more example because it involves slightly different mathematics. An auto manufacturer has certain requirements for the automobiles that it manufactures. The gasoline mileage, as the cars come off the assembly line and are tuned, is a Gaussian random variable, $M = N(25, (2)^2)$ mpg. The manufacturer then tests the mileage and remanufactures any unit that tests below 23 mpg. Determine the average mileage of all autos that meet the mileage specification.

Solution The problem calls for the conditional expectation $E[M|M \geq 23]$, which calls for the conditional PDF $f_{M|M\geq 23}(m)$. We calculate the probability of the conditioning event as

$$P[M \geq 23] = 1 - P[M < 23] = 1 - \Phi\left(\frac{23-25}{2}\right) = \Phi(1) = 0.8413 \tag{3.4.28}$$

The required conditional PDF is therefore

$$f_{M|M\geq 23}(m) = \frac{1}{0.8413} \times \frac{1}{2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m-25}{2}\right)^2\right], \quad 23 \leq m < +\infty \tag{3.4.29}$$

The expectation is

$$E[M|M \geq 23] = \int_{23}^{+\infty} m \times \frac{1}{0.8413} \times \frac{1}{2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m-25}{2}\right)^2\right] dm \tag{3.4.30}$$

We modify this integral by writing the first m as $(m-25)+25$ and thus split Eq. (3.4.30) into two integrals; call them I and II. The first is

$$\text{I:} \quad \int_{23}^{+\infty} (m-25) \frac{1}{0.8413} \times \frac{1}{2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m-25}{2}\right)^2\right] dm \tag{3.4.31}$$

We change variables to $w = \frac{m-25}{2}$, which yields

$$\text{I:} \int_{-1}^{+\infty} w \frac{1}{0.8413} \times \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}w^2\right] dw = \frac{1}{0.8413} \times \frac{2}{\sqrt{2\pi}} \int_{-1}^{+\infty} \exp\left[-\frac{1}{2}w^2\right] d\left(\frac{1}{2}w^2\right) \tag{3.4.32}$$

which gives the result

$$\sqrt{\frac{2}{\pi}} \frac{1}{0.8413} e^{-0.5} = 0.5752 \tag{3.4.33}$$

The second integral is easier because it reduces to Eq. (3.4.28):

$$\text{II: } 25 \times \int_{23}^{+\infty} \frac{1}{0.8413} \times \frac{1}{2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m-25}{2}\right)^2\right] dm = 25 \tag{3.4.34}$$

Thus we find $E[M|M \geq 23] = 25.5752$. We did the integral analytically to show you the technique. We can confirm it numerically with Mathematica.

```
NIntegrate[m PDF[NormalDistribution[25, 2], m],
  {m, 23, Infinity}] / CDF[NormalDistribution[0, 1], 1]
25.5752
```

Another way to integrate this is to have a self-normalizing integral. The critical part of the integral is the factor $\exp\left[-\frac{1}{2}\left(\frac{m-25}{2}\right)^2\right]$; the rest is a group of constants. Here, then, is another way to do the calculation.

```
NIntegrate[mExp[-0.5((m - 25)/2)^2], {m, 23, Infinity}]/
 NIntegrate[Exp[-0.5((m - 25)/2)^2], {m, 23, Infinity}]
25.5752
```

Note that we put the m in the numerator but not in the denominator. All the constants cancel, and we get the conditional expectation as required.

3.4.3 The Gaussian Used to Approximate Discrete Distributions

We showed in Sec. 3.3.5 that, as a consequence of the central limit theorem, we expect the Gaussian to approximate discrete distributions that represent the sum of a large number of random variables. We illustrated with some plots that compared the Gaussian with the binomial and the Pascal distributions.

Here we investigate a bit further and use the Gaussian to make calculations related to binomial trials. We introduce the calculation with a story.

Billy Bob Bojangles, having made a fortune in oil, decided to establish West Texas Airlines, headquartered at the Midland/Odessa Airport. Business was great and they sold all 140 seats on every flight. Some months after starting, Billy Bob's operations officer informed him that WTA flights were only 95% full because of last-minute cancellations and changes of plans, and so on. Billy Bob reached for the nearest envelope, turned it over, and made the following calculation: $\frac{140}{0.95} = 147.4$. Being conservative, Billy Bob rounded down and told the ticket people to sell 147 seats on every flight. Our job is to calculate the probability that a flight will be oversold, that is, that more than 140 people will show up to board the flight.

Our model. We model a flight as a series of 147 binomial trials. The probability of "success" is 0.95, and the random variable $K =$ the number of people that show up. The event $\{\text{oversold}\} = \{140 < K \leq 147\}$. Rather than do the calculation with the binomial distribution, however, we will use the Gaussian as an approximation.

Fitting a Gaussian to a binomial distribution. We make the two distributions have the same mean and variance. For the binomial, we derived the mean and variance [Eqs. (2.3.55, 2.3.56)] as

$$\mu_K = np = 147 \times 0.95 = 139.7 \text{ and } \sigma_K^2 = npq = 147 \times 0.95 \times 0.05 = 6.983 \qquad (3.4.35)$$

Thus our Gaussian model is $M = N(139.7, 6.983)$. We call this continuous random variable M because it is a continuous model for the discrete random variable K.

The "oversold flight" calculation. We will calculate the probability that a flight will be oversold, which corresponds to $K > 140$. This corresponds to the probability that $M > 140.5$. Why 140.5? Because M is a continuous random variable replacing a discrete random variable; so all the probability associated with $K = 141$ is modeled by $140.5 < M \leq 141.5$, and likewise for all the other discrete values. This is called a *continuity correction,* since we are approximating a discrete distribution with a continuous distribution. The required calculation is thus

$$P[\text{oversold}] = \int_{140.5}^{147.5} N(139.7, 6.983)\, dm = \Phi\left(\frac{147.5 - 139.7}{\sqrt{6.983}}\right) - \Phi\left(\frac{140.5 - 139.7}{\sqrt{6.983}}\right) \tag{3.4.36}$$

which is $0.9984 - 0.6133 = 0.3851$. Thus more than 38% of the WTA airlines flights are overbooked owing to Billy Bob's back-of-the-envelope calculation.

The exact calculations. Because we have the ready ability to make the exact calculation, let us see how accurate the approximate calculation is. The exact calculation is

```
Sum[PDF[BinomialDistribution[147, 0.95], k], {k, 141, 147}]
0.394075
```

In Eq. (3.4.36) we set the upper limit to 147.5, but infinity is more appropriate for calculating the probability above 140.5. Changing that limit to $+\infty$ raises the approximate calculation to 0.3867, which is closer to the true value but still 2% low.

You do it. You are now Billy Bob for a moment, so reach for the nearest envelope, turn it over, and calculate the number of tickets to be sold to have a 5% or less chance of being oversold. Enter your answer, an integer, in the cell box, and evaluate the cell to see how you did. Use the Gaussian approximation and let n be the unknown. Note that the mean and variance involve n.

```
myanswer = ;
```

```
Evaluate
```

For the answer, see endnote 36.

3.4.4 The Sum of Independent Gaussian Random Variables

The PDF of the sum of two independent normal random variables. We are concerned in this section with the sum of two or more independent Gaussian random variables. From the previous section we know that the PDF of the sum of two Gaussian random variables is