CHAPTER 2

Probability Distributions

THE BASICS OF PROBABILITY DISTRIBUTIONS

Imagine if you will that you are at a racetrack and you want to keep a log of the position in which the horses in a race finish. Specifically, you want to record whether the horse in the pole position came in first, second, and so on for each race of the day. You will record only 10 places. If the horse came in worse than in tenth place, you will record it as a tenth-place finish. If you do this for a number of days, you will have gathered enough data to see the *distribution* of finishing positions for a horse starting out in the pole position. Now you take your data and plot it on a graph. The horizontal axis represents where the horse finished, with the far left being the worst finishing position (tenth) and the far right being a win. The vertical axis will record how many times the pole-position horse finished in the position noted on the horizontal axis. You would begin to see a bell-shaped curve develop.

Under this scenario, there are 10 possible finishing positions for each race. We say that there are 10 *bins* in this distribution. What if, rather than using 10 bins, we used five? The first bin would be for a first- or second-place finish, the second bin for a third- or fourth-place finish, and so on. What would have been the result?

Using fewer bins on the same set of data would have resulted in a probability distribution with the same profile as one determined on the same

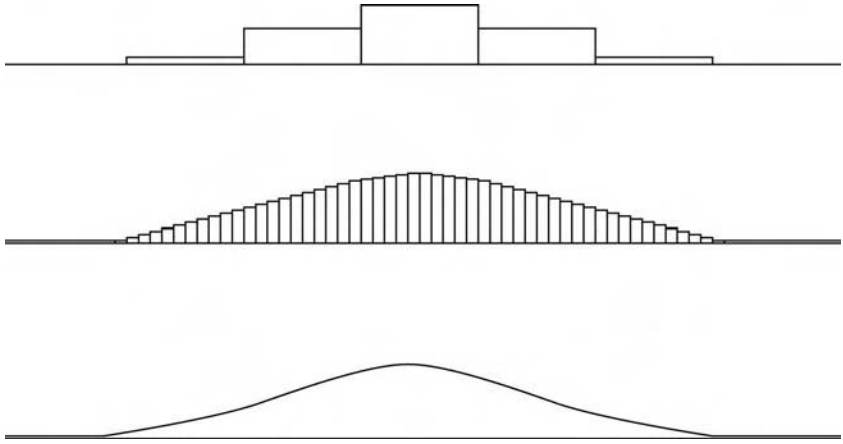

FIGURE 2.1 A continuous distribution is a series of infinitely thin bins

data with more bins. That is, they would look pretty much the same graphically. However, using fewer bins does reduce the information content of a distribution. Likewise, using more bins increases the information content of a distribution. If, rather than recording the finishing position of the pole-position horse in each race, we record the time the horse ran in, rounded to the nearest second, we will get more than 10 bins, and thus the information content of the distribution obtained will be greater.

If we recorded the exact finish time, rather than rounding finish times to use the nearest second, we would be creating what is called a *continuous* distribution. In a continuous distribution, there are no bins. Think of a continuous distribution as a series of infinitely thin bins (see Figure 2.1). A continuous distribution differs from a *discrete* distribution, the type we discussed first, in that a discrete distribution is a binned distribution. Although binning does reduce the information content of a distribution, in real life it is often necessary to bin data. Therefore, in real life it is often necessary to lose some of the information content of a distribution, while keeping the profile of the distribution the same, so that you can process the distribution. Finally, you should know that it is possible to take a continuous distribution and make it discrete by binning it, but it is not possible to take a discrete distribution and make it continuous.

When we are discussing the profits and losses of trades, we are essentially discussing a continuous distribution. A trade can take a multitude of values (although we could say that the data is binned to the nearest cent). In order to work with such a distribution, you may find it necessary to bin the data into, for example, $100-wide bins. Such a distribution would have

a bin for trades that made nothing to $99.99, the next bin would be for trades that made $100 to $199.99, and so on. There is a loss of information content in binning this way, yet the profile of the distribution of the trade profits and losses remains relatively unchanged.

DESCRIPTIVE MEASURES OF DISTRIBUTIONS

Most people are familiar with the average, or more specifically the *arithmetic mean*. This is simply the sum of the data points in a distribution divided by the number of data points:

$$A = \left(\sum_{i=1}^{N} X_i\right) \Big/ N \tag{2.01}$$

where: $A =$ The arithmetic mean.
$X_i =$ The ith data point.
$N =$ The total number of data points in the distribution.

The arithmetic mean is the most common of the types of measures of *location*, or *central tendency* of a body of data, a distribution. However, you should be aware that the arithmetic mean is not the only available measure of central tendency and often it is not the best. The arithmetic mean tends to be a poor measure when a distribution has very broad tails. Suppose you randomly select data points from a distribution and calculate their mean. If you continue to do this, you will find that the arithmetic means thus obtained converge poorly, if at all, when you are dealing with a distribution with very broad tails.

Another important measure of location of a distribution is the *median*. The median is described as the middle value when data are arranged in an array according to size. The median divides a probability distribution into two halves such that the area under the curve of one half is equal to the area under the curve of the other half. The median is frequently a better measure of central tendency than the arithmetic mean. Unlike the arithmetic mean, the median is *not* distorted by extreme outlier values. Further, the median can be calculated even for *open-ended* distributions. An open-ended distribution is a distribution in which all of the values in excess of a certain bin are thrown into one bin. An example of an open-ended distribution is the one we were compiling when we recorded the finishing position in horse racing for the horse starting out in the pole position. Any finishes

worse than tenth place were recorded as a tenth-place finish. Thus, we had an open distribution.

The third measure of central tendency is the *mode*—the most frequent occurrence. The mode is the peak of the distribution curve. In some distributions there is no mode and sometimes there is more than one mode. Like the median, the mode can often be regarded as a superior measure of central tendency. The mode is completely independent of extreme outlier values, and it is more readily obtained than the arithmetic mean or the median.

We have seen how the median divides the distribution into two equal areas. In the same way a distribution can be divided by three *quartiles* (to give four areas of equal size or probability), or nine *deciles* (to give 10 areas of equal size or probability) or 99 *percentiles* (to give 100 areas of equal size or probability). The 50th percentile is the median, and along with the 25th and 75th percentiles give us the quartiles. Finally, another term you should become familiar with is that of a *quantile.* A quantile is any of the $N-1$ variate-values that divide the total frequency into N equal parts.

We now return to the mean. We have discussed the arithmetic mean as a measure of central tendency of a distribution. You should be aware that there are other types of means as well. These other means are less common, but they do have significance in certain applications.

First is the *geometric mean*, which we saw how to calculate in the first chapter. The geometric mean is simply the Nth root of all the data points multiplied together.

$$G = \left(\prod_{i=1}^{N} X_i \right)^{1/N} \tag{2.02}$$

where: $G =$ The geometric mean.
$X_i =$ The ith data point.
$N =$ The total number of data points in the distribution.

The geometric mean cannot be used if any of the variate-values is zero or negative.

Another type of mean is the *harmonic mean.* This is the reciprocal of the mean of the reciprocals of the data points.

$$1/H = 1/N \sum_{i=1}^{N} 1/X_i \tag{2.03}$$

where: $H =$ The harmonic mean.
$X_i =$ The ith data point.
$N =$ The total number of data points in the distribution.

The final measure of central tendency is the *quadratic mean* or *root mean square.*

$$R^2 = 1/N \sum_{i=1}^{N} X_i^2 \tag{2.04}$$

where: R = The root mean square.
X_i = The ith data point.
N = The total number of data points in the distribution.

You should realize that the arithmetic mean (A) is always greater than or equal to the geometric mean (G), and the geometric mean is always greater than or equal to the harmonic mean (H):

$$H <= G <= A \tag{2.05}$$

where: H = The harmonic mean.
G = The geometric mean.
G = The arithmetic mean.

MOMENTS OF A DISTRIBUTION

The central value or location of a distribution is often the first thing you want to know about a group of data, and often the next thing you want to know is the data's variability or "width" around that central value. We call the measures of a distribution's central tendency the *first moment* of a distribution. The variability of the data points around this central tendency is called the *second moment* of a distribution. Hence, the second moment measures a distribution's dispersion about the first moment.

As with the measure of central tendency, many measures of dispersion are available. We cover seven of them here, starting with the least common measures and ending with the most common.

The *range* of a distribution is simply the difference between the largest and smallest values in a distribution. Likewise, the *10–90 percentile range* is the difference between the 90th and 10th percentile points. These first two measures of dispersion measure the spread from one extreme to the other. The remaining five measures of dispersion measure the departure from the central tendency (and hence measure the half-spread).

The *semi-interquartile range or quartile deviation* equals one half of the distance between the first and third quartiles (the 25th and 75th percentiles). This is similar to the 10–90 percentile range, except that with this measure the range is commonly divided by 2.

The *half-width* is an even more frequently used measure of dispersion. Here, we take the height of a distribution at its peak, the mode. If we find the point halfway up this vertical measure and run a horizontal line through it perpendicular to the vertical line, the horizontal line will touch the distribution at one point to the left and one point to the right. The distance between these two points is called the half-width.

Next, the *mean absolute deviation* or *mean deviation* is the arithmetic average of the absolute value of the difference between the data points and the arithmetic average of the data points. In other words, as its name implies, it is the average distance that a data point is from the mean. Expressed mathematically:

$$M = 1/N \sum_{i=1}^{N} ABS(X_i - A) \tag{2.06}$$

where: $M =$ The mean absolute deviation.
$N =$ The total number of data points.
$X_i =$ The ith data point.
$A =$ The arithmetic average of the data points.
$ABS(\,) =$ The absolute value function.

Equation (2.06) gives us what is known as the *population* mean absolute deviation. You should know that the mean absolute deviation can also be calculated as what is known as the *sample* mean absolute deviation. To calculate the sample mean absolute deviation, replace the term 1/N in Equation (2.06) with $1/(N - 1)$. You use the sample version when you are making judgments about the population based on a sample of that population.

The next two measures of dispersion, variance and standard deviation, are the two most commonly used. Both are used extensively, so we cannot say that one is more common than the other; suffice to say they are both the most common. Like the mean absolute deviation, they can be calculated two different ways, for a population as well as a sample. The population version is shown, and again it can readily be altered to the sample version by replacing the term 1/N with $1/(N - 1)$.

The *variance* is the same thing as the mean absolute deviation except that we square each difference between a data point and the average of the data points. As a result, we do not need to take the absolute value of each difference, since multiplying each difference by itself makes the result positive whether the difference was positive or negative. Further, since each distance is squared, extreme outliers will have a stronger effect on the

variance than they would on the mean absolute deviation. Mathematically expressed:

$$V = 1/N \sum_{i=1}^{N} (X_i - A)^2 \tag{2.07}$$

where: $V =$ The variance.
$N =$ The total number of data points.
$X_i =$ The ith data point.
$A =$ The arithmetic average of the data points.

Finally, the *standard deviation* is related to the variance (and hence the mean absolute deviation) in that the *standard deviation is simply the square root of the variance.*

The *third moment* of a distribution is called *skewness*, and it describes the extent of asymmetry about a distribution's mean (Figure 2.2). Whereas the first two moments of a distribution have values that can be considered *dimensional* (i.e., having the same units as the measured quantities), skewness is defined in such a way as to make it *nondimensional.* It is a pure number that represents nothing more than the shape of the distribution.

A positive value for skewness means that the tails are thicker on the positive side of the distribution and vice versa. A perfectly symmetrical distribution has a skewness of 0.

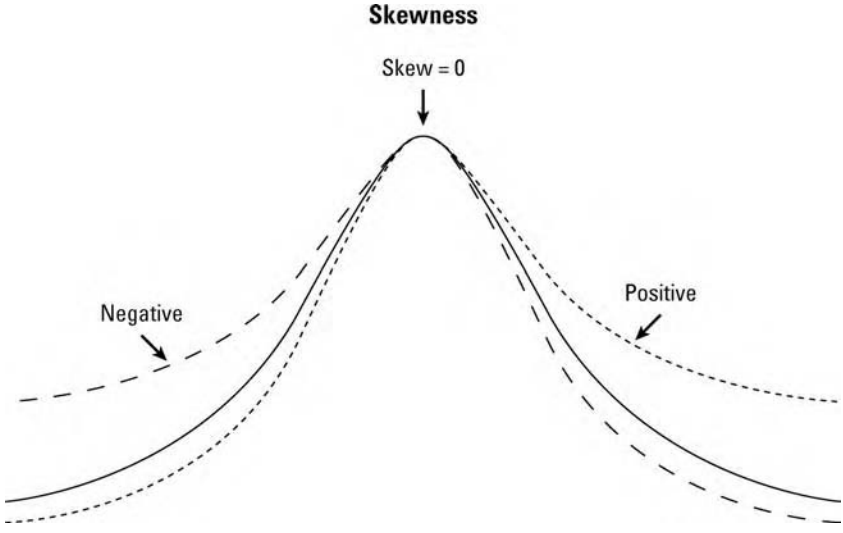

FIGURE 2.2 Skewness

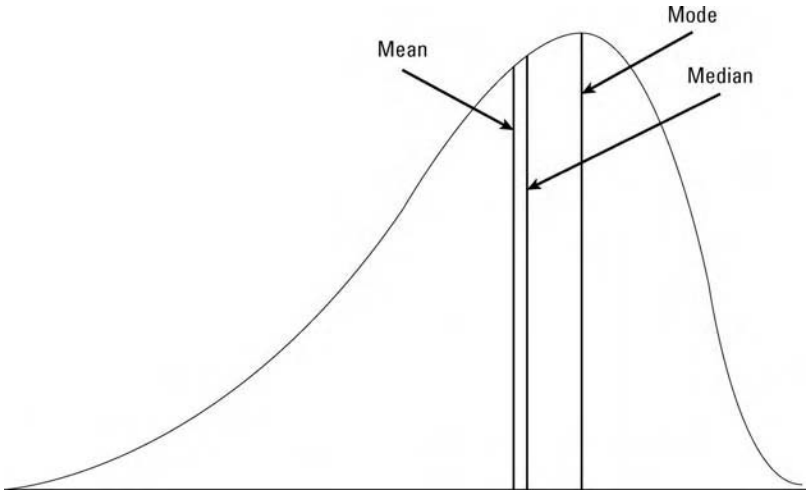

FIGURE 2.3 Skewness alters location

In a symmetrical distribution the mean, median, and mode are all at the same value. However, when a distribution has a nonzero value for skewness, this changes as depicted in Figure 2.3. The relationship for a skewed distribution (any distribution with a nonzero skewness) is:

$$\text{Mean} - \text{Mode} = 3 * (\text{Mean} - \text{Median}) \tag{2.08}$$

As with the first two moments of a distribution, there are numerous measures for skewness, which most frequently will give different answers. These measures now follow:

$$S = (\text{Mean} - \text{Mode})/\text{Standard Deviation} \tag{2.09}$$

$$S = (3 * (\text{Mean} - \text{Median}))/\text{Standard Deviation} \tag{2.10}$$

These last two equations, (2.09) and (2.10), are often referred to as Pearson's first and second coefficients of skewness, respectively. Skewness is also commonly determined as:

$$S = 1/N \sum_{i=1}^{N} ((X_i - A)/D)^3 \tag{2.11}$$

where: $S =$ The skewness.

$N =$ The total number of data points.

$X_i =$ The ith data point.

$A =$ The arithmetic average of the data points.

