TEACHING NOTE 00-06:
PROBABILITY AND STATISTICS REVIEW FOR FINANCE: PART II

Version date: October 24, 2000 C:\Classes\Tn00-06.wpd

This teaching note continues TN99-04. In that note, we looked at some of the mathematical theory that provides the foundations for the probability principles used in finance. Here we look at some of the important results dealing with the use of random variables, i.e., the concepts of expected values, variances and covariances.

Discrete Random Variables

A random variable is a variable that can take on many possible values representing uncertain outcomes whose frequencies of occurrence are governed by a probability distribution. A discrete random variable can take on only a finite number of values. For example, the number of people who respond "yes" to a survey or the number of laboratory mice who died following an experiment are examples of a discrete random variable. In contrast a continuous random variable can take on an infinite number of values. For example, the height of a person selected randomly or the amount of time elapsed following an event can always be expressed with an infinite number of decimal places.

For a discrete random variable there are a finite number of outcomes that we often call *states* or *states of the world*. For example, if the event is the selection of a person and we are interested in whether that person is male or female, we would have two states. Let $X = x_1$ if that person is male and $X = x_2$ if that person is female. In general, we specify n states and n possible values of X: x_1, $x_2, \ldots, x_n$. We often need to characterize the properties of the random variable X. Note that we are using upper case X to represent the random variable and lower case x with a subscript to represent a specific outcome or value of that random variable. We let p_i be the probability that state i occurs. Note that by definition $\sum_{i=1}^{n} p_i = 1$. It may well be the case that p_i is given by some mathematical function that might more appropriately be written as $f(x_i)$, but here we shall just leave the probability specification in the form p_i.

In the following sub-sections we derive the most important results for relationships regarding expected values, variances and covariances of discrete random variables.

1. Expected Value

The expected value, sometimes called the mean, is the probability-weighted average value of X. It is

$$E(X) = \sum_{i=1}^{n} x_i p_i.$$

Occasionally we shall have to work with a constant such as c in the following operations.

$$E(c) = \sum_{i=1}^{n} c p_i = c \sum_{i=1}^{n} p_i = c$$
$$E(cX) = c \sum_{i=1}^{n} x_i p_i = cE(X).$$

Thus, *the expected value of a constant is just the constant; the expected value of a constant times a random variable is the constant times the expected value of the random variable.* The Greek letter µ is often used for the expected value, but other Greek letters and many other symbols are also used as the expected value.

2. Variance

The variance is a measure of the dispersion of the distribution. It is defined as the probability-weighted squared deviation of each possible value from the expected value. Using that construction, we see that the variance can be converted to another useful version:

$$\begin{aligned} Var(X) &= \sum_{i=1}^{n} (x_i - E(X))^2 p_i \\ &= \sum_{i=1}^{n} (x_i - E(X))(x_i - E(X)) p_i = \sum_{i=1}^{n} (x_i^2 - 2E(X)x_i + [E(X)]^2) p_i \\ &= \sum_{i=1}^{n} x_i^2 p_i - 2E(X) \sum_{i=1}^{n} x_i p_i + [E(X)]^2 \sum_{i=1}^{n} p_i = E(X^2) - 2[E(X)]^2 + [E(X)]^2 \\ &= E(X^2) - [E(X)]^2. \end{aligned}$$

In other words *the variance is also the expected value of the squared value of X minus the square of the expected value of X.*

When we work with constants, the variance is affected in the following manner:

$$\begin{aligned} Var(cX) &= \sum_{i=1}^{n} (cx_i - E(cX))^2 p_i \\ &= E((cX)^2) - [E(cX)]^2 = E(c^2X^2) - c^2[E(X)]^2 \\ &= c^2E(X^2) - c^2[E(X)]^2 = c^2(E(X^2) - [E(X)]^2) \\ &= c^2 Var(X). \end{aligned}$$

