Random Variability: Variance and Standard Deviation

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

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

The units in which variance is measured can be hard to interpret. More frequently, for purposes of discussion we look at the standard deviation of X:

$$StDev(X) = \sqrt{Var(X)} .$$

A not-quite-correct, but usually-good-enough interpretation of the standard deviation of a random variable is that it indicates how far, on average, the actual observed value of X differs from (i.e., is above or below) the expected value of X.

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

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

When X and Y are independent, the last of the three terms is 0, so the variance of the sum is simply the sum of the variances. (This is not true in general, and we'll be discussing the general case later in the course.)

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

$$Var(D_1 +...+ D_{LT}) = E[LT]\cdot Var[D] + (E[D])^2\cdot Var[LT] .$$

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,

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

These “square root” effects come up in topic areas as diverse as option pricing, warehousing, and statistical sampling.