CLARK UNIVERSITY 1887

Covariance and Correlation
Math 217 Probability and Statistics
Prof. D. Joyce, Fall 2014

Covariance. Let X and Y be joint random variables. Their *covariance* $\mathrm{Cov}(X,Y)$ is defined by

$$\mathrm{Cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y)).$$

Notice that the variance of X is just the covariance of X with itself

$$\mathrm{Var}(X) = E((X-\mu_X)^2) = \mathrm{Cov}(X,X)$$

Analogous to the identity for variance

$$\mathrm{Var}(X) = E(X^2) - \mu_X^2$$

there is an identity for covariance

$$\mathrm{Cov}(X) = E(XY) - \mu_X\mu_Y$$

Here's the proof:

$$\begin{aligned}
&\mathrm{Cov}(X,Y) \\
= \; & E((X-\mu_X)(Y-\mu_Y)) \\
= \; & E(XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y) \\
= \; & E(XY) - \mu_X E(Y) - E(X)\mu_Y + \mu_X\mu_Y \\
= \; & E(XY) - \mu_X\mu_Y
\end{aligned}$$

Covariance can be positive, zero, or negative. Positive indicates that there's an overall tendency that when one variable increases, so doe the other, while negative indicates an overall tendency that when one increases the other decreases.

If X and Y are independent variables, then their covariance is 0:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E(XY) - \mu_X\mu_Y \\
&= E(X)E(Y) - \mu_X\mu_Y = 0
\end{aligned}$$

The converse, however, is *not* always true. $\mathrm{Cov}(X,Y)$ can be 0 for variables that are not independent.

For an example where the covariance is 0 but X and Y aren't independent, let there be three outcomes, $(-1,1)$, $(0,-2)$, and $(1,1)$, all with the same probability $\frac{1}{3}$. They're clearly not independent since the value of X determines the value of Y. Note that $\mu_X = 0$ and $\mu_Y = 0$, so

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) \\
&= E(XY) \\
&= \tfrac{1}{3}(-1) + \tfrac{1}{3}0 + \tfrac{1}{3}1 = 0
\end{aligned}$$

We've already seen that when X and Y are independent, the variance of their sum is the sum of their variances. There's a general formula to deal with their sum when they aren't independent. A covariance term appears in that formula.

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$

Here's the proof

$$\begin{aligned}
&\mathrm{Var}(X+Y) \\
= \; & E((X+Y)^2) - E(X+Y)^2 \\
= \; & E(X^2 + 2XY + Y^2) - (\mu_X + \mu_Y)^2 \\
= \; & E(X^2) + 2E(XY) + E(Y^2) \\
& - \mu_X^2 - 2\mu_X\mu_Y - \mu_Y^2 \\
= \; & E(X^2) - \mu_X^2 + 2(E(XY) - \mu_X\mu_Y) \\
& + E(Y^2) - \mu_Y^2 \\
= \; & \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)
\end{aligned}$$

Bilinearity of covariance. Covariance is linear in each coordinate. That means two things. First, you can pass constants through either coordinate:

$$\mathrm{Cov}(aX,Y) = a\,\mathrm{Cov}(X,Y) = \mathrm{Cov}(X,aY).$$

Second, it preserves sums in each coordinate:

$$\mathrm{Cov}(X_1+X_2,Y) = \mathrm{Cov}(X_1,Y) + \mathrm{Cov}(X_2,Y)$$

and

$$\mathrm{Cov}(X,Y_1+Y_2) = \mathrm{Cov}(X,Y_1) + \mathrm{Cov}(X,Y_2).$$

Here's a proof of the first equation in the first condition:

$$\begin{aligned}\operatorname{Cov}(aX,Y) &= E((aX-E(aX))(Y-E(Y)))\\ &= E(a(X-E(X))(Y-E(Y)))\\ &= aE((X-E(X))(Y-E(Y)))\\ &= a\operatorname{Cov}(X,Y)\end{aligned}$$

The proof of the second condition is also straightforward.

Correlation. The *correlation* ρ_{XY} of two joint variables X and Y is a normalized version of their covariance. It's defined by the equation

$$\rho_{XY} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y}.$$

Note that independent variables have 0 correlation as well as 0 covariance.

By dividing by the product $\sigma_X\sigma_Y$ of the standard deviations, the correlation becomes bounded between plus and minus 1.

$$-1 \le \rho_{XY} \le 1.$$

There are various ways you can prove that inequality. Here's one. We'll start by proving

$$0 \le \operatorname{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right) = 2(1 \pm \rho_{XY}).$$

There are actually two equations there, and we can prove them at the same time.

First note the "$0 \le$" parts follow from the fact variance is nonnegative. Next use the property proved above about the variance of a sum.

$$\begin{aligned}&\operatorname{Var}\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right)\\ =\ &\operatorname{Var}\left(\frac{X}{\sigma_X}\right) + \operatorname{Var}\left(\frac{\pm Y}{\sigma_Y}\right) + 2\operatorname{Cov}\left(\frac{X}{\sigma_X}, \frac{\pm Y}{\sigma_Y}\right)\end{aligned}$$

Now use the fact that $\operatorname{Var}(cX) = c^2\operatorname{Var}(X)$ to rewrite that as

$$\frac{1}{\sigma_X^2}\operatorname{Var}(X) + \frac{1}{\sigma_Y^2}\operatorname{Var}(\pm Y) + 2\operatorname{Cov}\left(\frac{X}{\sigma_X}, \pm\frac{Y}{\sigma_Y}\right)$$

But $\operatorname{Var}(X) = \sigma_X^2$ and $\operatorname{Var}(-Y) = \operatorname{Var}(Y) = \sigma_Y^2$, so that equals

$$2 + 2\operatorname{Cov}\left(\frac{X}{\sigma_X}, \frac{\pm Y}{\sigma_Y}\right)$$

By the bilinearity of covariance, that equals

$$2 \pm \frac{2}{\sigma_x\sigma_Y}\operatorname{Cov}(X,Y) = 2 \pm 2\rho_{XY})$$

and we've shown that

$$0 \le 2(1 \pm \rho_{XY}.$$

Next, divide by 2 move one term to the other side of the inequality to get

$$\mp\rho_{XY} \le 1,$$

so

$$-1 \le \rho_{XY} \le 1.$$

This exercise should remind you of the same kind of thing that goes on in linear algebra. In fact, it is the same thing exactly. Take a set of real-valued random variables, not necessarily independent. Their linear combinations form a vector space. Their covariance is the inner product (also called the dot product or scalar product) of two vectors in that space.

$$X \cdot Y = \operatorname{Cov}(X,Y)$$

The norm $\|X\|$ of X is the square root of $\|X\|^2$ defined by

$$\|X\|^2 = X \cdot X = \operatorname{Cov}(X,X) = V(X) = \sigma_X^2$$

and, so, the angle θ between X and Y is defined by

$$\cos\theta = \frac{X \cdot Y}{\|X\|\,\|Y\|} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X\sigma_Y} = \rho_{XY}$$

that is, θ is the arccosine of the correlation ρ_{XY}.

Math 217 Home Page at
`http://math.clarku.edu/~djoyce/ma217/`