Random Variability: Covariance and Correlation

What of the variance of the sum of two random variables? If you work through the algebra, you'll find that

$$Var[X+Y] = Var[X] + Var[Y] + 2\cdot(E[XY] - E[X]\cdot E[Y]) .$$

This means that variances add when the random variables are independent, but not necessarily in other cases. The *covariance* of two random variables is $Cov[X,Y] = E[(X-E[X])\cdot(Y-E[Y])] = E[XY] - E[X]\cdot E[Y]$. We can restate the previous equation as

$$Var[X+Y] = Var[X] + Var[Y] + 2\cdot Cov[X,Y] .$$

Note that the covariance of a random variable with itself is just the variance of that random variable.

While variance is usually easier to work with when doing computations, it is somewhat difficult to interpret because it is expressed in squared units. For this reason, the *standard deviation* of a random variable is defined as the square-root of its variance. A practical (although not quite precise) interpretation is that the standard deviation of X indicates roughly how far from E[X] you'd expect the actual value of X to be.

Similarly, covariance is frequently "de-scaled," yielding the *correlation* between two random variables:

$$Corr(X,Y) = Cov[X,Y] / (StdDev(X) \cdot StdDev(Y)) .$$

The correlation between two random variables will always lie between -1 and 1, and is a measure of the strength of the linear relationship between the two variables.

Example: Let X be the percentage change in value of investment A in the course of one year (i.e., the annual *rate of return* on A), and let Y be the percentage change in value of investment B. Assume that you have $1 to invest, and you decide to put a dollars into investment A, and $1-a$ dollars into B. Then your return on investment from your portfolio will be $aX+(1-a)Y$, your expected return on investment will be $a\cdot E[X] + (1-a)\cdot E[Y]$, and the variance in your return on investment (a measure of the risk inherent in your portfolio) will be

$$a^2\cdot Var[X] + (1-a)^2\cdot Var[Y] + 2a(1-a)\cdot Cov[X,Y] .$$

For example, if you put all of your dollar into investment A, you'll have an expected return of $E[X]$, with a variance of $Var[X]$, while if you split your money between A and B, you'll have an expected return of $0.5 \cdot E[X] + 0.5 \cdot E[Y]$, with a variance of $0.25 \cdot Var[X] + 0.25 \cdot Var[Y] + 0.5 \cdot Cov[X,Y]$. Assume that both investments have equal expected returns and variances, i.e., $E[X] = E[Y]$ and $Var[X] = Var[Y]$. If X and Y are independent, then the expected return from the balanced portfolio is the same as the expected return from an investment in A alone. But the variance is only half as large! This observation lies at the heart of much of modern finance: Diversification can reduce risk. [Note that, if the covariance of X and Y is positive — if, for example, A and B are investments in similar industries — some of the advantage of diversification is lost. But if the covariance is negative, an even greater reduction in risk is achieved.]

More generally,

$$Var(X_1+X_2+X_3) = Var(X_1)+Var(X_2)+Var(X_3)+2\ Cov(X_1,X_2)+2\ Cov(X_1,X_3)+2\ Cov(X_2,X_3)\ ,$$

And even more generally, the variance of a sum is the sum of the individual variances, added to twice every pairwise covariance. This result is essential when determining the amount of risk inherent in an investment in any portfolio,