In other words, *the variance of a constant times a random variable is the constant squared times the variance of the random variable*. *The variance of a constant is zero* as shown by the following:

$$\begin{aligned} Var(c) &= \sum_{i=1}^{n} (c - E(c))^2 p_i \\ &= \sum_{i=1}^{n} (c - c)^2 p_i \\ &= 0. \end{aligned}$$

Finally, we also show that *the variance of a constant plus a random variable is the variance of the random variable*:

$$\begin{aligned} Var(c+X) &= \sum_{i=1}^{n} (c + x_i - E(c+X))^2 p_i \\ &= \sum_{i=1}^{n} (c - E(c) + x_i - E(X))^2 p_i = \sum_{i=1}^{n} (c - c + x - E(X))^2 p_i \\ &= \sum_{i=1}^{n} (x_i - E(X))^2 p_i \\ &= Var(X). \end{aligned}$$

The square root of the variance is the standard deviation. Using σ as its symbol, we have σ(cX) = c(σ(X)).

3. Covariance

An important concept somewhat similar to the variance is the *covariance*. It measures the extent to which two random variables move together. Now we need another random variable, which we shall refer to as Y. The covariance between X and Y is given as

$$
\begin{aligned}
Cov(X,Y) &= \sum_{i=1}^{n} (x_i - E(X))(y_i - E(Y))p_i \\
&= \sum_{i=1}^{n} (x_i y_i - E(X)y_i - x_i E(Y) + E(X)E(Y))p_i \\
&= \sum_{i=1}^{n} x_i y_i p_i - E(X)\sum_{i=1}^{n} y_i p_i - E(Y)\sum_{i=1}^{n} x_i p_i + E(X)E(Y)\sum_{i=1}^{n} p_i \\
&= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y).
\end{aligned}
$$

The covariance by itself is difficult to interpret. A positive (negative) covariance means that the two variables tend to move together (opposite) in a linear fashion. A zero covariance implies that there is no linear relationship between the two variables, but that does not rule out a non-linear relationship. A more useful measure of association is the *correlation*, defined as

$$
\rho(XY) = \frac{Cov(XY)}{\sigma(X)\sigma(Y)},
$$

where σ(X) and σ(Y) are the standard deviations of X and Y, respectively. The correlation ranges between -1 and +1.

An important result, which is sometimes seen in finance, is that *the covariance of a variable with itself is the variance*. We can easily see this by letting Y also be X:

$$
\begin{aligned}
Cov(X,X) &= \sum_{i=1}^{n} (x_i - E(X))(x_i - E(X))p_i \\
&= Var(X).
\end{aligned}
$$

The covariance of a random variable and a constant is zero:

$$\begin{aligned} Cov(X,c) &= \sum_{i=1}^{n} (x_i - E(X))(c_i - E(c))p_i \\ &= \sum_{i=1}^{n} (x_i - E(X))(c - c)p_i \\ &= 0. \end{aligned}$$

The covariance concept facilitates the understanding of the variance of a combination of more than one random variable. Consider a weighted sum of variables X and Y, obtained by multiplying X by a and Y by b. We wish to find the variance of aX + bY:

$$\begin{aligned} Var(aX+bY) &= \sum_{i=1}^{n} (ax_i+by_i)^2 p_i - \left(\sum_{i=1}^{n} (ax_i p_i + by_i p_i)\right)^2 \\ &= \sum_{i=1}^{n} (ax_i+by_i)(ax_i+by_i)p_i - \left(a\sum_{i=1}^{n} x_i p_i + b\sum_{i=1}^{n} y_i p_i\right)^2 \\ &= \sum_{i=1}^{n} (a^2x_i^2 + 2abx_iy_i + b^2y_i^2)p_i - \big(aE(X) + bE(Y)\big)^2 \\ &= a^2\sum_{i=1}^{n} x_i^2 p_i + 2ab\sum_{i=1}^{n} x_i y_i p_i + b^2\sum_{i=1}^{n} y_i^2 p_i - \big(a^2[E(X)]^2 + 2abE(X)E(Y) + b^2[E(Y)]^2\big) \\ &= a^2E(X^2) + 2abE(XY) + b^2E(Y^2) - a^2[E(X)]^2 - 2abE(X)E(Y) - b^2[E(Y)]^2 \\ &= a^2E(X^2) - a^2[E(X)]^2 + b^2E(Y^2) - b^2[E(Y)]^2 + 2abE(XY) - 2abE(X)E(Y) \\ &= a^2(E(X^2) - [E(X)]^2) + b^2(E(Y^2) - [E(Y)]^2) + 2ab(E(XY) - E(X)E(Y)) \\ &= a^2Var(X) + b^2Var(Y) + 2abCov(X,Y) \end{aligned}$$