$D =$ The population standard deviation of the data points.

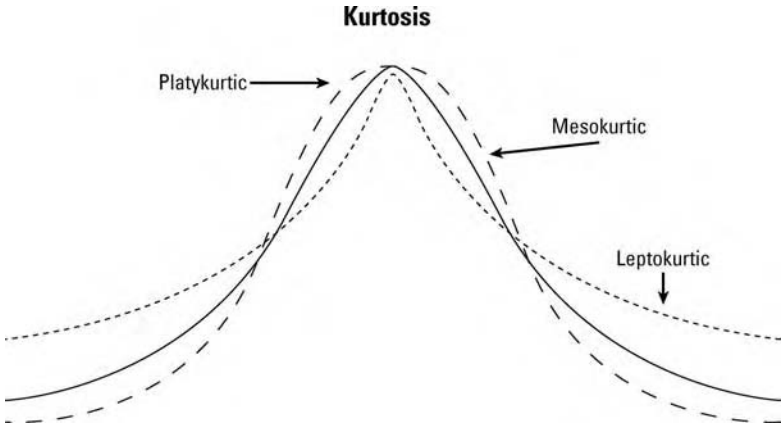

FIGURE 2.4 Kurtosis

Finally, the *fourth moment* of a distribution, *kurtosis* (see Figure 2.4), measures the peakedness or flatness of a distribution (relative to the Normal Distribution). Like skewness, it is a nondimensional quantity. A curve less peaked than the Normal is said to be *platykurtic* (kurtosis will be negative), and a curve more peaked than the Normal is called *leptokurtic* (kurtosis will be positive). When the peak of the curve resembles the Normal Distribution curve, kurtosis equals zero, and we call this type of peak on a distribution *mesokurtic*.

Like the preceding moments, kurtosis has more than one measure. The two most common are:

$$K = Q/P \tag{2.12}$$

where: $K =$ The kurtosis.
$Q =$ The semi-interquartile range.
$P =$ The 10–90 percentile range.

$$K = \left(1/N \sum_{i=1}^{N} ((X_i - A)/D)^4\right) - 3 \tag{2.13}$$

where: $K =$ The kurtosis.
$N =$ The total number of data points.
$X_i =$ The ith data point.
$A =$ The arithmetic average of the data points.
$D =$ The population standard deviation of the data points.

Finally, it should be pointed out there is a lot more "theory" behind the moments of a distribution than is covered here. The depth of discussion about the moments of a distribution presented here will be more than adequate for our purposes throughout this text.

Thus far, we have covered data distributions in a general sense. Now we will cover the specific distribution called the Normal Distribution.

THE NORMAL DISTRIBUTION

Frequently, the Normal Distribution is referred to as the Gaussian distribution, or de Moivre's distribution, after those who are believed to have discovered it—Karl Friedrich Gauss (1777–1855) and, about a century earlier and far more obscurely, Abraham de Moivre (1667–1754).

The Normal Distribution is considered to be the most useful distribution in modeling. This is due to the fact that the Normal Distribution accurately models many phenomena. Generally speaking, we can measure heights, weights, intelligence levels, and so on from a population, and these will very closely resemble the Normal Distribution.

Let's consider what is known as Galton's board (Figure 2.5). This is a vertically mounted board in the shape of an isosceles triangle. The board is studded with pegs, one on the top row, two on the second, and so on. Each row down has one more peg than the previous row. The pegs are arranged in a triangular fashion such that when a ball is dropped in, it has a 50/50 probability of going right or left with each peg it encounters. At the base of the board is a series of troughs to record the exit gate of each ball.

The balls falling through Galton's board and arriving in the troughs will begin to form a Normal Distribution. The "deeper" the board is (i.e., the more rows it has) and the more balls are dropped through, the more closely the final result will resemble the Normal Distribution.

The Normal is useful in its own right, but also because it tends to be the limiting form of many other types of distributions. For example, if X is distributed binomially, then as N tends toward infinity, X tends to be Normally distributed. Further, the Normal Distribution is also the limiting form of a number of other useful probability distributions such as the Poisson, the Student's, or the T distribution. In other words, as the data (N) used in these other distributions increases, these distributions increasingly resemble the Normal Distribution.

THE CENTRAL LIMIT THEOREM

One of the most important applications for statistical purposes involving the Normal Distribution has to do with the distribution of averages. The

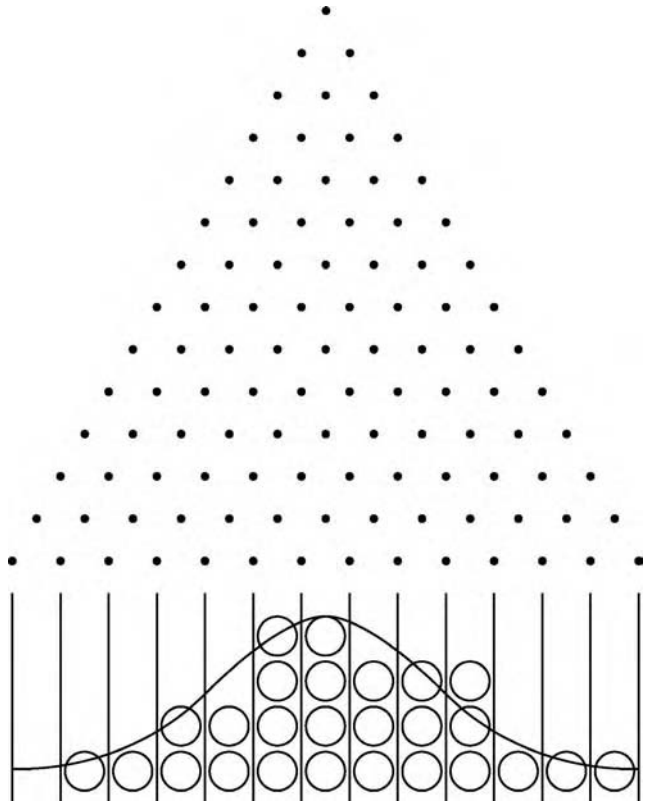

FIGURE 2.5 Galton's board

averages of samples of a given size, taken such that each sampled item is selected independently of the others, will yield a distribution that is close to Normal. This is an extremely powerful fact, for it means that you can generalize about an actual random process from averages computed using sample data.

Thus, we can state that *if N random samples are drawn from a population, then the sums (or averages) of the samples will be approximately Normally distributed, regardless of the distribution of the population from which the samples are drawn. The closeness to the Normal Distribution improves as N (the number of samples) increases.*

As an example, consider the distribution of numbers from 1 to 100. This is what is known as a *uniform distribution*: All elements (numbers in this case) occur only once. The number 82 occurs once and only once, as does 19, and so on. Suppose now that we take a sample of five elements and we take the average of these five sampled elements (we can just as

well take their sums). Now, we replace those five elements back into the population, and we take another sample and calculate the sample mean. If we keep on repeating this process, we will see that the sample means are Normally distributed, even though the population from which they are drawn is uniformly distributed.

Furthermore, this is true *regardless* of how the population is distributed! The Central Limit Theorem allows us to treat the distribution of sample means as being Normal without having to know the distribution of the population. This is an enormously convenient fact for many areas of study.

If the population itself happens to be Normally distributed, then the distribution of sample means will be exactly (not approximately) Normal. This is true because how quickly the distribution of the sample means approaches the Normal, as N increases, is a function of how close the population is to Normal. As a general rule of thumb, if a population has a *unimodal distribution*—any type of distribution where there is a concentration of frequency around a single mode, and diminishing frequencies on either side of the mode (i.e., it is convex)—or is uniformly distributed, using a value of 20 for N is considered sufficient, and a value of 10 for N is considered *probably* sufficient. However, if the population is distributed according to the Exponential Distribution (Figure 2.6), then it may be necessary to use an N of 100 or so.

The Central Limit Theorem, this amazingly simple and beautiful fact, validates the importance of the Normal Distribution.

WORKING WITH THE NORMAL DISTRIBUTION

In using the Normal Distribution, we most frequently want to find the percentage of area under the curve at a given point along the curve. In the parlance of calculus this would be called the integral of the function for the curve itself. Likewise, we could call the function for the curve itself the derivative of the function for the area under the curve. Derivatives are often noted with a prime after the variable for the function. Therefore, if we have a function, N(X), that represents the percentage of area under the curve at a given point, X, we can say that the derivative of this function, N′(X) (called N prime of X), is the function for the curve itself at point X.

We will begin with the formula for the curve itself, N′(X). This function is represented as:

$$\begin{aligned} \mathrm{N}'(\mathrm{X}) = {} & 1/(\mathrm{S} * \sqrt{2 * 3.1415926536} \\ & * \mathrm{EXP}(-(\mathrm{X} - \mathrm{U})^2/2 * \mathrm{S}^2) \end{aligned} \tag{2.14}$$

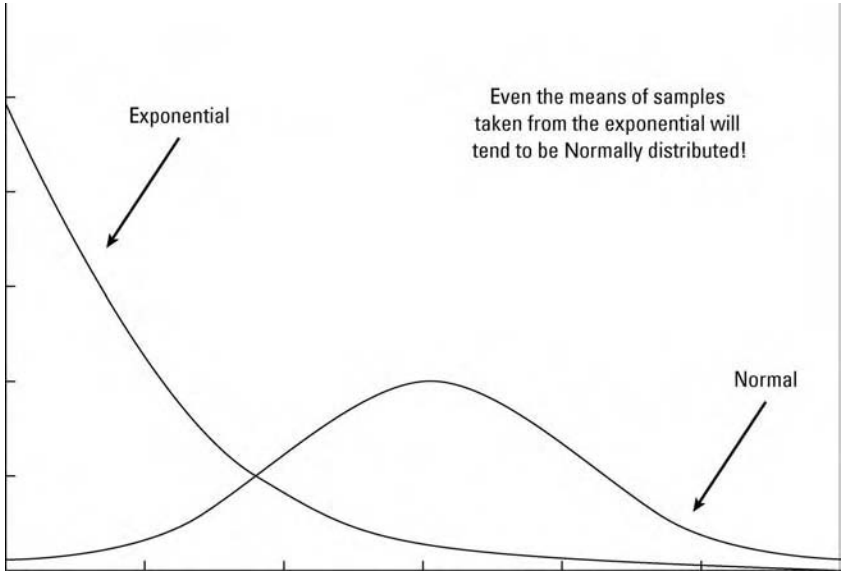

FIGURE 2.6 The Exponential Distribution and the Normal

where: U = The mean of the data.
S = The standard deviation of the data.
X = The observed data point.
EXP( ) = The exponential function.

This formula will give us the Y axis value, or the height of the curve if you will, at any given X axis value.

Often, it is easier to refer to a point along the curve with reference to its X coordinate in terms of how many standard deviations it is away from the mean. Thus, a data point that was 1 standard deviation away from the mean would be said to be one *standard unit* from the mean.

Further, it is often easier to subtract the mean from all of the data points, which has the effect of shifting the distribution so that it is centered over zero rather than over the mean. Therefore, a data point that was 1 standard deviation to the right of the mean would now have a value of 1 on the X axis.

When we make these conversions, subtracting the mean from the data points, then dividing the difference by the standard deviation of the data points, we are converting the distribution to what is called the *standardized normal*, which is the Normal Distribution with mean = 0 and

variance = 1. Now, N′(Z) will give us the Y axis value (the height of the curve) for any value of Z:

$$\begin{aligned} N'(Z) &= 1/\sqrt{2 * 3.1415926536 * EXP(-(Z^2/2))} \\ &= .398942 * EXP(-(Z^2/2)) \end{aligned} \tag{2.15a}$$

where: $Z = (X - U)/S$

and $U =$ The mean of the data.

$S =$ The standard deviation of the data.

$X =$ The observed data point.

$EXP\ (\ ) =$ The exponential function.

Equation (2.16) gives us the number of *standard units* that the data point corresponds to—in other words, how many standard deviations away from the mean the data point is. When Equation (2.16) equals 1, it is called the *standard normal deviate*. A standard deviation or a standard unit is sometimes referred to as a *sigma*. Thus, when someone speaks of an event's being a "five sigma event," they are referring to an event whose probability of occurrence is the probability of being beyond 5 standard deviations.

Consider Figure 2.7, which shows this equation for the Normal curve. Notice that the height of the standard Normal curve is .39894. From Equation (2.15a), the height is:

$$\begin{aligned} N'(Z) &= .398942 * EXP(-(Z^2/2)) \\ N'(0) &= .398942 * EXP(-(0^2/2)) \\ N'(0) &= .398942 \end{aligned}$$

Notice that the curve is *continuous*—that is, there are no "breaks" in the curve as it runs from minus infinity on the left to positive infinity on the right. Notice also that the curve is symmetrical, the side to the right of the peak being the mirror image of the side to the left of the peak.

Suppose we had a group of data where the mean of the data was 11 and the standard deviation of the group of data was 20. To see where a data point in that set would be located on the curve, we could first calculate it as a standard unit. Suppose the data point in question had a value of −9. To calculate how many standard units this is, we first must subtract the mean from this data point:

$$-9 - 11 = -20$$

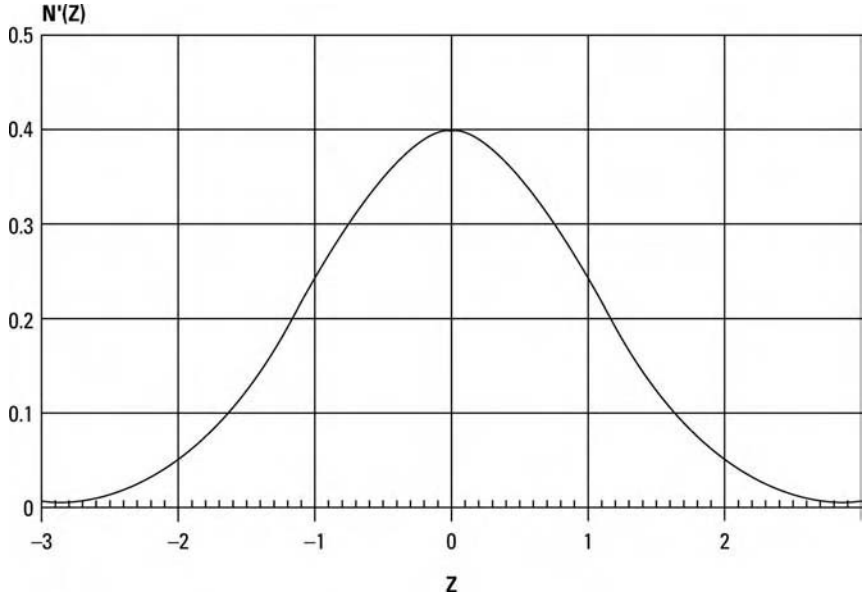

FIGURE 2.7 The Normal Probability density function

Next, we need to divide the result by the standard deviation:

$$-20/20 = -1$$

We can therefore say that the number of standard units is -1, when the data point equals -9, and the mean is 11, and the standard deviation is 20. In other words, we are 1 standard deviation away from the peak of the curve, the mean, and since this value is negative we know that it means we are 1 standard deviation to the left of the peak. To see where this places us on the curve itself (i.e., how high the curve is at 1 standard deviation left of center, or what the Y axis value of the curve is for a corresponding X axis value of -1), we need to now plug this into Equation (2.15a):

$$\begin{aligned} \mathrm{N}'(\mathrm{Z}) &= .398942 * \mathrm{EXP}(-(\mathrm{Z}^2/2)) \\ &= .398942 * 2.7182818285(-(-1^2/2)) \\ &= .398942 * 2.7182818285^{-1/2} \\ &= .398942 * .6065307 \\ &= .2419705705 \end{aligned}$$

Thus, we can say that the height of the curve at $X = -1$ is .2419705705. The function $N'(Z)$ is also often expressed as:

$$\begin{aligned} N'(Z) &= EXP(-(Z^2/2))/\sqrt{8 * ATN(1)} \\ &= EXP(-(Z^2/2))/\sqrt{8 * .7853983} \\ &= EXP(-(Z^2/2))/2.506629 \end{aligned} \tag{2.15b}$$

where: $$Z = (X - U)/S \tag{2.16}$$

and

$ATN( ) =$ The arctangent function.
$U =$ The mean of the data.
$S =$ The standard deviation of the data.
$X =$ The observed data point.
$EXP ( ) =$ The exponential function.

