Covariance

When discussing a single RV, we used the notion of variance to capture how much that RV could differ from its expected value. With two RVs we have a similar notion called the *covariance*, defined as

$$\operatorname{cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y].$$

Note that $\operatorname{cov}(X, X) = \operatorname{var}(X)$. It is straightforward to show that if two RVs are independent, then $\operatorname{cov}(X, Y) = 0$.

The covariance gives a way to quantify how two variables move together when they are not independent. When $\operatorname{cov}(X, X) > 0$, then X and Y tend to have the same sign (i.e., one being positive indicates that the other is likely positive). In contrast, when $\operatorname{cov}(X, X) < 0$, then X and Y tend to have the opposite sign (i.e., one being positive indicates that the other is likely negative).

To show that the covariance actually quantifies something important, remember our rule that $\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y)$ *only when X and Y are independent*? What about when the RVs are dependent? In general, we can say that:

$$\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X, X).$$

Correlation coefficient

The covariance discussed above seems like a very helpful way to quantify the dependence between RVs. One glaring problem is that it's not normalized. While the covariance can be bounded using the individual variances, any bound derived in this way would depends on the distribution of the variables (or even the scaling of the variables). So, it's hard to compare from one situation to the next.

To alleviate this, we can introduce the *correlation coefficient*:

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

which can be shown to be $-1 \leq \rho(X,Y) \leq 1$. When $\rho(X,Y) = 0$ we call the RVs uncorrelated.

The correlation coefficient gives us a way to quantify dependence that is on a normalized scale and can be compared from one situation to the next. The sign of $\rho(X,Y)$ tells you whether the RVs tend to go in the same or opposite directions. The size of $\rho(X,Y)$ tells you the extent to which this tendency is true. A value of $\rho(X,Y) = 1$ means that one RV can be exactly determined from the other (i.e, there is no randomness once you know one of the RVs).

We call X and Y **uncorrelated** if

$$\text{E}[XY] = \text{E}[X] \cdot \text{E}[Y].$$

It is easy to see that uncorrelated RVs have $\rho(X,Y) = 0$.

Independence and correlation/covariance

It is straightforward to show that:

$$X, Y \text{ independent} \quad \Rightarrow \quad X, Y \text{ uncorrelated and } \text{cov}(X,X) = 0$$

$$X, Y \text{ uncorrelated or } \text{cov}(X,X) = 0 \quad \not\Rightarrow \quad X, Y \text{ independent.}$$

So independent random variables are uncorrelated (and have covariance of zero), but uncorrelated random variables are not necessarily independent.

While the above statement that uncorrelated RVs are not independent is true in general, there is one (very important) special case. When X, Y are Gaussian RVs, then $\rho(X, Y) = 0$ actually does imply that the RVs are independent!

Many people find the above general statement hard to believe and want an example where two RVs are uncorrelated but not independent. The simplest example is from discrete RVs:

Example:
The pair of RVs (X, Y) take the values (1,0), (0,1), (-1,0) and (0,-1) with equal probability (1/4). Because of the symmetric distributions, we can easily show that $\mathrm{E}[X] = \mathrm{E}[Y] = 0$. For all possible pairs, one of the RVs is zero which implies that $\mathrm{E}[XY] = 0$. Therefore, $\mathrm{cov}(X, Y) = \rho(X, Y) = 0$ and the RVs are uncorrelated.

BUT, these RVs also cannot be independent. Consider that knowing X takes a nonzero value tells you exactly that Y has to be zero.