Thus, we see that *the variance of a weighted sum of random variables is a weighted average of the variances of the individual random variables plus twice the weighted covariance*.

Now let us find the covariance between the weighted variables, i.e., Cov(aX,bY):

$$
\begin{aligned}
Cov(aX,bY) &= \sum_{i=1}^{n} (ax_i - E(aX))(by_i - E(bY))p_i \\
&= \sum_{i=1}^{n} (ax_i - aE(X))(by_i - bE(Y))p_i \\
&= \sum_{i=1}^{n} (ax_i by_i - ax_i bE(Y) - aE(X)by_i + abE(X)E(Y))p_i \\
&= ab\sum_{i=1}^{n} x_i y_i p_i - abE(Y)\sum_{i=1}^{n} x_i p_i - abE(X)\sum_{i=1}^{n} y_i p_i + abE(X)E(Y)\sum_{i=1}^{n} p_i \\
&= abE(XY) - abE(Y)E(X) - abE(X)E(Y) + abE(X)E(Y) \\
&= ab(E(XY) - E(X)E(Y)) \\
&= abCov(X,Y).
\end{aligned}
$$

We may also wish to know what happens to the covariance if we add constants to X and Y. Let c and g be constants:

$$
\begin{aligned}
Cov(c+X,g+Y) &= \sum_{i=1}^{n} (c + x_i - E(c+X))(g + y_i - E(g+Y))p_i \\
&= \sum_{i=1}^{n} (c - E(c) + x_i - E(X))(g - E(g) + y_i - E(Y))p_i \\
&= \sum_{i=1}^{n} (c - c + x_i - E(X)(g - g + y_i - E(Y))p_i \\
&= \sum_{i=1}^{n} (x_i - E(X))(y_i - E(Y))p_i \\
&= Cov(X,Y).
\end{aligned}
$$

Thus, *adding constants does not change the covariance between two variables*.

Finally, we might be interested in the covariance between the sum of two random variables and a third random variable. Letting the third random variable be Z, we want to know Cov(X +Y,Z):

$$
\begin{aligned}
Cov(X+Y,Z) &= \sum_{i=1}^{n} (x_i + y_i - E(X+Y))(z_i - E(Z))p_i \\
&= \sum_{i=1}^{n} [(x_i - E(X)) + (y_i - E(Y))](z_i - E(Z))p_i \\
&= \sum_{i=1}^{n} (x_i - E(X))(z_i - E(Z))p_i + \sum_{i=1}^{n} (y_i - E(Y))(z_i - E(Z))p_i \\
&= Cov(X,Z) + Cov(Y,Z).
\end{aligned}
$$

Continuous Random Variables

For continuous random variables, the operations are somewhat different but the results are identical. The expected value and variance of a continuous random variable X are given as

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
Var(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx, \\
Cov(X,Y) &= \int_{-\infty}^{\infty} (x - E(X))(y - E(Y)) f(xy) dxy,
\end{aligned}
$$

where f(x) is the density function that describes the outcomes x of the random variable X in terms of the infinite number of states of the world. In the third equation above, f(xy) is the joint density of X and Y, which gives the probability of outcomes of both X and Y occurring. The single integral in the covariance equation means to integrate over all joint outcomes. Finally, let us make sure we remember that when we integrate the density function over all outcomes, we obtain the result:

$$
\int_{-\infty}^{\infty} f(x) dx = 1.
$$

In the following sub-sections we derive the most important results for relationships regarding expected values, variances and covariances of continuous random variables.