Nonstatisticians often find the concept of the standard deviation (or its square, *variance*) hard to envision. A remedy for this is to use what is known as the mean absolute deviation and convert it to and from the standard deviation in these equations. The mean absolute deviation is exactly what its name implies. The mean of the data is subtracted from each data point. The absolute values of each of these differences are then summed, and this sum is divided by the number of data points. What you end up with is the average distance each data point is away from the mean. The conversion for mean absolute deviation and standard deviation are given now:

$$\begin{aligned} \text{Mean Absolute Deviation} &= S * \sqrt{2/3.1415926536} \\ &= S * .7978845609 \end{aligned} \tag{2.17}$$

where: $M =$ The mean absolute deviation.
$S =$ The standard deviation.

Thus, we can say that in the Normal Distribution, the mean absolute deviation equals the standard deviation times .7979. Likewise:

$$\begin{aligned} S &= M * 1/.7978845609 \\ &= M * 1.253314137 \end{aligned} \tag{2.18}$$

where: $S =$ The standard deviation.
$M =$ The mean absolute deviation.

So we can also say that in the Normal Distribution the standard deviation equals the mean absolute deviation times 1.2533. Since the variance

is always the standard deviation squared (and standard deviation is always the square root of variance), we can make the conversion between variance and mean absolute deviation.

$$\begin{aligned} M &= \sqrt{V} * \sqrt{2/3.1415926536} \\ &= \sqrt{V} * .7978845609 \end{aligned} \tag{2.19}$$

where: $M =$ The mean absolute deviation.
$V =$ The variance.

$$V = (M * 1.253314137)^2 \tag{2.20}$$

where: $V =$ The variance.
$M =$ The mean absolute deviation.

Since the standard deviation in the standard normal curve equals 1, we can state that the mean absolute deviation in the standard normal curve equals .7979.

Further, in a bell-shaped curve like the Normal, the semi-interquartile range equals approximately two-thirds of the standard deviation, and therefore the standard deviation equals about 1.5 times the semi-interquartile range. This is true of most bell-shaped distributions, not just the Normal, as are the conversions given for the mean absolute deviation and standard deviation.

NORMAL PROBABILITIES

We now know how to convert our raw data to standard units and how to form the curve N′(Z) itself (i.e., how to find the height of the curve, or Y coordinate for a given standard unit) as well as N′(X) (Equation (2.14), the curve itself without first converting to standard units). To really use the Normal Probability Distribution, though, we want to know what the probabilities of a certain outcome's happening are. This is *not* given by the height of the curve. Rather, the probabilities correspond to the area under the curve. These areas are given by the integral of this N′(Z) function that we have thus far studied. We will now concern ourselves with N(Z), the integral to N′(Z), to find the areas under the curve (the probabilities).[1]

[1]The actual integral to the Normal probability density does not exist in closed form, but it can very closely be approximated by Equation (2.21).

$$\begin{aligned} N(Z) = 1 - N'(Z) * ((1.330274429 * Y\,\hat{}\,5) - (1.821255978 * Y\,\hat{}\,4) \\ + (1.781477937 * Y\,\hat{}\,3) - (.356563782 * Y\,\hat{}\,2) \\ + (.31938153 * Y)) \end{aligned} \tag{2.21}$$

If $Z < 0$, then $N(Z) = 1 - N(Z)$. Now recall Equation (2.15a):

$$N'(Z) = .398942 * EXP(-(Z^2/2))$$

where: $Y = 1/(1 + 2316419 * ABS(Z))$

and $ABS(\,) =$ The absolute value function.

$EXP(\,) =$ The exponential function.

We will always convert our data to standard units when finding probabilities under the curve. That is, we will not describe an N(X) function, but rather we will use the N(Z) function where:

$Z = (X - U)/S$

and $U =$ The mean of the data.

$S =$ The standard deviation of the data.

$X =$ The observed data point.

Refer now to Equation (2.21). Suppose we want to know what the probability is of an event's not exceeding +2 standard units ($Z = +2$).

$$\begin{aligned} Y &= 1/(1 + 2316419 * ABS(+2)) \\ &= 1/1.4632838 \\ &= .68339443311 \end{aligned}$$

$$\begin{aligned} N'(Z) &= .398942 * EXP(-(Z^2/2)) \\ &= .398942 * EXP(-2) \\ &= .398942 * .1353353 \\ &= .05399093525 \end{aligned}$$

Notice that this tells us the height of the curve at −2 standard units. Plugging these values for Y and N′(Z) into Equation (2.21) we can obtain the probability of an event's not exceeding +2 standard units:

$$\begin{aligned} N(Z) = 1 - N'(Z) * ((1.330274429 * Y\,\hat{}\,5) - (1.821255978 * Y\,\hat{}\,4) \\ +(1.781477937 * Y\,\hat{}\,3) - (.356563782 * Y\,\hat{}\,2) \\ +(.31938153 * Y)) \end{aligned}$$

$$= 1 - .05399093525 * \begin{pmatrix} 1.330274429 * .68339443311^5 \\ -1.821255978 * .68339443311^4 \\ +1.781477937 * .68339443311^3 \\ -.356563782 * .68339443311^2 \\ +.31928153 * .68339443311 \end{pmatrix}$$

$$= 1 - .05399093525 * \begin{pmatrix} 1.330274429 * .1490587 - 1.821255978 * \\ .2181151 + 1.781477937 * .3191643 \\ -.356563782 * .467028 + .31928153 * \\ .68339443311 \end{pmatrix}$$

$$= 1 - .05399093525 * (.198299977 - .3972434298$$
$$+.5685841587 - .16652527 + .2182635596)$$
$$= 1 - .05399093525 * .4213679955$$
$$= 1 - .02275005216$$
$$= .9772499478$$

Thus, we can say that we can expect 97.72% of the outcomes in a Normally distributed random process to fall shy of +2 standard units. This is depicted in Figure 2.8.

If we wanted to know what the probabilities were for an event's equaling or exceeding a prescribed number of standard units (in this case +2), we would simply amend Equation (2.21), taking out the 1− in the beginning of the equation and doing away with the −Z provision (i.e., doing away with

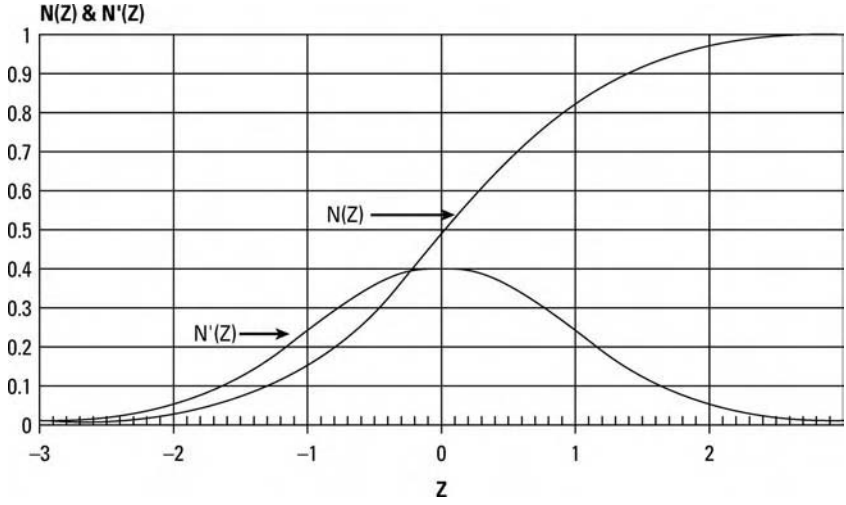

FIGURE 2.8 Equation (2.21) showing probability with Z = +2

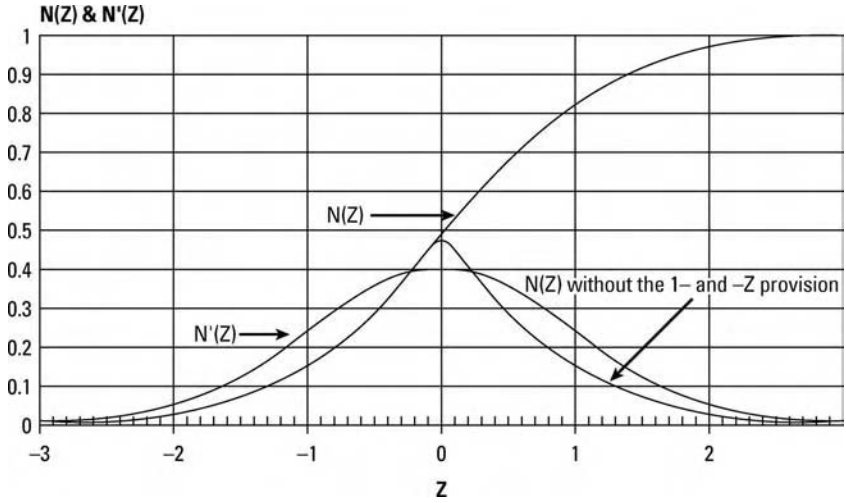

FIGURE 2.9 Doing away with the 1− and −Z provision in Equation (2.21)

"If Z < 0 then N(Z) = 1 – N(Z)"). Therefore, the second to last line in the last computation would be changed from

$$= 1 - .02275005216$$

to simply

$$.02275005216$$

We would therefore say that there is about a 2.275% chance that an event in a Normally distributed random process would equal or exceed +2 standard units. This is shown in Figure 2.9.

Thus far we have looked at areas under the curve (probabilities) where we are dealing only with what are known as "one-tailed" probabilities. That is to say we have thus far looked to solve such questions as, "What are the probabilities of an event's being less (more) than such-and-such standard units from the mean?" Suppose now we were to pose the question as, "What are the probabilities of an event's being *within* so many standard units of the mean?" In other words, we wish to find out what the "2-tailed" probabilities are.

Consider Figure 2.10. This represents the probabilities of being within 2 standard units of the mean. Unlike Figure 2.8, this probability computation does not include the extreme left tail area, the area of less than −2 standard units. To calculate the probability of being within Z standard units of the mean, you must first calculate the one-tailed probability of the absolute value of Z with Equation (2.21). This will be your input to the next

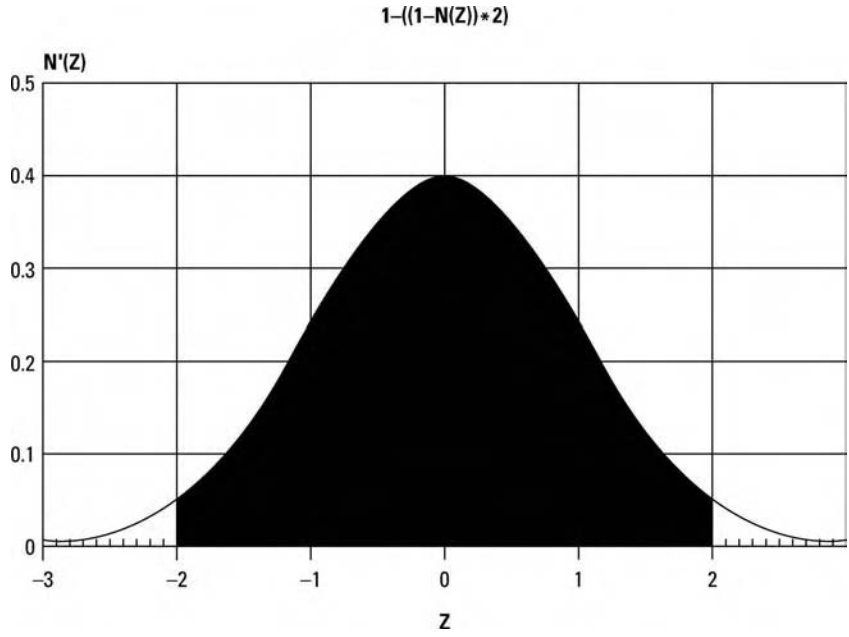

FIGURE 2.10 A two-tailed probability of an event's being + or −2 sigma

Equation, (2.22), which gives us the two-tailed probabilities (i.e., the probabilities of being within ABS(Z) standard units of the mean):

$$\text{Two-tailed probability} = 1 - ((1 - \text{N}(\text{ABS}(\text{Z}))) * 2) \qquad (2.22)$$

If we are considering what our probabilities of occurrence within 2 standard deviations are (Z = 2), then from Equation (2.21) we know that N(2) = .9772499478, and using this as input to Equation (2.22):

$$\begin{aligned}\text{Two-tailed probability} &= 1 - ((1 - .9772499478) * 2)\\ &= 1 - (.02275005216 * 2)\\ &= 1 - .04550010432\\ &= .9544998957\end{aligned}$$

Thus, we can state from this equation that the probability of an event in a Normally distributed random process falling within 2 standard units of the mean is about 95.45%.

Just as with Equation (2.21), we can eliminate the leading 1− in Equation (2.22) to obtain (1 – N(ABS(Z))) * 2, which represents the probabilities of an event's falling outside of ABS(Z) standard units of the mean. This is depicted in Figure 2.11. For the example where Z = 2, we can state that the

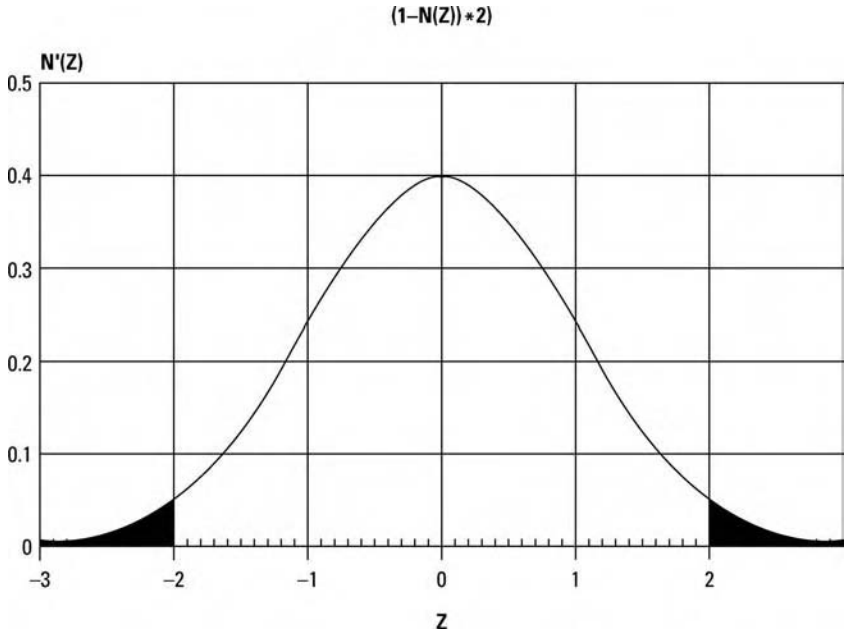

FIGURE 2.11 Two-tailed probability of an event's being beyond 2 sigma

probabilities of an event in a Normally distributed random process falling *outside* of 2 standard units is:

$$\begin{aligned}\text{Two-tailed probability (outside)} &= (1 - .9772499478) * 2 \\ &= .02275005216 * 2 \\ &= .04550010432\end{aligned}$$

Finally, we come to the case where we want to find what the probabilities (areas under the N′(Z) curve) are for two different values of Z.

