Lecture 40: more on the sign of covariance. By definition:

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

(\*)

Suppose, for simplicity, that \(E(X) = E(Y) = 0\).

Then (\*) becomes simply

$$\text{Cov}(X,Y) = E(XY).$$

(\**)

What is the meaning of the sign of \(E(XY)\)?

Remark: given two numbers \(x\) and \(y\), we have:

- \(xy > 0 \iff x\) and \(y\) have the same sign.
- \(xy < 0 \iff x\) and \(y\) have opposite sign.

So, if \(\text{Cov}(X,Y) = E(XY) > 0\), then the r.v.'s of \(X\) and \(Y\) have, on average, the same sign; that is, on the \((x,y)\) plane on which the joint pdf \(f(x,y)\) is defined, with higher probability \(X\) and \(Y\) will be either in the 1\(^{\text{st}}\) quadrant or in the 3\(^{\text{rd}}\) quadrant (or both).

E.g.: if \(X\) and \(Y\) are jointly uniform on the set \(D\) on the right then \(E(XY) > 0\), because \(XY\) is most likely positive.
On the other hand, if \( \text{Cov}(X,Y) = E(XY) < 0 \),
then the product \( XY \) is more likely to be negative;
so with higher probability the pair \((X,Y)\) is likely
to fall in the 3rd or 4th quadrants:

So if we know that \( X \) is positive, then \( Y \) is more
likely to be negative.

- Finally, if \( \text{Cov}(X,Y) = E(XY) = 0 \), then on average
\( X \) and \( Y \) have the same sign. This is the case, for
example, of \( X \sim \text{Normal}(0, \sigma_x^2) \), \( Y \sim \text{Normal}(0, \sigma_y^2) \), independent.

Let's go back to the general case \((*)\), with \( \mu_X \) & \( \mu_Y \) not
necessarily zero.

**Lemma:** the covariance \((*)\) may be written as:

\[
\text{Cov}(X,Y) = E\left[ (X-\mu_X)(Y-\mu_Y) \right]
\]

where \( \mu_X = E(X) \) and \( \mu_Y = E(Y) \).

**Proof:** by direct computation.

So, if \( \text{Cov}(X,Y) > 0 \) then, on average,
\( X-\mu_X \) and \( Y-\mu_Y \) have the same sign!

In other words,
- if \( X-\mu_X > 0 \), i.e. \( X > \mu_X \), then it is
  likely that \( Y-\mu_Y > 0 \), i.e. \( Y > \mu_Y \).
- on the other hand, if \( X < \mu_X \), then
  it is likely that \( Y < \mu_Y \).
Graphically, on the \((x,y)\) plane on which the joint density \(f(x,y)\) is defined, we have, with higher probability, that: \(X > \mu_x\) and \(Y > \mu_y\), or \(X < \mu_x\) and \(Y < \mu_y\), since in both cases \((x-\mu_x)(y-\mu_y) > 0\).

Typically:

\[
\begin{array}{c|c}
& x \\
\hline
y & \\
\end{array}
\]

\(\mu_y\)

\(\mu_x\)

\(x\)

Note: either of \(\mu_x\) or \(\mu_y\), or both, could be negative.

So: "above average values of \(X\) are associated with above average values of \(Y\)"; and vice-versa.

For example, if \((X,Y)\) is uniformly distributed on the set \(D\) illustrated on the right, then \(\text{Cov}(X,Y) > 0\).

Examples of random variables with positive covariance:

- \(X = \) height of a person
  \(Y = \) height of a person's father

- \(X = \) a person's income
  \(Y = \) \# of years of education completed by a person

- \(X = \) \# of hours spent studying
  \(Y = \) GPA

- \(X = \) height of a person
  \(Y = \) weight of a person
Similarly, if $\text{Cov}(X,Y) < 0$, then, on average, $X - \mu_X$ and $Y - \mu_Y$ have opposite signs!

In other words,
- if $X > \mu_X$, then it is likely that $Y < \mu_Y$;
- if $X < \mu_X$, then it is likely that $Y > \mu_Y$.

So, above average values of $X$ are associated with below average values of $Y$.