1. Expected Value

We already presented the definition of an expected value in continuous time. If X is actually a constant c, then we have the following results.

$$E(c) = \int_{-\infty}^{\infty} cf(x)dx = c\int_{-\infty}^{\infty} f(x)dx = c$$

$$E(cX) = \int_{-\infty}^{\infty} cxf(x)dx = c\int_{-\infty}^{\infty} xf(x)dx = cE(X).$$

2. Variance

The key results for the variance are given as follows. First note that the variance itself can be restated as

$$\begin{aligned}
Var(X) &= \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx \\
&= \int_{-\infty}^{\infty} (x^2 - 2xE(X) + [E(X)]^2) f(x)dx \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - 2E(X)\int_{-\infty}^{\infty} xf(x)dx + [E(X)]^2 \int_{-\infty}^{\infty} f(x)dx \\
&= E(X^2) - 2[E(X)]^2 + [E(X)]^2 \\
&= E(X^2) - [E(X)]^2.
\end{aligned}$$

If x is a constant, we have the following results:

$$
\begin{aligned}
Var(c) &= \int_{-\infty}^{\infty} (c - E(c))^2 f(x)dx \\
&= \int_{-\infty}^{\infty} (c - c)^2 f(x)dx \\
&= 0 \\
Var(cX) &= \int_{-\infty}^{\infty} (cx - E(cX))(cx - E(cX))f(x)dx \\
&= \int_{-\infty}^{\infty} c(x - E(X))c(x - E(X))f(x)dx = c^2 \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx \\
&= c^2 Var(X). \\
Var(c + X) &= \int_{-\infty}^{\infty} (c + x - E(c + X))^2 f(x)dx \\
&= \int_{-\infty}^{\infty} (c - E(c) + x - E(X))^2 f(x)dx = \int_{-\infty}^{\infty} (c - c + x - E(X))^2 f(x)dx \\
&= \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx = Var(X).
\end{aligned}
$$

3. Covariance

The covariance is also defined as follows:

$$Cov(X,Y) = \int_{-\infty}^{\infty} (x - E(X))(y - E(Y))f(xy)dxy.$$

At this point it will be useful to introduce a substitution for the integral and the density function.

$$\int_{-\infty}^{\infty} f(xy)dxy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y|x)f(x)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x|y)f(y)dydx.$$

Here we substitute the conditional and marginal densities for the joint density. When we do so, we must have an integral and a differential for each density. Also note that any time we see any of the above expressions with no other terms, it equals 1.0. In other words,

$$\int_{-\infty}^{\infty} f(y|x)dy = \int_{-\infty}^{\infty} f(x|y)dx = 1.$$

Now our covariance can be written as

$$\begin{aligned}
Cov(X,Y) &= \int_{-\infty}^{\infty} (x - E(X))(y - E(Y))f(xy)dxy \\
&= \int_{-\infty}^{\infty} (xy - xE(Y) - yE(X) - E(X)E(Y))f(xy)dxy \\
&= \int_{-\infty}^{\infty} xyf(xy)dxy - E(Y)\int_{-\infty}^{\infty} xf(xy)dxy \\
&\quad - E(X)\int_{-\infty}^{\infty} yf(xy)dxy + E(X)E(Y)\int_{-\infty}^{\infty} f(xy)dxy. \\
&= E(XY) - E(Y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(y|x)f(x)dxdy \\
&\quad - E(X)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x|y)f(y)dxdy + E(X)E(Y) \\
&= E(XY) - E(Y)\int_{-\infty}^{\infty} f(y|x)dy\int_{-\infty}^{\infty} xf(x)dx - E(X)\int_{-\infty}^{\infty} f(x|y)dx\int_{-\infty}^{\infty} yf(y)dy + E(X)E(Y) \\
&= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y)
\end{aligned}$$