Suppose we want to find the area under the N′(Z) curve between −1 standard unit and +2 standard units. There are a couple of ways to accomplish this. To begin with, we can compute the probability of not exceeding +2 standard units with Equation (2.21), and from this we can subtract the probability of not exceeding −1 standard units (see Figure 2.12). This would give us:

$$.9772499478 - .1586552595 = .8185946883$$

Another way we could have performed this is to take the number 1, representing the entire area under the curve, and then subtract the sum of

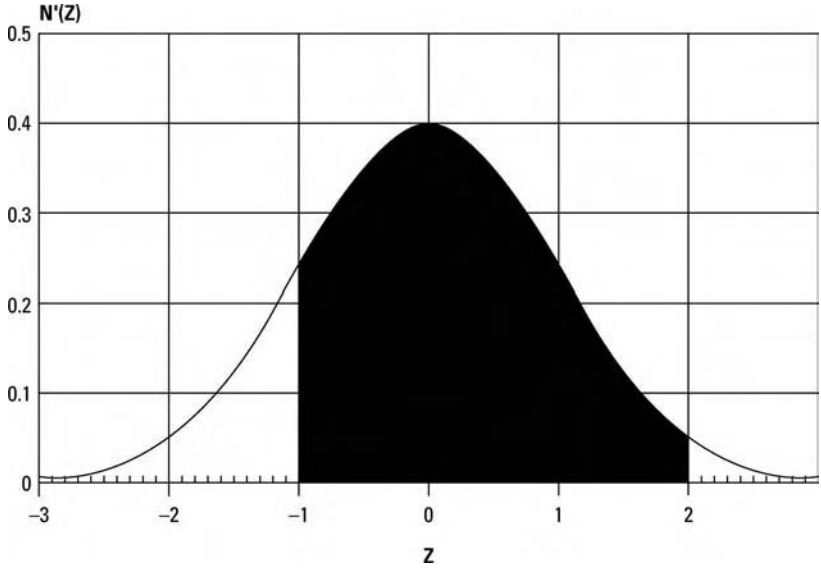

FIGURE 2.12 The area between −1 and +2 standard units

the probability of not exceeding −1 standard unit and the probability of exceeding 2 standard units:

$$= 1 - (.022750052 + .1586552595)$$
$$= 1 - .1814053117$$
$$= .8185946883$$

With the basic mathematical tools regarding the Normal Distribution thus far covered in this chapter, you can now use your powers of reasoning to figure any probabilities of occurrence for Normally distributed random variables.

FURTHER DERIVATIVES OF THE NORMAL

Sometimes you may want to know the second derivative of the N(Z) function. Since the N(Z) function gives us the area under the curve at Z, and the N′(Z) function gives us the height of the curve itself at Z, then the N″(Z) function gives us the *instantaneous slope* of the curve at a given Z:

$$N''(Z) = -Z/2.506628274 * EXP(-(Z^2)/2) \qquad (2.23)$$

where: EXP( ) = The exponential function.

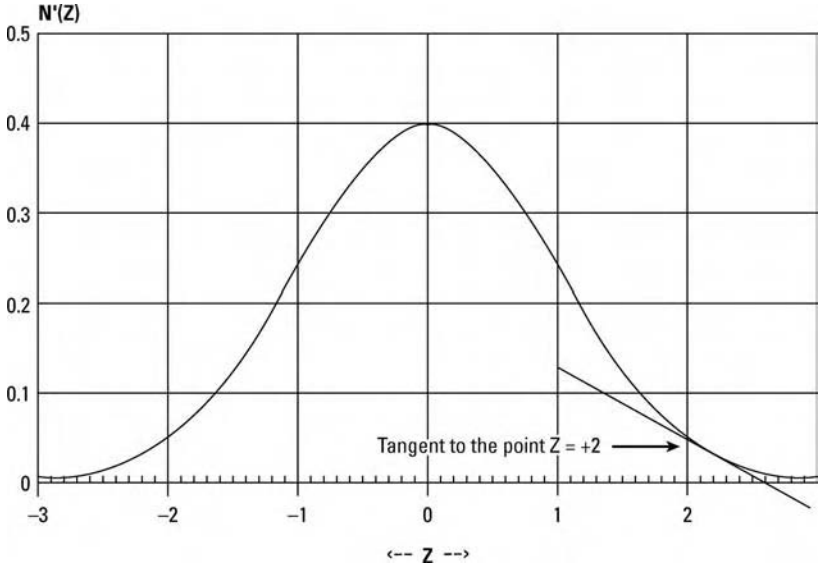

FIGURE 2.13 N″(Z) giving the slope of the line tangent tangent to N′(Z) at Z = +2

To determine what the slope of the N′(Z) curve is at +2 standard units:

$$\begin{aligned} N''(Z) &= -2/2.506628274 * EXP(-(2^2)/2) \\ &= -2/2.506628274 * EXP(-2) \\ &= -2/2.506628274 * .1353353 \\ &= -.1079968336 \end{aligned}$$

Therefore, we can state that the instantaneous rate of change in the N′(Z) function when Z = +2 is −.1079968336. This represents rise/run, so we can say that when Z = +2, the N′(Z) curve is rising −.1079968336 for every 1 unit run in Z. This is depicted in Figure 2.13.

For the reader's own reference, further derivatives are now given. These will not be needed throughout the remainder of this text, but are provided for the sake of completeness:

$$N'''(Z) = (Z^2 - 1)/2.506628274 * EXP(-(Z^2)/2)) \quad (2.24)$$

$$N''''(Z) = ((3 * Z) - Z^3)/2.506628274 * EXP(-(Z^2)/2)) \quad (2.25)$$

$$N'''''(Z) = (Z^4 - (6 * Z^2) + 3)/2.506628274 * EXP(-(Z^2)/2)) \quad (2.26)$$

As a final note regarding the Normal Distribution, you should be aware that the distribution is nowhere near as "peaked" as the graphic examples

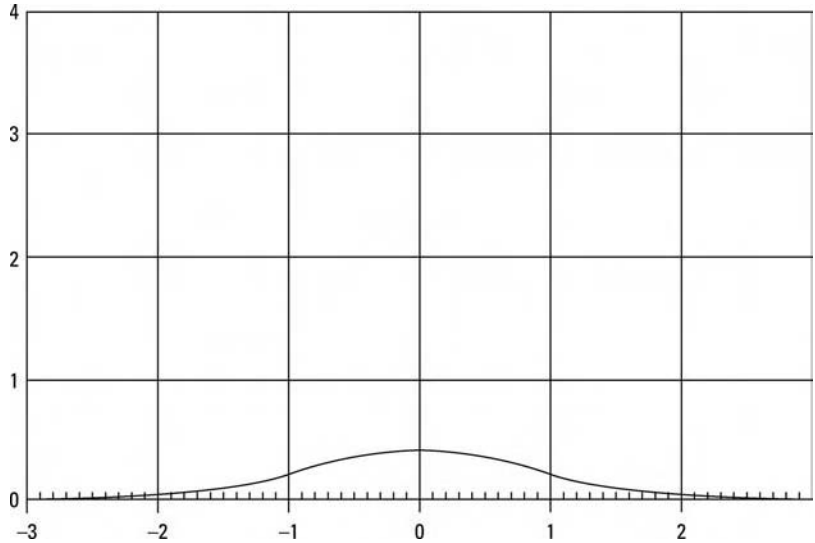

FIGURE 2.14 The real shape of the Normal Distribution

presented in this chapter imply. The real shape of the Normal Distribution is depicted in Figure 2.14.

Notice that here the scales of the two axes are the same, whereas in the other graphic examples they differ so as to exaggerate the shape of the distribution.

THE LOGNORMAL DISTRIBUTION

Many of the real-world applications in trading require a small but crucial modification to the Normal Distribution. This modification takes the Normal, and changes it to what is known as the Lognormal Distribution.

Consider that the price of any freely traded item has zero as a lower limit.[2] Therefore, as the price of an item drops and approaches zero, it

[2]This idea that the lowest an item can trade for is zero is not always entirely true. For instance, during the stock market crash of 1929 and the ensuing bear market, the shareholders of many failed banks were held liable to the depositors in those banks. Persons who owned stock in such banks not only lost their full investment, they also realized liability *beyond* the amount of their investment. The point here isn't to say that such an event can or cannot happen again. Rather, we cannot always say that zero is the absolute low end of what a freely traded item can be priced at, although it usually is.

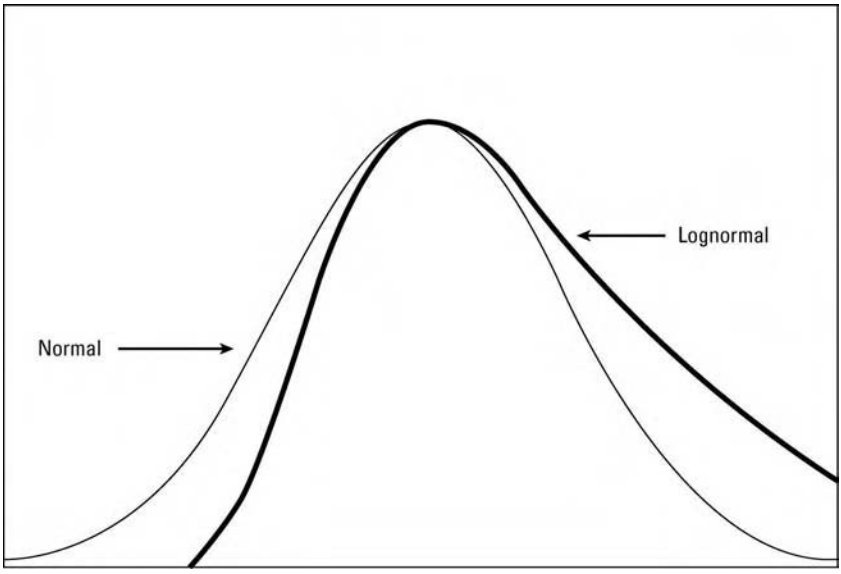

FIGURE 2.15 The Normal and Lognormal Distributions

should in theory become progressively more difficult for the item to get lower. For example, consider the price of a hypothetical stock at $10 per share. If the stock were to drop $5, to $5 per share, a 50% loss, then according to the Normal Distribution it could just as easily drop from $5 to $0. However, under the Lognormal, a similar drop of 50% from a price of $5 per share to $2.50 per share would be about as probable as a drop from $10 to $5 per share.

The Lognormal Distribution, Figure 2.15, works exactly like the Normal Distribution except that with the Lognormal we are dealing with percentage changes rather than absolute changes.

Consider now the upside. According to the Lognormal, a move from $10 per share to $20 per share is about as likely as a move from $5 to $10 per share, as both moves represent a 100% gain.

That isn't to say that we won't be using the Normal Distribution. The purpose here is to introduce you to the Lognormal, show you its relationship to the Normal (the Lognormal uses percentage price changes rather than absolute price changes), and point out that it usually is used when talking about price moves, or anytime that the Normal would apply but be bounded on the low end at zero.

To use the Lognormal Distribution, you simply convert the data you are working with to natural logarithms.[3] Now the converted data will be Normally distributed if the raw data was Lognormally distributed.

For instance, if we are discussing the distribution of price changes as being Lognormal, we can use the Normal distribution on it. First, we must divide each closing price by the previous closing price. Suppose in this instance we are looking at the distribution of monthly closing prices (we could use any time period—hourly, daily, yearly, or whatever). Suppose we now see \$10, \$5, \$10, \$10, then \$20 per share as our first five months closing prices. This would then equate to a loss of 50% going into the second month, a gain of 100% going into the third month, a gain of 0% going into the fourth month, and another gain of 100% into the fifth month. Respectively, then, we have quotients of .5, 2, 1, and 2 for the monthly price changes of months 2 through 5. We must now convert to natural logarithms in order to study their distribution under the math for the Normal Distribution. Thus, the natural log of .5 is −.6931473, of 2 it is .6931471, and of 1 it is 0. We are now able to apply the mathematics pertaining to the Normal Distribution to this converted data.[4]

THE UNIFORM DISTRIBUTION

The *Uniform Distribution*, sometimes referred to as the *Rectangular Distribution* from its shape, occurs when all items in a population have equal frequency. A good example is the 10 digits 0 through 9. If we were to randomly select one of these digits, each possible selection has an equal chance of occurrence. Thus, the Uniform Distribution is used to model truly random events. A particular type of Uniform Distribution where A = 0 and B = 1 is called the *Standard Uniform Distribution*, and it is used extensively in generating random numbers.

---

[3]The distinction between common and natural logarithms is reiterated here. A common log is a log base 10, while a natural log is a log base e, where e = 2.7182818285. The common log of X is referred to mathematically as log(X) while the natural log is referred to as ln(X). The distinction gets blurred when we observe BASIC programming code, which often utilizes a function LOG(X) to return the *natural* log.

[4]This is diametrically opposed to mathematical convention. BASIC does not have a provision for common logs, but the natural log can be converted to the common log by multiplying the natural log by .4342917. Likewise, we can convert common logs to natural logs by multiplying the common log by 2.3026.