Typically, the regions of the $(x,y)$ plane of highest probability are the following:

Examples of random variables with negative covariance:
- $X =$ median household income in a neighborhood
  $Y =$ crime rate in a neighborhood
- $X =$ median temperature for a given winter
  $Y =$ money spent on heating per household
- $X =$ # of hours spent playing video games
  $Y =$ GPA

$\text{Cov}(X,Y)$ is zero when positive values and negative values of $(X - \mu_X)(Y - \mu_Y)$ balance each other, and its expectation is zero.
So much for the sign of covariance. What about its amplitude?

Def.: Let \( X, Y \) be two random variables. The CORRELATION of \( X \) and \( Y \) is:

\[
\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{SD}(X) \text{SD}(Y)}
\]

Theorem: \[ -1 \leq \text{Corr}(X, Y) \leq 1 \]

Proof: Let \( \mu_X = E(X), \mu_Y = E(Y) \)
\( \sigma_X = \text{SD}(X), \sigma_Y = \text{SD}(Y) \)

Remember: \( X^* = \frac{X - \mu_X}{\sigma_X} \) is "\( X \) in standard units", and it has the properties: \( E(X^*) = \frac{1}{\sigma_X} (E(X) - \mu_X) = 0 \), while \( \text{SD}(X^*) = \frac{1}{\sigma_X} \text{SD}(X) = 1 \). Similarly, \( Y^* = \frac{Y - \mu_Y}{\sigma_Y} \).

We have:

\[
E[X^*Y^*] = E\left[ \frac{X - \mu_X}{\sigma_X} \cdot \frac{Y - \mu_Y}{\sigma_Y} \right] = E\left[ \frac{(X - \mu_X)(Y - \mu_Y)}{\sigma_X \sigma_Y} \right] = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}
\]

\[ \implies E[X^*Y^*] = \text{Corr}(X, Y), \quad (****) \]

But: for any two random variables \( Z \) and \( W \) we have:

\[
\text{Var}(Z + W) = \text{Var}(Z) + \text{Var}(W) + 2 \text{Cov}(Z, W); \quad (***)
\]

if \( E(Z) = 0 \) & \( E(Y) = 0 \), \( \text{Var}(Z) = E(Z^2) \)
\( \text{Var}(W) = E(W^2) \)
\( \text{Var}(Z + W) = E[(Z + W)^2] \)
\( \text{Cov}(Z, W) = E(ZW) \)

so \((***)\) becomes:

\[
E[(Z + W)^2] = E(Z^2) + E(W^2) + 2 E(ZW),
\]
Take: \( Z = X^* \) and \( W = Y^* \). (***) becomes:

\[
0 \leq E[(X^* + Y^*)^2] = E[(X^*)^2] + E[(Y^*)^2] + 2E(X^*Y^*) \geq 1 + 1 + 2E(X^*Y^*)
\]

so \( E(X^*Y^*) \geq -1 \)

Now take: \( Z = X^* \) and \( W = -Y^* \). (***) becomes:

\[
0 \leq E[(X^* - Y^*)^2] = E[(X^*)^2] + E[(Y^*)^2] - 2E(X^*Y^*) \geq 1 + 1 - 2E(X^*Y^*)
\]

so \( E(X^*Y^*) \leq 1 \). In conclusion:

\[-1 \leq E(X^*Y^*) \leq 1\]

therefore by (***) , \(-1 \leq \text{Corr}(X,Y) \leq 1\).

**Corollary:** if \( X^* = \frac{X - \mu_X}{\sigma_X} \), where \( \mu_X = E(X) \) and \( \sigma_X = SD(X) \),

and \( Y^* = \frac{Y - \mu_Y}{\sigma_Y} \), where \( \mu_Y = E(Y) \) and \( \sigma_Y = SD(Y) \),

then \( \text{Corr}(X,Y) = E(X^*Y^*) \).

(it's a byproduct of the above proof).

**Corollary:** we have \( |\text{Corr}(X,Y)| \leq 1 \), therefore

\[ |\text{Cov}(X,Y)| \leq SD(X)SD(Y) \].