So overall we have

$$Cov(X,Y) = E(XY) - E(X)E(Y).$$

With respect to constants, covariances are zero as indicated by the following:

$$
\begin{aligned}
Cov(c,X) &= \int_{-\infty}^{\infty}(c-E(c))(x-E(X))f(xy)dxy \\
&= \int_{-\infty}^{\infty}(c-c)(x-E(X))f(xy)dxy = 0.
\end{aligned}
$$

Below we show the important result that the variance of a weighted combination of random variables is a weighted combination of their variances and all possible pairwise covariances.

$$
\begin{aligned}
Var(aX+bY) &= \int_{-\infty}^{\infty}(ax+by-E(aX-bY))^2 f(xy)dxy \\
&= \int_{-\infty}^{\infty}(ax-E(aX)+by-E(bY))^2 f(xy)dxy \\
&= \int_{-\infty}^{\infty}[a(x-E(X))(b(y-E(Y)))]^2 f(xy)dxy \\
&= \int_{-\infty}^{\infty}[a^2(x-E(X))^2 + b^2(y-E(Y))^2 + 2ab(x-E(X))(y-E(Y))]f(xy)dxy \\
&= a^2\int_{-\infty}^{\infty}(x-E(X))^2 f(xy)dxy + b^2\int_{-\infty}^{\infty}(y-E(Y))^2 f(xy)dxy \\
&\quad + 2ab\int_{-\infty}^{\infty}(x-E(X))(y-E(Y))f(xy)dxy.
\end{aligned}
$$

Recall that earlier we substituted the product of the conditional and marginal densities for the joint density and split the single integral into the product of two integrals. Using these tricks, we obtain a^2Var(X) for the first term, b^2Var(Y) for the second and 2abCov(X,Y). Thus,

$$Cov(X,Y) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y).$$

The covariance between two weighted random variables is simply their covariance times the product of their weights:

$$
\begin{aligned}
Cov(aX,bY) &= abCov(X,Y)\\
&= \int_{-\infty}^{\infty}(ax-E(aX))(by-E(bY))f(xy)dxy\\
&= \int_{-\infty}^{\infty}(a(x-E(X))b(y-E(Y)))f(xy)dxy\\
&= ab\int_{-\infty}^{\infty}(x-E(X))(y-E(Y))f(xy)dxy\\
&= abCov(X,Y).
\end{aligned}
$$

The covariance between the sum of a constant and random variables times another sum of a constant and a random variable is simply the covariance of the random variables:

$$
\begin{aligned}
Cov(c+X,g+Y) &= \int_{-\infty}^{\infty}(c+x-E(c+X))(g+y-E(g+Y))f(xy)dxy\\
&= \int_{-\infty}^{\infty}(c-E(c)+x-E(X))(g-E(g)+y-E(Y))f(xy)dxy\\
&= \int_{-\infty}^{\infty}(c-c+x-E(X))(g-g+y-E(Y))f(xy)dxy\\
&= \int_{-\infty}^{\infty}(x-E(X))(y-E(Y))f(xy)dxy\\
&= Cov(X,Y).
\end{aligned}
$$

One of the more complex results is the covariance between the sum of two random variables and a third random variable:

$$
\begin{aligned}
Cov(X+Y,Z) &= \int_{-\infty}^{\infty}(x+y-E(X+Y))(z-E(Z))f(xyz)dxyz \\
&= \int_{-\infty}^{\infty}(x-E(X)+y-E(Y))(z-E(Z))f(xyz)dxyz \\
&= \int_{-\infty}^{\infty}(xz-xE(Z)-E(X)z+E(X)E(Z)+yz - yE(Z) \\
&\quad - E(Y)z + E(Y)E(Z))f(xyz)dxyz \\
&= \int_{-\infty}^{\infty}xzf(xyz)dxyz - E(Z)\int_{-\infty}^{\infty}xf(xyz)dxyz \\
&\quad - E(X)\int_{-\infty}^{\infty}zf(xyz)dxyz + E(X)E(Z)\int_{-\infty}^{\infty}f(xyz)dxyz \\
&\quad + \int_{-\infty}^{\infty}yzf(xyz)dxyz - E(Z)\int_{-\infty}^{\infty}yf(xyz)dxyz \\
&\quad - E(Y)\int_{-\infty}^{\infty}zf(xyz)dxyz + E(Y)E(Z)\int_{-\infty}^{\infty}f(xyz)dxyz.
\end{aligned}
$$