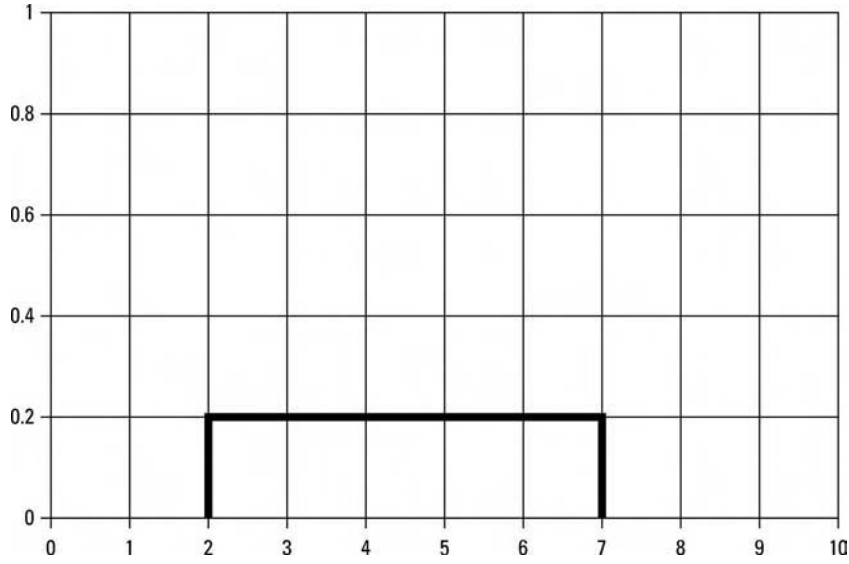

FIGURE 2.16 Probability density functions for the Uniform Distribution (A = 2, B = 7)

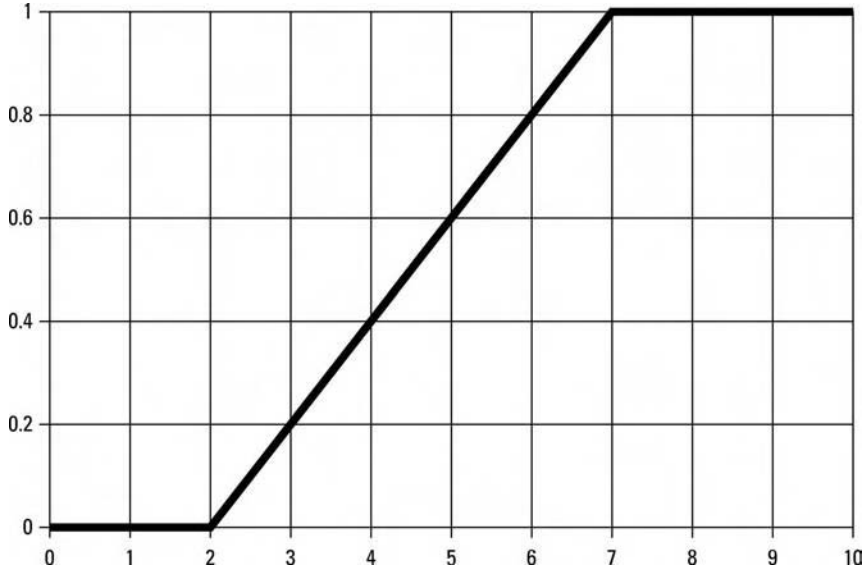

FIGURE 2.17 Cumulative probability functions for the Uniform Distribution (A = 2, B = 7)

The Uniform Distribution is a *continuous* distribution. The probability density function, N′(X), is described as:

$$N'(X) = 1/(B - A) \quad \text{for } A <= X <= B \tag{2.27}$$

else

$$N'(X) = 0$$

where: B = The rightmost limit of the interval AB.
A = The leftmost limit of the interval AB.

The cumulative density of the Uniform is given by:

$$N(X) = 0 \qquad \text{for } X < A \tag{2.28}$$

else

$$N(X) = (X - A)/(B - A) \qquad \text{for } A <= X <= B$$

else

$$N(X) = 1 \qquad \text{for } X > B$$

where: B = The rightmost limit of the interval AB.
A = The leftmost limit of the interval AB.

Figures 2.16 and 2.17 illustrate the probability density and cumulative probability (i.e., cdf) respectively of the Uniform Distribution.

Other qualities of the Uniform Distribution are:

$$\text{Mean} = (A + B)/2 \tag{2.29}$$

$$\text{Variance} = (B - A)^2/12 \tag{2.30}$$

where: B = The rightmost limit of the interval AB.
A = The leftmost limit of the interval AB.

THE BERNOULLI DISTRIBUTION

Another simple, common distribution is the *Bernoulli Distribution.* This is the distribution when the random variable can have only two possible values. Examples of this are heads and tails, defective and nondefective

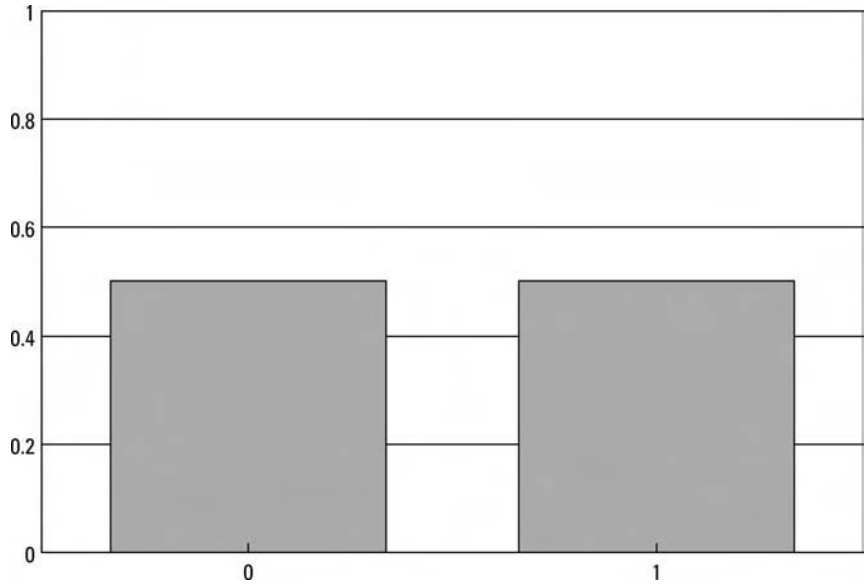

FIGURE 2.18 Probability density functions for the Binomial Distribution (P = .5)

articles, success or failure, hit or miss, and so on. Hence, we say that the Bernoulli Distribution is a *discrete distribution* (as opposed to being a continuous distribution). The distribution is completely described by one parameter, P, which is the probability of the first event occurring. The variance in the Bernoulli is:

$$\text{Variance} = P * Q \tag{2.31}$$

where:

$$Q = P - 1 \tag{2.32}$$

Figure 2.18 and 2.19 illustrate the probability density and cumulative probability (i.e., cdf) respectively of the Bernoulli Distribution.

THE BINOMIAL DISTRIBUTION

The *Binomial Distribution* arises naturally when sampling from a Bernoulli Distribution. The probability density function, N′(X), of the Binomial (the probability of X successes in N trials or X defects in N items or

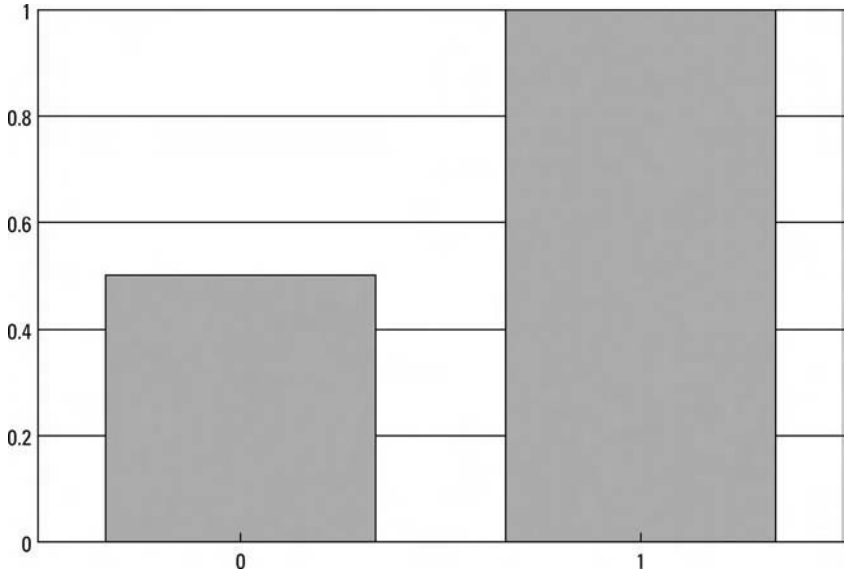

FIGURE 2.19 Cumulative probability function of the Bernoulli Distribution (P = .5)

X heads in N coin tosses, etc.) is:

$$N'(X) = (N!/(X! * (N - X)!)) * P^X * Q^{(N-X)} \tag{2.33}$$

where: N = The number of trials.
X = The number of successes.
P = The probability of a success on a single trial.
Q = 1 − P.

$$X! = X * (X - 1) * (X - 2) * \ldots * 1 \tag{2.34}$$

which can be also written as:

$$X! = \prod_{J=0}^{X-1} X - J \tag{2.34a}$$

Further, by convention:

$$0! = 1 \tag{2.34b}$$

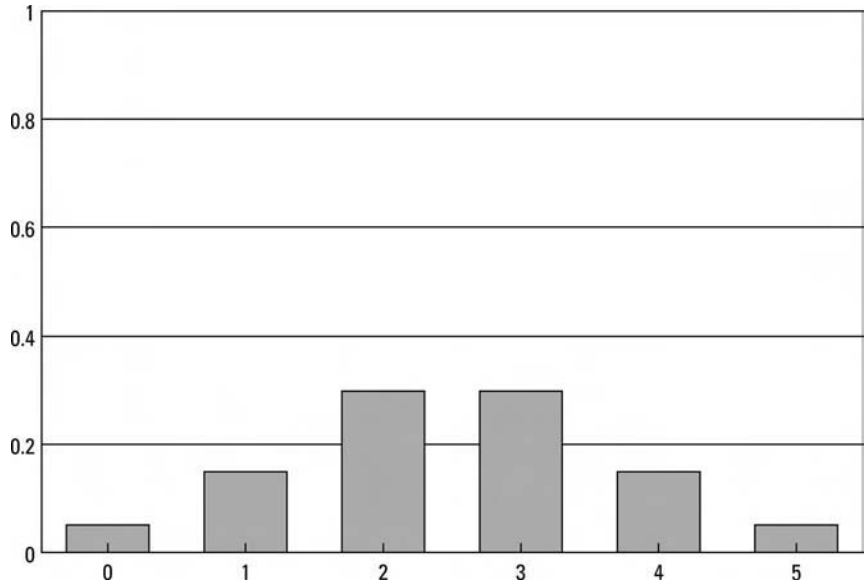

FIGURE 2.20 Probability density functions for the Binomial Distribution (N = 5, P = .5)

The cumulative density function for the Binomial is:

$$N(X) = \sum_{J=0}^{X} (N!/(J! * (N - J)!)) * P^{J} * Q^{(N-J)} \tag{2.35}$$

where: N = The number of trials.
X = The number of successes.
P = The probability of a success on a single trial.
Q = 1 − P.

Figures 2.20 and 2.21 illustrate the probability density and cumulative probability (i.e., cdf), respectively, of the Bionomial Distribution.

The Binomial is also a discrete distribution. Other properties of the Binomial Distribution are:

$$\text{Mean} = N * P \tag{2.36}$$

$$\text{Variance} = N * P * Q \tag{2.37}$$

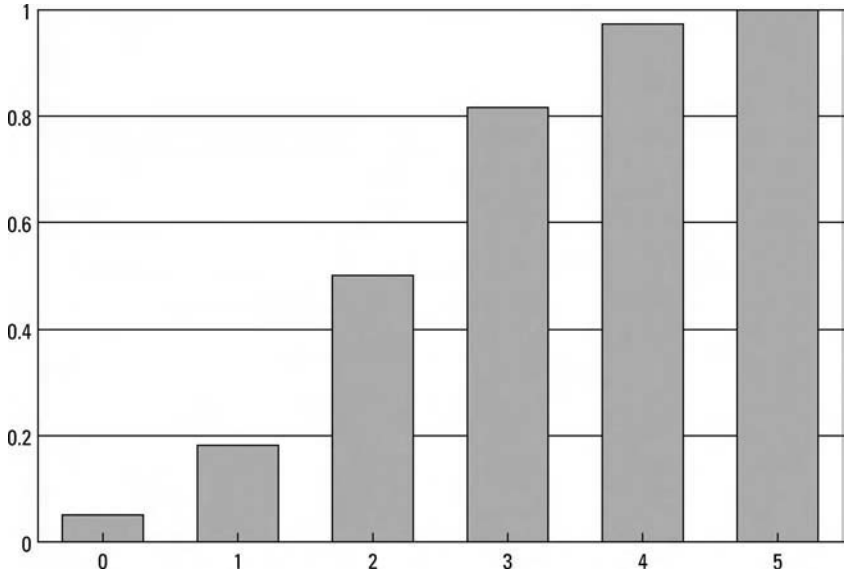

FIGURE 2.21 Cumulative probability functions for the Binomial Distribution (N = 5, P = .5)

where: N = The number of trials.
P = The probability of a success on a single trial.
Q = 1 − P.

As N becomes large, the Binomial tends to the Normal Distribution, with the Normal being the limiting form of the Binomial. Generally, if N * P and N * Q are both greater than 5, you could use the Normal in lieu of the Binomial as an approximation.

The Binomial Distribution is often used to statistically validate a gambling system. An example will illustrate. Suppose we have a gambling system that has won 51% of the time. We want to determine what the winning percentage would be if it performs in the future at a level of 3 standard deviations worse. Thus, the variable of interest here, X, is equal to .51, the probability of a winning trade. The variable of interest need not always be for the probability of a win. It can be the probability of an event being in one of two mutually exclusive groups. We can now perform the first necessary equation in the test:

$$L = P - Z * \sqrt{(P * (P - 1))/(N - 1)} \tag{2.38}$$

where: L = The lower boundary for P to be at Z standard deviations.
P = The variable of interest representing the probability of being in one of two mutually exclusive groups.
Z = The selected number of standard deviations.
N = The total number of events in the sample.

Suppose our sample consisted of 100 plays. Thus:

$$\begin{aligned} L &= .51 - 3 * \sqrt{(.51 * (1 - .51))/(100 - 1)} \\ &= .51 - 3 * \sqrt{(.51 * .49)/99} \\ &= .51 - 3 * \sqrt{.2499/99} \\ &= .51 - 3 * \sqrt{.0025242424} \\ &= .51 - 3 * .05024183938 \\ &= .51 - .1507255181 \\ &= .3592744819 \end{aligned}$$

Based on our history of 100 plays which generated a 51% win rate, we can state that it would take a three-sigma event for the population of plays (the future if we play an infinite number of times into the future) to have less than 35.92744819% winners.

What kind of a confidence level does this represent? That is a function of N, the total number of plays in the sample. We can determine the confidence level of achieving 35 or 36 wins in 100 tosses by Equation (2.35). However, (2.35) is clumsy to work with as N gets large because of all of the factorial functions in (2.35). Fortunately, the Normal distribution, Equation (2.21) for one-tailed probabilities, can be used as a very close approximation for the Binomial probabilities. In the case of our example, using Equation (2.21), 3 standard deviations translates into a 99.865% confidence. Thus, if we were to play this gambling system over an infinite number of times, we could be 99.865% sure that the percentage of wins would be greater than or equal to 35.92744819%.

This technique can also be used for statistical validation of trading systems. However, this method is only valid when the following assumptions are true. First, the N events (trades) are all independent and randomly selected. This can easily be verified for any trading system. Second, the N events (trades) can all be classified into two mutually exclusive groups (wins and losses, trades greater than or less than the median trade, etc.). This assumption, too, can easily be satisfied. The third assumption is that the probability of an event being classified into one of the two mutually exclusive groups is constant from one event to the next. This is not necessarily true in trading, and the technique becomes inaccurate to the degree

that this assumption is false. Be that as it may, the technique still can have value for traders.

Not only can it be used to determine the confidence level for a certain method's being profitable; the technique can also be used to determine the confidence level for a given market indicator. For instance, if you have an indicator that will forecast the direction of the next day's close, you then have two mutually exclusive groups: correct forecasts and incorrect forecasts. You can now express the reliability of your indicator to a certain confidence level.

This technique can also be used to discern how many trials are necessary for a system to be profitable to a given confidence level. For example, suppose we have a gambling system that wins 51% of the time on a game that pays 1 to 1. We want to know how many trials we must observe to be certain to a given confidence level that the system will be profitable in an asymptotic sense. Thus, we can restate the problem as, "If the system wins 51% of the time, how many trials must I witness, and have it show a 51% win rate, to know that it will be profitable to a given confidence level?"

