Random Variability

For any random variable X , the *variance* of X is the expected value of the squared difference between X and its expected value:

$$Var[X] = E[(X-E[X])^2] = E[X^2] - (E[X])^2 .$$

(The second equation is the result of a bit of algebra: $E[(X-E[X])^2] = E[X^2 - 2\cdot X\cdot E[X] +(E[X])^2] = E[X^2] - 2\cdot E[X]\cdot E[X] + (E[X])^2$.) Variance comes in squared units (and adding a constant to a random variable, while shifting its values, doesn't affect its variance), so

$$Var[kX+c] = k^2\cdot Var[X] .$$

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.]

Example: Consider the randomly-varying demand for a product over a fixed number LT (short for "leadtime") of days. Day-to-day demand varies independently, with each day's demand having the same probability distribution. Total demand is $D_1 + ... + D_{LT}$. Then the expected total demand is $LT \cdot E[D]$, and the variance is $LT \cdot Var[D]$, where D represents any single day's demand.

Now, assume that the daily demand D is constant, but the length of the leadtime is uncertain. Then the total demand is $D \cdot LT$, with expected value $D \cdot E[LT]$ and variance $D^2 \cdot Var[LT]$.

Finally, combine these two cases, and consider the total demand when both day-to-day demand and the length of the leadtime are random variables (so the total is a sum of a random number of random variables). As long as the length of the leadtime is independent of the daily demands, the expected total demand will be $E[D] \cdot E[LT]$, and the variance will be $E[LT] \cdot Var[D] + (E[D])^2 \cdot Var[LT]$. This result is useful in analyzing buffer inventories (safety stocks).

Summarizing the important special case which arises when the leadtime is constant, i.e., LT = n ,

$$E[D_1 + ... + D_n] = n \cdot E[D] ,$$
$$Var(D_1 + ... + D_n) = n \cdot Var(D) , \text{ and}$$
$$StdDev(D_1 + ... + D_n) = \sqrt{n} \cdot StdDev(D) .$$

Similarly, for averages,

$$E[(D_1 + ... + D_n) / n] = E[D] ,$$
$$Var((D_1 + ... + D_n) / n) = Var(D) / n , \text{ and}$$
$$StdDev((D_1 + ... + D_n) / n) = StdDev(D) / \sqrt{n} .$$