As we did previously when working with only two random variables, we need to convert the joint density into the product of the marginal and conditional densities. Now, however, we have three variables. We use the following relationships:

$$
\begin{aligned}
f(xyz) &= f(y|xz)f(xz) \\
&= f(x|yz)f(yz) \\
&= f(z|xy)f(xy) \\
&= f(yz|x)f(x) \\
&= f(xz|y)f(y) \\
&= f(xy|z)f(z).
\end{aligned}
$$

As we have previously noted, the integral of any density function over its entire domain is 1.0. Thus,

$$\int_{-\infty}^{\infty} f(yz|x)dyz = \int_{-\infty}^{\infty} f(xz|y)dxz = \int_{-\infty}^{\infty} f(xy|z)dxy = 1.0.$$

Of the eight expressions for the covariance, the first is, therefore,

$$\begin{aligned} + \int_{-\infty}^{\infty} xzf(xyz)dxyz &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xzf(y|xz)f(xz)dydxz \\ &= \int_{-\infty}^{\infty} f(y|xz)dy \int_{-\infty}^{\infty} xzf(xz)dxz \\ &= E(XZ). \end{aligned}$$

The second is

$$\begin{aligned} - E(Z)\int_{-\infty}^{\infty} xf(xyz)dxyz &= - E(Z)\int_{-\infty}^{\infty} f(yz|x)dyz \int_{-\infty}^{\infty} xf(x)dx \\ &= - E(Z)E(X). \end{aligned}$$

The third is

$$\begin{aligned} - E(X)\int_{-\infty}^{\infty} zf(xyz)dxyz &= - E(X)\int_{-\infty}^{\infty} f(xy|z)dxy \int_{-\infty}^{\infty} zf(z)dz \\ &= - E(X)E(Z). \end{aligned}$$

The fourth expression is clearly E(X)E(Z). The fifth expression is

$$\begin{aligned} &+ \int_{-\infty}^{\infty} yzf(xyz)dxyz \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yzf(x|yz)f(yz)dxdyz = \int_{-\infty}^{\infty} f(x|yz)dx \int_{-\infty}^{\infty} yzf(yz)dyz \\ &= E(YZ). \end{aligned}$$

The sixth expression is

$$
\begin{aligned}
& -E(Z)\int_{-\infty}^{\infty} yf(xyz)dxyz = -E(Z)\int_{-\infty}^{\infty} f(xz|y)dxz\int_{-\infty}^{\infty} yf(y)dy \\
& = -E(Z)E(Y).
\end{aligned}
$$

The seventh expression is

$$
\begin{aligned}
& -E(Y)\int_{-\infty}^{\infty} zf(xyz)dxyz = -E(Y)\int_{-\infty}^{\infty} f(xy|z)dxy\int_{-\infty}^{\infty} zf(z)dz \\
& = -E(Y)E(Z).
\end{aligned}
$$

The eight expression is clearly E(Y)E(Z). Thus, overall our covariance is

$$
\begin{aligned}
E(XZ) & - E(Z)E(X) - E(X)E(Z) + E(X)E(Z) + E(YZ) \\
& - E(Z)E(Y) - E(Y)E(Z) + E(Y)E(Z). \\
& = E(XZ) - E(X)E(Y) + E(YZ) - E(Y)E(Z) \\
& = Cov(X,Z) + Cov(Y,Z).
\end{aligned}
$$

Some General Results in Probability Theory

In the following sub-sections, we look at some of the general results from probability theory that occasionally are helpful. By the term “general,” we mean that these results are not dependent on any particular probability distribution. When a specific probability distribution is known, usually a stronger statement can be made.