Since the payoff is 1:1, the system must win in excess of 50% of the time to be considered profitable. Let's say we want the given confidence level to again be 99.865, or 3 standard deviations (although we are using 3 standard deviations in this discussion, we aren't restricted to that amount; we can use any number of standard deviations that we want). How many trials must we now witness to be 99.865% confident that at least 51% of the trials will be winners?

If $.51 - X = .5$, then $X = .01$. Therefore, the right factors of Equation (2.38), $Z * \sqrt{P * (P - 1)/(N - 1)}$, must equal .01. Since $Z = 3$ in this case, and $.01/3 = .0033$, then:

$$\sqrt{P * (1 - P)/(N - 1)}$$

We know that P equals .51, thus:

$$.51 * \sqrt{(1 - .51)/(N - 1)}$$

Squaring both sides gives us:

$$((.51 * (1 - .51))/(N - 1)) = .00001111$$

To continue:

$$(.51 * .49)/(N - 1) = .00001111$$

$$.2499/(N - 1) = .00001111$$

$$.2499/.00001111 = N - 1$$
$$.2499/.00001111 + 1 = N$$
$$22,491 + 1 = N$$
$$N = 22,492$$

Thus, we need to witness a 51% win rate over 22,492 trials to be 99.865% certain that we will see at least 51% wins.

THE GEOMETRIC DISTRIBUTION

Like the Binomial, the *Geometric Distribution*, also a discrete distribution, occurs as a result of N independent Bernoulli trials. The Geometric Distribution measures the number of trials before the first success (or failure). The probability density function, $N'(X)$, is:

$$N'(X) = Q^{(X-1)} * P \tag{2.39}$$

where: P = The probability of success for a given trial.
Q = The probability of failure for a given trial.

In other words, $N'(X)$ here measures the number of trials until the first success. The cumulative density function for the Geometric is therefore:

$$N(X) = \sum_{J=1}^{X} Q^{(J-1)} * P \tag{2.40}$$

where: P = The probability of success for a given trial.
Q = The probability of failure for a given trial.

Figures 2.22 and 2.23 illustrate the probability density and cumulative probability ability (i.e., cdf), respectively, of the Geometric Distribution.

Other properties of the Geometric are:

$$\text{Mean} = 1/P \tag{2.41}$$
$$\text{Variance} = Q/P^2 \tag{2.42}$$

where: P = The probability of success for a given trial.
Q = The probability of failure for a given trial.

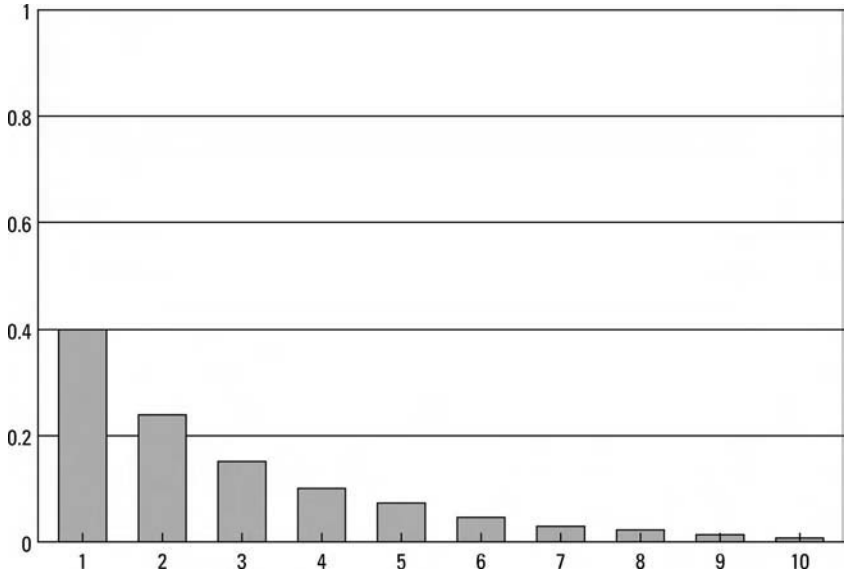

FIGURE 2.22 Probability density functions for the Geometric Distribution (P = .6)

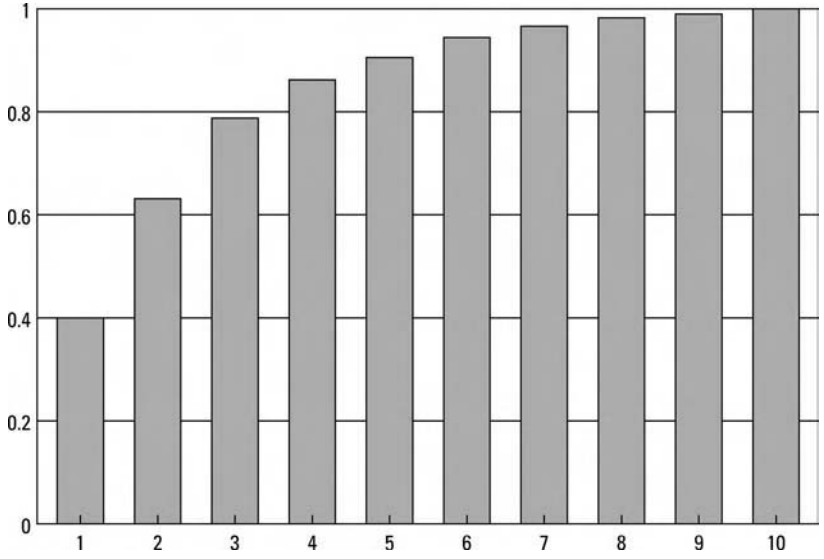

FIGURE 2.23 Cumulative probability functions for the Geometric Distribution (P = .6)

Suppose we are discussing tossing a single die. If we are talking about having the outcome of 5, how many times will we have to toss the die, on average, to achieve this outcome? The mean of the Geometric Distribution tells us this. If we know the probability of throwing a 5 is 1/6 (.1667), then the mean is 1/.1667 = 6. Thus, we would expect, on average, to toss a die six times in order to get a 5. If we kept repeating this process and recorded how many tosses it took until a 5 appeared, plotting these results would yield the Geometric Distribution function formulated in (2.39).

THE HYPERGEOMETRIC DISTRIBUTION

Another type of discrete distribution related to the preceding distributions is termed the *Hypergeometric Distribution*. Recall that in the Binomial Distribution it is assumed that each draw in succession from the population has the same probabilities. That is, suppose we have a deck of 52 cards; 26 of these cards are black and 26 are red. If we draw a card and record whether it is black or red, we then put the card back into the deck for the next draw. This "sampling with replacement" is what the Binomial Distribution assumes. Now, for the next draw, there is still a .5 (26/52) probability of the next card's being black (or red).

The Hypergeometric Distribution assumes almost the same thing, except there is no replacement after sampling. Suppose we draw the first card and it is red, and we *do not* replace it back into the deck. Now, the probability of the next draw's being red is reduced to 25/51 or .4901960784. In the Hypergeometric Distribution there is *dependency*, in that the probabilities of the next event are dependent on the outcome(s) of the prior event(s). Contrast this to the Binomial Distribution, where an event is *independent* of the outcome(s) of the prior event(s).

The basic functions N′(X) and N′(X) of the Hypergeometric are the same as those for the Binomial, (2.33) and (2.35), respectively, except that with the Hypergeometric the variable P, the probability of success on a single trial, changes from one trial to the next.

It is interesting to note the relationship between the Hypergeometric and Binomial Distributions. As N becomes larger, the differences between the computed probabilities of the Hypergeometric and the Binomial draw closer to each other. Thus, we can state that as N approaches infinity, the Hypergeometric approaches the Binomial as a limit.

If you want to use the Binomial probabilities as an approximation of the Hypergeometric, as the Binomial is far easier to compute, how big must the population be? It is not easy to state with any certainty, since the

desired accuracy of the result will determine whether the approximation is successful or not. Generally, though, a population to sample size of 100 to 1 is usually sufficient to permit approximating the Hypergeometric with the Binomial.

THE POISSON DISTRIBUTION

The *Poisson Distribution* is another important discrete distribution. This distribution is used to model arrival distributions and other seemingly random events that occur repeatedly yet haphazardly. These events can occur at points in time or at points along a wire or line (one dimension), along a plane (two dimensions), or in any N-dimensional construct. Figure 2.24 shows the arrival of events (the Xs) along a line, or in time.

The Poisson Distribution was originally developed to model incoming telephone calls to a switchboard. Other typical situations that can be modeled by the Poisson are the breakdown of a piece of equipment, the completion of a repair job by a steadily working repairman, a typing error, the growth of a colony of bacteria on a Petri plate, a defect in a long ribbon or chain, and so on.

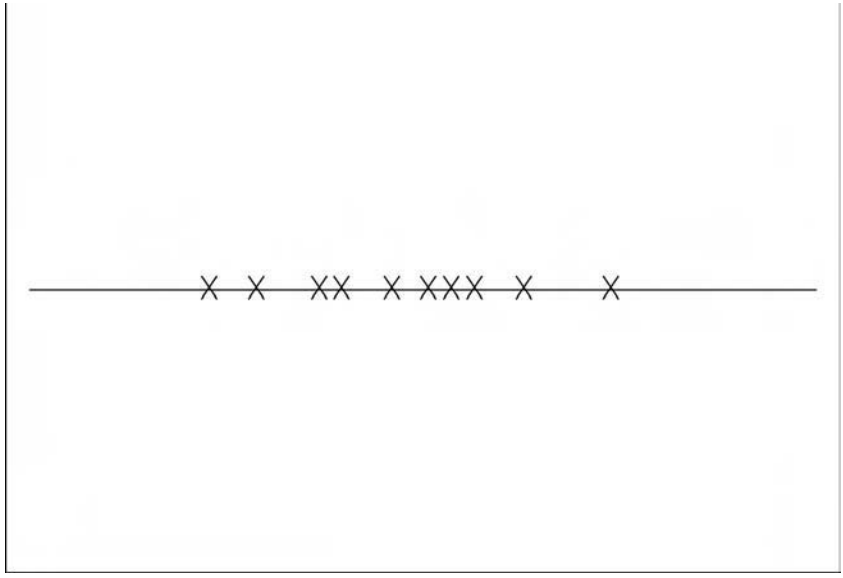

FIGURE 2.24 Sequence of haphazard events in time

The main difference between the Poisson and the Binomial distributions is that the Binomial is not appropriate for events that can occur more than once within a given time frame. Such an example might be the probability of an automobile accident over the next 6 months. In the Binomial we would be working with two distinct cases: Either an accident occurs, with probability P, or it does not, with probability Q (i.e., 1 – P). However, in the Poisson Distribution we can also account for the fact that more than one accident can occur in this time period.

The probability density function of the Poisson, N′(X), is given by:

$$N'(X) = (L^X * EXP(-L))/X! \tag{2.43}$$

where: L = The parameter of the distribution.
EXP( ) = The exponential function.

Note that X must take discrete values.

Suppose that calls to a switchboard average four calls per minute (L = 4). The probability of three calls (X = 3) arriving in the next minute is:

$$\begin{aligned} N'(3) &= (4^3 * EXP(-4))/3! \\ &= (64 * EXP(-4))/(3 * 2) \\ &= (64 * .01831564)/6 \\ &= 1.17220096/6 \\ &= .1953668267 \end{aligned}$$

So we can say there is about a 19.5% chance of getting three calls in the next minute. Note that this is not cumulative—that is, this is not the probability of getting three calls or fewer, it is the probability of getting exactly three calls. If we wanted to know the probability of getting three calls or fewer, we would have had to use the N(3) formula [which is given in (2.46)].

Other properties of the Poisson Distribution are:

$$\text{Mean} = L \tag{2.44}$$

$$\text{Variance} = L \tag{2.45}$$

where: L = The parameter of the distribution.

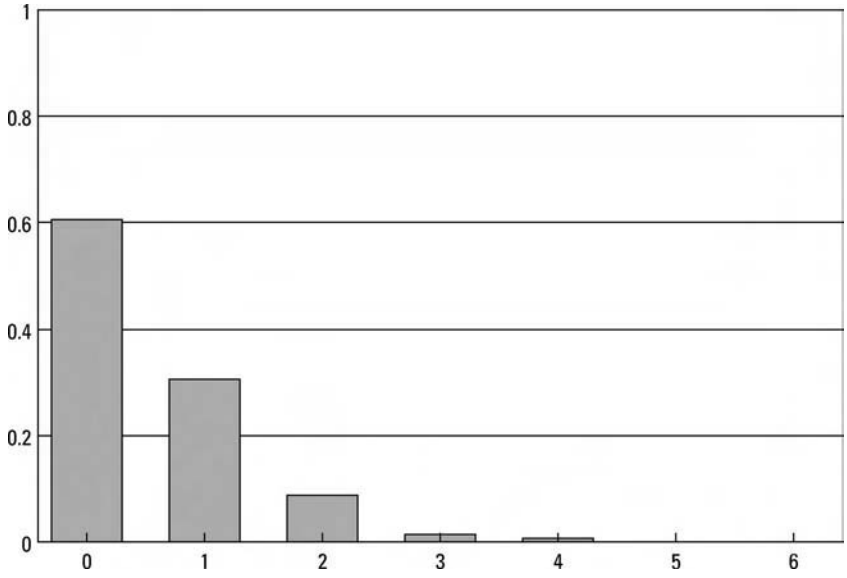

FIGURE 2.25 Probability density functions for the Poisson Distribution (L = .5)

In the Poisson Distribution, both the mean and the variance equal the parameter L. Therefore, in our example case we can say that the mean is four calls and the variance is four calls (or, the standard deviation is 2 calls—the square root of the variance, 4).

When this parameter, L, is small, the distribution is shaped like a reversed J, and when L is large, the distribution is not dissimilar to the Binomial. Actually, the Poisson is the limiting form of the Binomial as N approaches infinity and P approaches O. Figures 2.25 through 2.28 show the Poisson Distribution with parameter values of .5 and 4.5.

The cumulative density function of the Poisson, N(X), is given by:

$$N(X) = \sum_{J=0}^{X} (L^J * EXP(-L))/J! \tag{2.46}$$

where:

L = The parameter of the distribution.

EXP( ) = The exponential function.

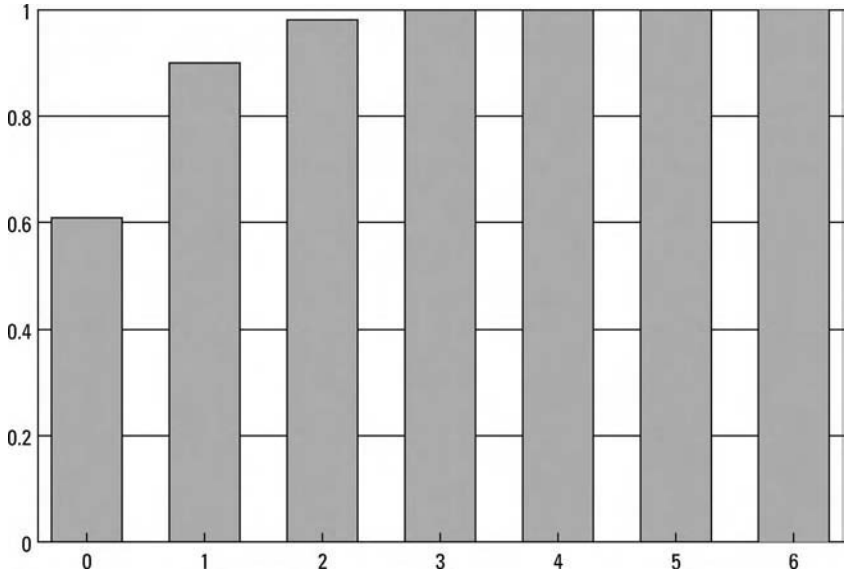

FIGURE 2.26 Cumulative probability functions for the Poisson Distribution (L = .5)

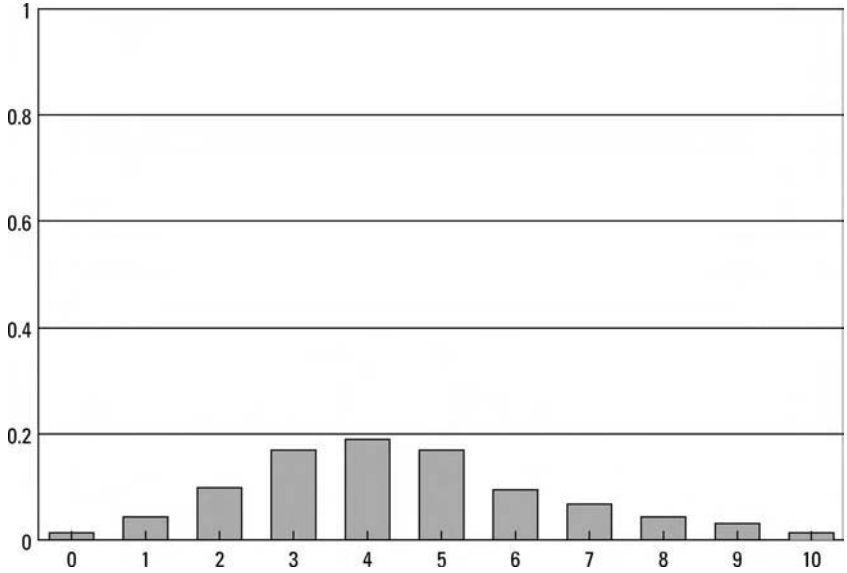

FIGURE 2.27 Probability density functions for the Poisson Distribution (L = 4.5)

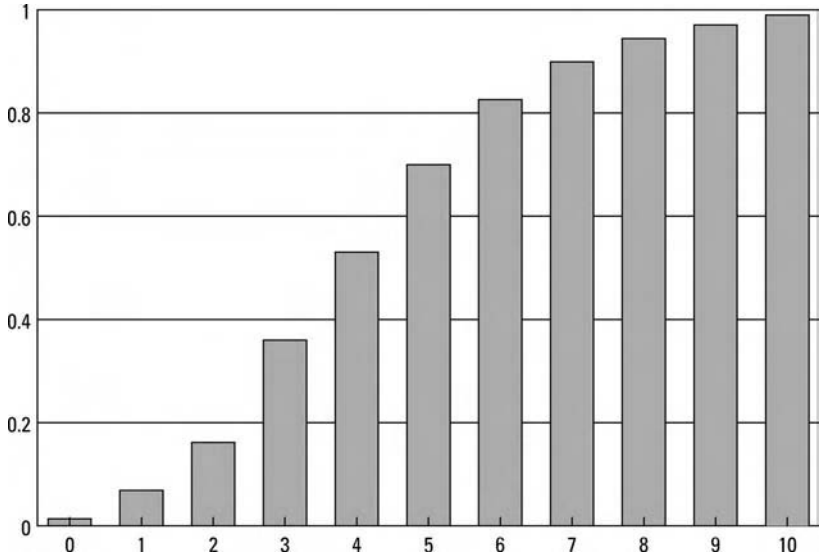

FIGURE 2.28 Cumulative probability functions for the Poisson Distribution (L = 4.5)

THE EXPONENTIAL DISTRIBUTION

Related to the Poisson Distribution is a continuous distribution with a wide utility called the *Exponential Distribution*, sometimes also referred to as the *Negative Exponential Distribution*. This distribution is used to model interarrival times in queuing systems; service times on equipment; and sudden, unexpected failures such as equipment failures due to manufacturing defects, light bulbs burning out, the time that it takes for a radioactive particle to decay, and so on. (There is a very interesting relationship between the Exponential and the Poisson Distributions. The arrival of calls to a queuing system follows a Poisson Distribution, with arrival rate L. The interarrival distribution (the time between the arrivals) is Exponential with parameter 1/L.)

The probability density function N′(X) for the Exponential Distribution is given as:

$$N'(X) = A * EXP(-A * X) \tag{2.47}$$

where: A = The single parametric input, equal to 1/L in the Poisson Distribution. A must be greater than 0.

EXP( ) = The exponential function.

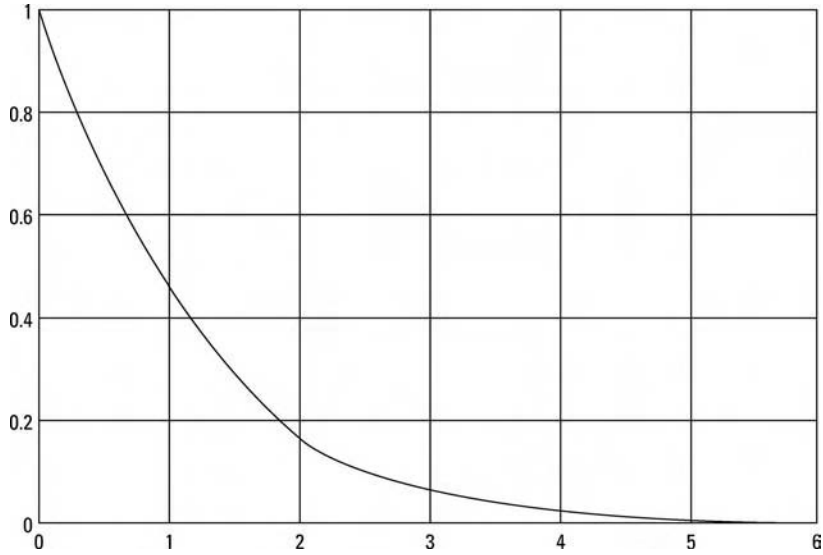

FIGURE 2.29 Probability density functions for the Exponential Distribution (A = 1)

The integral of (2.47), N(X), the cumulative density function for the Exponential Distribution is given as:

$$N(X) = 1 - EXP(-A * X) \tag{2.48}$$

where: A = The single parametric input, equal to 1/L in the Poisson Distribution. A must be greater than 0.

EXP( ) = The exponential function.

Figures 2.29 and 2.30 show the functions of the Exponential Distribution. Note that once you know A, the distribution is completely determined.

where: A = The single parametric input, equal to 1/L in the Poisson Distribution. A must be greater than 0.

EXP( ) = The exponential function.

Figures 2.29 and 2.30 show the functions of the Exponential Distribution. Note that once you know A, the distribution is completely determined.

The mean and variance of the Exponential Distribution are:

$$\text{Mean} = 1/A \tag{2.49}$$

$$\text{Variance} = 1/A^2 \tag{2.50}$$

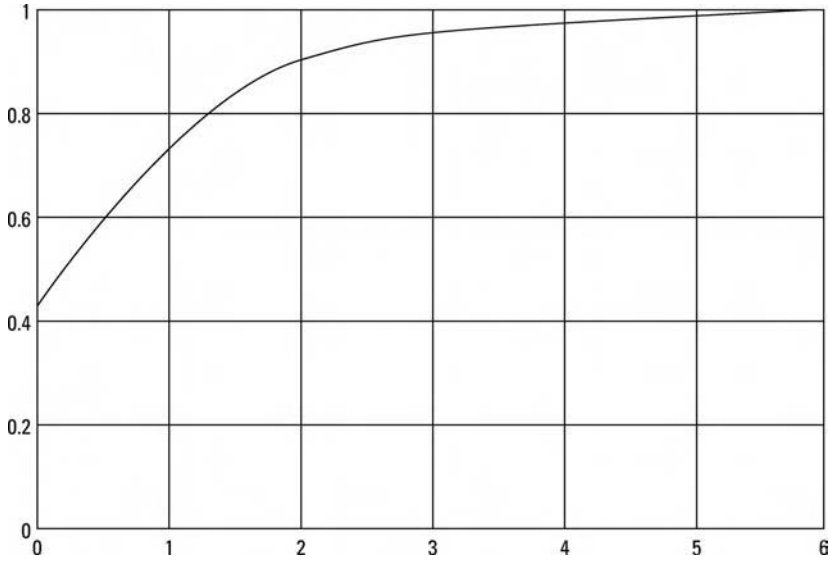

FIGURE 2.30 Cumulative probability functions for the Exponential Distribution (A = 1)

Again A is the single parametric input, equal to 1/L in the Poisson Distribution, and must be greater than 0.

Another interesting quality about the Exponential Distribution is that it has what is known as the "forgetfulness property." In terms of a telephone switchboard, this property states that the probability of a call in a given time interval is not affected by the fact that no calls may have taken place in the preceding interval(s).

THE CHI-SQUARE DISTRIBUTION

A distribution that is used extensively in goodness-of-fit testing is the *Chi-Square Distribution* (pronounced ki square, from the Greek letter X (chi) and hence often represented as the X^2 distribution).

Assume that K is a standard normal random variable (i.e., it has mean 0 and variance 1). If we say that K equals the square root of J ($J = K^2$) , then we know that K will be a continuous random variable. However, we know that K will not be less than zero, so its density function will differ from the Normal. The Chi-Square Distribution gives us the density function of K:

$$N'(K) = K^{V/2-1} * EXP(-V/2)/2^{V/2} * GAM(V/2) \qquad (2.51)$$

where: $K =$ The chi-square variable X^2.

$V =$ The number of degrees of freedom, which is the single input parameter.

$EXP(\,) =$ The exponential function.

$GAM(\,) =$ The standard gamma function.

A few notes on the gamma function are in order. This function has the following properties:

1. $GAM(O) \neq 1$
2. $GAM(1/2) \neq$ The square root of pi, or 1.772453851
3. $GAM(N) \neq (N - 1) * GAM(N - 1)$; therefore, if N is an integer, $GAM(N) = (N-1)!$

Notice in Equation (2.51) that the only input parameter is V, the number of degrees of freedom. Suppose that rather than just taking one independent random variable squared (K^2), we take M independent random variables squared, and take their sum:

$$JM = K_1^2 + K_2^2 \ldots K_M^2$$

Now J_M is said to have the Chi-Square Distribution with M degrees of freedom. It is the number of degrees of freedom that determines the shape of a particular Chi-Square Distribution. When there is one degree of freedom, the distribution is severely asymmetric and resembles the Exponential Distribution (with $A = 1$). At two degrees of freedom the distribution begins to look like a straight line going down and to the right, with just a slight concavity to it. At three degrees of freedom, a convexity starts taking shape and we begin to have a unimodal-shaped distribution. As the number of degrees of freedom increases, the density function gradually becomes more and more symmetric. As the number of degrees of freedom becomes very large, the Chi-Square Distribution begins to resemble the Normal Distribution per the Central Limit Theorem.

THE CHI-SQUARE "TEST"

Do not confuse the Chi-Square "Test" with the Chi-Square Distribution. The former is a hypothesis testing procedure (one of many such procedures). Mention is made of it here, but it should not be confused with the distributional form of the same name.

There exist a number of statistical tests designed to determine if two samples come from the same population. Essentially, we want to know if

two distributions are different. Perhaps the most well known of these tests is the chi-square test, devised by Karl Pearson around 1900. It is perhaps the most popular of all statistical tests used to determine whether two distributions are different.

The chi-square statistic, X^2, is computed as:

$$X^2 = \sum_{i=1}^{N} (O_i - E_i)^2 / E_i \tag{2.52}$$

where: $N =$ The total number of bins.
$O_i =$ The number of events observed in the ith bin.
$E_i =$ The number of events expected in the ith bin.

A large value for the chi-square statistic indicates that it is unlikely that the two distributions are the same (i.e., the two samples are not drawn from the same population). Likewise, the smaller the value for the chi-square statistic, the more likely it is that the two distributions are the same (i.e., the two samples were drawn from the same population).

Note that the observed values, the O_i's, will always be integers. However, the expected values, the E_i's, can be nonintegers. Equation (2.52) gives the chi-square statistic when both the expected and observed values are integers. When the expected values, the E_i's, are permitted to be nonintegers, we must use a different equation, known as *Yates' correction*, to find the chi-square statistic:

$$X^2 = \sum_{i=1}^{N} (ABS\,(O_i - E_i) - .5)^2 / E_i \tag{2.53}$$

where: $N =$ The total number of bins.
$O_i =$ The number of events observed in the ith bin.
$E_i =$ The number of events expected in the ith bin.
$ABS() =$ The absolute value function.

We can convert a chi-square statistic such as 37.5336 to a *sigfinicance level*. In the sense we are using here, a significance level is a number between 0, representing that the two distributions are different, and 1, meaning that the two distributions are the same. We can never be 100% certain that two distributions are the same (or different), but we can determine how alike or different two distributions are to a certain significance level. There are two ways in which we can find the significance level. This first and by far the simplest way is by using tables. The second way to convert a chi-square statistic to a significance level is to perform the math yourself

(which is how the tables were drawn up in the first place). However, the math requires the use of incomplete gamma functions, not be treated in this text. However, most readers who would want to know how to calculate a significance level from a given chi-square statistic would want to know this because tables are rather awkward to use from a programming standpoint. Therefore, what follows is a snippet of Java code to convert from a given chi-square statistic to a significance level.

```
Public void ChiSquareTest(int nmbrOfBins, double chiSquareStatistic){
      double confidenceLevel = 1.0;
      double a = 0.0, b = 0.0, c=1.0, d = 0.0, e = 0.0,
      f = 1.0;
      int nbins = nmbrOfBins -3;
      System.out.println("Chi-Square Statistic at " + nbins + "degrees of
freedom is "+chiSquareStatistic);
      if(chiSquareStatistic < 31.0 || nbins > 2){
        e = nbins/2 -1;
        a = 1;
        for(int i = 1;i <=nbons/2 - .5; i++){
          a *=e;
          e -=1.0;
        }
        if(nbins% 2 !=0){
          a *= 1.77245374942627;
        }
        b = Math.pow((chiSquareStatistic/2.0), (double) (nbins/2)) * 2.0/
(Math.exp(chiSquareStatistic/2.0) * a * nbins);
        d = nbins + 2;
        do{
        c *=chiSquareStatistic/d;
        f+=c;
        d+=2.0;
        }while(c > 0.0);
        confidenceLevel = 1.0 - b *f;
      }
      System.out.println("For a Significance level of "+confidenceLevel);
}
```

Whether you determine your significance levels via a table or calculate them yourself, you will need two parameters to determine a significance level. The first of these parameters is, of course, the chi-square statistic itself. The second is the number of *degrees of freedom*. Generally, the number of degrees of freedom is equal to the number of bins minus 1 minus the number of population parameters that have to be estimated for the sample statistics. What follows is a small table for converting between chi-square values and degrees of freedom to significance levels:

Values of X^2

Degrees of Freedom	Significance Level			
	.20	.10	.05	.01
1	1.6	2.7	3.8	6.6
2	3.2	4.6	6.0	9.2
3	4.6	6.3	7.8	11.3
4	6.0	7.8	9.5	13.3
5	7.3	9.2	11.1	15.1
10	13.4	16.0	18.3	23.2
20	25.0	28.4	31.4	37.6

You should be aware that the chi-square test can do a lot more than is presented here. For instance, you can use the chi-square test on a 2×2 contingency table (actually on any $N \times M$ contingency table).