1. The Central Limit Theorem

The *Central Limit Theorem* is a powerful statement that tells us that sum or average of independent samples drawn from any given probability distribution becomes normally distributed in the limit, that is, as the sample size approaches infinity. Thus, provided the sample size is large enough, the Central Limit Theorem allows us to use the rules associated with normal probability theory when drawing inferences about sample means. How large the sample size must be to rely on the Central Limit Theorem is not known, but a common rule of thumb has always been at least 30.

2. Chebyshev's Inequality

Chebyshev's Inequality, sometimes called *Chebyshev's Theorem*, allows us to make a sometimes useful, but still rather weak, statement about the probability of a sample value deviating from the population mean. More precisely, let X be a random variable with mean μ and variance σ^2, which can come from any probability distribution. For any real number t > 0,

$$\text{Prob}(*X - \mu* \ \$ \ t\sigma) \ \# \ t^2, \text{ assuming } t \ \$1.$$

Thus, Chebyshev's theorem gives an upper bound on how much the observed value deviates from the mean in terms of an arbitrary value t. For example, let X be the average height in inches of a randomly drawn male university student. Let μ = 70 and σ = 3. For t = various values, we have the following results:

t	tσ	Maximum Prob(*X - μ* $ tσ)
5	15	.04
4	12	.0625
3	9	.1111

In other words, the probability that a sample value will deviate from the mean by more than 15 inches is less than 0.04. The probability that it will deviate by more than 12 inches is at most 0.0625. The probability that it will deviate by more than 9 inches is at most .1111. These rules hold for any distribution, though knowing the exact distribution usually allows one to make more precise statements.[1]

3. The Law of Large Numbers

The *Law of Large Numbers* provides information about the accuracy with which a sample mean approximates a population mean. Basically it says that the probability that the difference between the sample mean and the population mean is greater than an arbitrarily chosen small value δ is zero as the sample size goes to infinity. This law holds as long as the sample consists of independently selected random variables from the same distribution. The Law of Large Numbers

[1]For example, if the distribution is normal, then we can say that the probability of being more than one standard deviation away from the mean, in either direction, is about .32.

more or less says that if we take sufficiently large samples, our sample mean estimate converges to the population mean.

4. The Law of Iterated Expectations (The Tower Law)

The *Law of Iterated Expectations*, sometimes called the *Tower Law*, is used to specify the expected value assessed at a given time t in terms of an expected value at another time t + i, which is taken as an expected value at a later time t + j. In other words, we are taking the expectation of an expectation. For example, suppose we are at time t + i and are calculating the expected value at time t + j. We might denote this as $E_{t+i}[X_{t+j}]$,which simply means that, using the information at time t+i, we assess the expected value of X to occur at a later time t+j. Now step back to time t and try to assess the expected value at time t+i. In other words, we want $E_t[E_{t+i}[X_{t+j}]]$. The Law of Iterated Expectations simply says

$$E_t[E_{t+i}[X_{t+j}]] = E_t[X_{t+j}].$$

In other words, the expectation is iterated from the later time to the earlier time.

5. The Law of Total Probability

The *Law of Total Probability* is a simple statement about the conditional probabilities and marginal probabilities. Specifically let Y and X be random variables where X is bifurcated into values greater than b or less than or equal to b. Then the Law of Total Probability is that

Prob(Y) = Prob(Y*X > b) + Prob(Y*X # b).

The statement Prob(Y) is simply any specification of Y, such as Prob(Y > c) or Prob(a # Y # c). So the probability of any event associated with Y can be broken down into the probability of that event for Y, conditional on one event occurring for X, and the probability of that event for Y, conditional on that event not occurring for X.