Finally, there is the problem of the arbitrary way we have chosen our bins as regards both their number and their range. Recall that binning data involves a certain loss of information about that data, but generally the profile of the distribution remains relatively the same. If we choose to work with only three bins, or if we choose to work with 30, we will likely get somewhat different results. It is often a helpful exercise to bin your data in several different ways when conducting statistical tests that rely on binned data. In so doing, you can be rather certain that the results obtained were not due solely to the arbitrary nature of how you chose your bins.

In a purely statistical sense, in order for our number of degrees of freedom to be valid, it is necessary that the number of elements in each of the expected bins, the E_i's, be at least five. When there is a bin with less than five expected elements in it, theoretically the number of bins should be reduced until all of the bins have at least five expected elements in them. Often, when only the lowest and/or highest bin has less than five expected elements in it, the adjustment can be made by making these groups "all less than" and "all greater than" respectively.

THE STUDENT'S DISTRIBUTION

The *Student's Distribution*, sometimes called the *t Distribution* or *Student's t*, is another important distribution used in hypothesis testing that is related to the Normal Distribution. When you are working with less than 30 samples of a near-Normally distributed population, the Normal Distribution can no longer be accurately used. Instead, you must use the Student's Distribution. This is a symmetrical distribution with one parametric input, again the degrees of freedom. The degrees of freedom usually equals the number of elements in a sample minus one (N – 1).

The shape of this distribution closely resembles the Normal except that the tails are thicker and the peak of the distribution is lower. As the number of degrees of freedom approaches infinity, this distribution approaches the Normal in that the tails lower and the peak increases to resemble the Normal Distribution. When there is one degree of freedom, the tails are at their thickest and the peak at its smallest. At this point, the distribution is called *Cauchy*.

It is interesting that if there is only one degree of freedom, then the mean of this distribution is said not to exist. If there is more than one degree of freedom, then the mean does exist and is equal to zero, since the distribution is symmetrical about zero. The variance of the Student's Distribution is infinite if there are fewer than three degrees of freedom.

The concept of *infinite variance* is really quite simple. Suppose we measure the variance in daily closing prices for a particular stock for the last month. We record that value. Now we measure the variance in daily closing prices for that stock for the next year and record that value. Generally, it will be greater than our first value, of simply last month's variance. Now let's go back over the last five years and measure the variance in daily closing prices. Again, the variance has gotten larger. The farther back we go—that is, the more data we incorporate into our measurement of variance—the greater the variance becomes. Thus, the variance increases without bound as the size of the sample increases. This is infinite variance. The distribution of the log of daily price changes appears to have infinite variance, and thus the Student's Distribution is sometimes used to model the log of price changes. (That is, if C_0 is today's close and C_1 yesterday's close, then $\ln(C_0/C_1)$ will give us a value symmetrical about 0. The distribution of these values is sometimes modeled by the Student's distribution).

If there are three or more degrees of freedom, then the variance is finite and is equal to:

$$\text{Variance} = V/(V-2) \qquad \text{for } V > 2 \tag{2.54}$$

$$\text{Mean} = 0 \qquad \text{for } V > 1 \tag{2.55}$$

where: $V =$ The degrees of freedom.

Suppose we have two independent random variables. The first of these, Z, is standard normal (mean of 0 and variance of 1). The second of these, which we call J, is Chi-Square distributed with V degrees of freedom. We can now say that the variable T, equal to Z/(J/V), is distributed according to the Student's Distribution. We can also say that the variable T will follow the Student's Distribution with N – 1 degrees of freedom if:

$$T = \sqrt{N} * (X - U)/S$$

where: X = A sample mean.
S = A sample standard deviation.
N = The size of a sample.
U = The population mean.

The probability density function for the Student's Distribution, N′(X), is given as:

$$N'(X) = \frac{GAM((V+1)/2)}{\sqrt{V * P} * GAM(V/2)} * (1 + X^2/V)^{-(V+1)/2} \tag{2.56}$$

where: P = pi, or 3.1415926536.
V = The degrees of freedom.
GAM( ) = The standard gamma function.

The mathematics of the Student's Distribution are related to the incomplete beta function. Since we aren't going to plunge into functions of mathematical physics such as the incomplete beta function, we will leave the Student's Distribution at this point. Before we do, however, you still need to know how to calculate probabilities associated with the Student's Distribution for a given number of standard units (Z score) and degrees of freedom. As the following snippet of java code to discern the probabilities. You'll note that as the degrees of freedom variable, DEGFDM, approaches infinity, the values returned, the probabilities, converge to the Normal as given by Equation (2.22):

```
public void StudentsT2TailProbs(double zScore, int degreesOfFreedom){
    double confidenceLevel = 1.0;
    double st = Math.abs(zScore);
    double r8 = Math.atan(st/Math.sqrt((double)degreesOfFreedom));
    double rc8 = Math.cos(r8);
    double x8 = 1.0;
```

```
double r28 = rc8 * rc8;
double rs8 = Math.sin(r8);
double y8 = r8;
double y8 = r8;
if(degreesOfFreedom %2 !=0 ){
  if(degreesOfFreedom !=1){
    y8 = rc8;
    for(int i =3;i<=degreesOfFreedom - 2; i+=2){
      x8 = x8 * r28 * (double)((i - 1)/i);
      y8 = y8 + x8 * rc8;
    }
    y8 = r8 + rs8 * y8;
  }
  confidenceLevel = y8 * 0.6366197723657157;
}else{
  y8=1.0;
  for(int i =2;i <=degreesOfFreedom -2; i+=2){
    x8 = x8 * r28 * (double)((i - 1)/i);
    y8 += x8;
  }
  confidenceLevel = y8 *rs8;
  }
  System.out.println("The two-tailed probabilities associated with
the T distribution for a Z score of "+zScore+" and "+degreesOfFreedom+
"degrees freedom is "+confidenceLevel);
}
```

Next, we come to another distribution, related to the Chi-Square Distribution, that also has important uses in statistics. The *F Distribution*, sometimes referred to as *Snedecor's Distribution* or *Snedecor's F*, is useful in hypothesis testing. Let A and B be independent chi-square random variables with degrees of freedom of M and N, respectively. Now the random variable:

$$F = (A/M)/(B/N)$$

can be said to have the F Distribution with M and N degrees of freedom. The density function, $N'(X)$, of the F Distribution is given as:

$$N'(X) = \frac{GAM((M+N)/2) * (M/N)^{M/2}}{(GAM(M/2) * GAM(N/2) * (1+M/N))^{(M+N)/2}} \quad (2.57)$$

where: M = The number of degrees of freedom of the first parameter.
N = The number of degrees of freedom of the second parameter.
GAM( ) = The standard gamma function.

THE MULTINOMIAL DISTRIBUTION

The *Multinomial Distribution* is related to the Binomial, and likewise is a discrete distribution. Unlike the Binomial, which assumes two possible outcomes for an event, the Multinomial assumes that there are M different outcomes for each trial. The probability density function, N′(X), is given as:

$$N'(X) = \left(N! \Big/ \prod_{i=1}^{M} N_i! \right) * \prod_{i=1}^{M} P_i^{N_i} \quad (2.58)$$

where: N = The total number of trials.
N_i = The number of times the ith trial occurs.
P_i = The probability that outcome number i will be the result of any one trial. The summation of all P_i's equals 1.
M = The number of possible outcomes on each trial.

For example, consider a single die where there are six possible outcomes on any given roll (M = 6). What is the probability of rolling a 1 once, a 2 twice, and a 3 three times out of 10 rolls of a fair die? The probabilities of rolling a 1, a 2, or a 3 are each 1/6. We must consider a fourth alternative to keep the sum of the probabilities equal to 1, and that is the probability of not rolling a 1, 2, or 3, which is 3/6. Therefore, $P_1 = P_2 = P_3 = 1/6$, and $P_4 = 3/6$. Also, $N_1 = 1$, $N_2 = 2$, $N_3 = 3$, and $N_4 = 10 - 3 - 2 - 1 = 4$. Therefore, Equation (2.58) can be worked through as:

$$\begin{aligned} N'(X) &= (10!/(1! * 2! * 3! * 4!)) * (1/6)^1 * (1/6)^2 * (1/6)^3 * (3/6)^4 \\ &= (3628800/(1 * 2 * 6 * 24)) * .1667 * .0278 * .00463 * .0625 \\ &= (3628800/288) * .000001341 \\ &= 12600 * .000001341 \\ &= .0168966 \end{aligned} \quad (2.58)$$

Note that this is the probability of rolling exactly a 1 once, a 2 twice, and a 3 three times, not the cumulative density. This is a type of distribution that uses more than one random variable; hence, its cumulative density cannot be drawn out nicely and neatly in two dimensions as you could with the other distributions discussed thus far. We will not be working with other distributions that have more than one random variable, but you should be aware that such distributions and their functions do exist.

THE STABLE PARETIAN DISTRIBUTION

The *Stable Paretian Distribution* is actually an entire class of distributions, sometimes referred to as "Pareto-Levy" distributions. The probability density function N′(U) is given as:

$$\ln(N''(U)) = i * D * U - V * ABS(U)^{A} * Z \tag{2.59}$$

where:
- U = The variable of the stable distribution.
- A = The kurtosis parameter of the distribution.
- B = The skewness parameter of the distribution.
- D = The location parameter of the distribution.
- V = This is also called the scale parameter.
- i = The imaginary unit, $\sqrt{-1}$
- Z = $1 - i * B * (U/ABS(U)) * \tan(A * 3.1415926536/2)$ when $A > < 1$ and $1 + i * B * (U/ABS(U)) * 2/3.1415926536 * \log(ABS(U))$ when $A = 1$.
- ABS( ) = The absolute value function.
- tan( ) = The tangent function.
- ln( ) = The natural logarithm function.

The limits on the parameters of Equation (2.59) are:

$$0 < A <= 2 \tag{2.60}$$

$$-1 <= B <= 1 \tag{2.61}$$

$$0 <= V \tag{2.62}$$

The four parameters of the distribution—A, B, D, and V—allow the distribution to assume a great many different shapes.

The variable A measures the height of the tails of the distribution. Thus, we can say that A represents the kurtosis variable of the distribution. A is also called the *characteristic exponent* of the distribution. When A equals 2, the distribution is Normal, and when A equals 1 the distribution

is Cauchy. For values of A that are less than 2, the tails of the distribution are higher than with the Normal Distribution. The total probability in the tails increases as A decreases. When A is less than 2, the variance is infinite. The mean of the distribution exists only if A is greater than 1.

The variable B is the *index of skewness*. When B equals zero, the distribution is perfectly symmetrical. The degree of skewness is larger the larger the absolute value of B. Notice that when A equals 2, W(U,A) equals 0; hence, B has no effect on the distribution. In this case, when A equals 2, no matter what B is, we still have the perfectly symmetrical Normal Distribution. The *scale parameter*, V, is sometimes written as a function of A, in that $V = C^A$, therefore $C = V^{1/A}$. When A equals 2, V is one-half the variance. When A equals 1, the Cauchy Distribution, V is equal to the semi-interquartile range. D is the *location parameter*. When A is equal to 2, the arithmetic mean is an unbiased estimator of D; when A is equal to 1, the median is.

The cumulative density functions for the stable Paretian are not known to exist in closed form. For this reason, evaluation of the parameters of this distribution is complex, and work with this distribution is made more difficult. It is interesting to note that the stable Paretian parameters A, B, C, and D correspond to the fourth, third, second, and first moments of the distribution, respectively. This gives the stable Paretian the power to model many types of real-life distributions—in particular, those where the tails of the distribution are thicker than they would be in the Normal, or those with infinite variance (i.e., when A is less than 2). For these reasons, the stable Paretian is an extremely powerful distribution with applications in economics and the social sciences, where data distributions often have those characteristics (fatter tails and infinite variance) that the stable Paretian addresses.

This infinite variance characteristic makes the Central Limit Theorem inapplicable to data that is distributed per the stable Paretian distribution when A is less than 2. This is a very important fact if you plan on using the Central Limit Theorem.

One of the major characteristics of the stable Paretian is that it is invariant under addition. This means that the sum of independent stable variables with characteristic exponent A will be stable, with approximately the same characteristic exponent. Thus, we have the Generalized Central Limit Theorem, which is essentially the Central Limit Theorem, except that the limiting form of the distribution is the stable Paretian rather than the Normal, and the theorem applies even when the data has infinite variance (i.e., $A < 2$), which is when the Central Limit Theorem does not apply. For example, the heights of people have finite variance. Thus, we could model the heights of people with the Normal Distribution. The distribution of people's incomes, however, does not have finite variance and is therefore modeled by the stable Paretian distribution rather than the Normal Distribution.

It is because of this Generalized Central Limit Theorem that the stable Paretian Distribution is believed by many to be representative of the distribution of price changes.

There are many more probability distributions that we could still cover (Negative Binomial Distribution, Gamma Distribution, Beta Distribution, etc.); however, they become increasingly more obscure as we continue from here. The distributions we have covered thus far are, by and large, the main common probability distributions.

Efforts have been made to catalogue the many known probability distributions. Undeniably, one of the better efforts in this regard has been done by Karl Pearson, but perhaps the most comprehensive work done on cataloguing the many known probability distributions has been presented by Frank Haight.[5] Haight's "Index" covers almost all of the known distributions on which information was published prior to January, 1958. Haight lists most of the mathematical functions associated with most of the distributions. More important, references to books and articles are given so that a user of the index can find what publications to consult for more in-depth matter on the particular distribution of interest. Haight's index categorizes distributions into ten basic types: (1) Normal; (2) Type III; (3) Binomial; (4) Discrete; (5) Distributions on (A, B); (6) Distributions on (0, infinity); (7) Distributions on (–infinity, infinity); (8) Miscellaneous Univariate; (9) Miscellaneous Bivariate; (10) Miscellaneous Multivariate.

Of the distributions we have covered in this Chapter, the Chi-Square and Exponential (Negative Exponential) are categorized by Haight as Type III. The Binomial, Geometric, and Bernoulli are categorized as Binomial. The Poisson and Hypergeometric are categorized as Discrete. The Rectangular is under Distributions on (A, B), the F Distribution as well as the Pareto are under Distributions on (0, infinity), the Student's Distribution is regarded as a Distribution on (–infinity, infinity), and the Multinomial as a Miscellaneous Multivariate. It should also be noted that not all distributions fit cleanly into one of these ten categories, as some distributions can actually be considered subclasses of others. For instance, the Student's distribution is catalogued as a Distribution on (–infinity, infinity), yet the Normal can be considered a subclass of the Student's, and the Normal is given its own category entirely. As you can see, there really isn't any "clean" way to categorize distributions. However, Haight's index is quite thorough. Readers interested in learning more about the different types of distributions should consult Haight as a starting point.

---

[5]Haight, F. A., "Index to the Distributions of Mathematical Statistics," *Journal of Research of the National Bureau of Standards-B. Mathematics and Mathematical Physics* 65B No. 1, pp. 23–60, January–March 1961.