Covariance of Random Variables

April 1, 2020

1 Covariance

Note: This material will not be on Midterm 2, but will be on the Final. It is an application of Section 5.5 and 5.6, so it does provide additional problems related to that.

The covariance between 2 random variables gives a measure of how related they are. Let's begin with the definition, then we'll consider some special cases to see why this definition tells us how random variables are related.

DEFINITION 1.1. *Let X and Y be random variables, with $\mathbb{E}[X]=\mu_X$ and $\mathbb{E}[Y]=\mu_Y$. The* ***covariance*** *of X and Y is*

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

Let's consider some special cases

EXAMPLE 1.2. *Let X and Y be* ***independent*** *random variables. Compute* $\mathrm{Cov}(X,Y)$.

Solution:

$$\mathrm{Cov}(X,Y)=\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]=\mathbb{E}[X-\mu_X]\mathbb{E}[Y-\mu_Y]=0.$$

In the third equality we used our theorem from the last section that $\mathbb{E}[g(X)h(Y)]=\mathbb{E}[g(X)]\mathbb{E}[h(Y)]$ when the r.v.'s are independent, applying to the functions $g(X)=X-\mu_X$, $h(Y)=Y-\mu_Y$. The final equality used that

$$\mathbb{E}[X-\mu_X]=\mathbb{E}[X]-\mathbb{E}[\mu_X]=\mu_X-\mu_X=0.$$

So the covariance of independent random variables is always 0.

EXAMPLE 1.3. *Let X be a random variable. Compute* $\mathrm{Cov}(X,X)$.

Solution:

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

So the covariance of a random variable with itself is just the variance.

These are the two extreme examples: if the random variables are not related at all then their covariance is 0 and if they are equal to each other then their covariance is their variance.

Intuitively, you can think of having a small covariance as begin close to independent, but this is not the whole picture, as the next few examples will show. You would also like to of think of a large

Covariance as meaning the random variables are strongly related, this is also not quite right, but after a few examples, we will introduce a better measure of how related they are, called the correlation coefficient.

Example 1.4. *Let Y_1 and Y_2 be discrete random variables with joint pmf:*

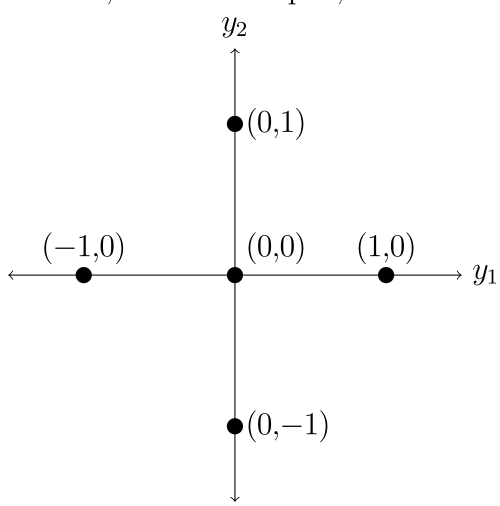
$$p_{Y_1,Y_2}(-1,0)=1/5 \; p_{Y_1,Y_2}(1,0)=1/5 \; p_{Y_1,Y_2}(0,1)=1/5 \; p_{Y_1,Y_2}(0,-1)=1/5 \; p_{Y_1,Y_2}(0,0)=1/5$$

and $p_{Y_1,Y_2}(y_1,y_2)=0$ otherwise.

(a) *What is* $\mathrm{Cov}[Y_1,Y_2]$*?*

(b) *Show that Y_1,Y_2 are not independent.*

Solution: This is similar to an example from the last lecture, but a bit simpler, each of the 5 points



is equally likely to occur. Once again the picture is:

(a) We begin by computing $\mathbb{E}[Y_1]$,

$$\mathbb{E}[Y_1]=-1*1/5+0*3/5+1*1/5=0$$

where the first term is the value Y_1 can take and the second term is the probability that happens. $\mathbb{E}[Y_2]=0$, by the same computation. So

$$\mathrm{Cov}(Y_1,Y_2)=\mathbb{E}[(Y_1-0)(Y_2-0)]=\mathbb{E}[Y_1Y_2]=\sum_{y_1=-1}^{1}\sum_{y_2=-1}^{1} y_1y_2p_{Y_1,Y_2}(y_1,y_2)=0$$

The last equality just uses that if $p_{Y_1,Y_2}(y_1,y_2)>0$ then at least one of the y_1,y_2 must be 0. So

$$\mathrm{Cov}(Y_1,Y_2)=0$$

(b) We should suspect that the random variables are not independent, because is $Y_1=1$ then Y_2 must be 0. Which seems like a strong constraint. In order to show, 2 r.v.'s are not independent we just need to show there exists a point (y_1,y_2) where $p_{Y_1,Y_2}(y_1,y_2)\neq p_{Y_1}(y_1)p_{Y_2}(y_2)$. To do this, one proceeds by guess and check, if you guess and check a few points and equalities then you should suspect the r.v.'s are in fact independent and compute their marginal distributions.

Let's try $y_1=1,y_2=1$, $p_{Y_1,Y_2}(1,1)=0$, but $p_{Y_1}(1)=1/5$ and $p_{Y_2}(1)=1/5$, since $0\neq 1/5^2$ we conclude the random variables are not independent.

For a more complicated example of the same phenomenon as the last example, you can check that the r.v.'s in Example 1.8 from the March 27 (Section 5.4) lecture notes are also uncorrelated.

As earlier mentioned a larger covariance seems to imply the random variables are strongly related, the only problem with his intuition is that it is possible one of the random variables has a very large variance that is causing the large covariance. This issue is handled by the following definition.

DEFINITION 1.5. *Let X and Y be random variables, the **correlation coefficient** of X and Y is*

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X\,\mathrm{Var}Y}}$$

The notation ρ is pretty standard, if it is ambiguous which random variables are being compared you can write ρ_{XY}.

The correlations coefficient is always between -1 and 1 (inclusively). So "large" correlations coefficients are near 1, and this happens is the covariance is close to the variance of both the random variables.

The proof of the fact that $\rho \in [-1,1]$ follows from general facts from in linear algebra. I'll mention them here, in the case you've taken an advanced linear algebra class, but if not this will not be so important.

THEOREM 1.6.

$$\rho \in [-1,1]$$

Idea of proof: The operation $\mathrm{Cov}(X,Y)$ is an inner product (which is the generalization of the dot product to arbitrary vector spaces). Therefore it satisfies the Cauchy-Schwarz inequality:

$$|\mathrm{Cov}(X,Y)|^2 \leq \mathrm{Cov}(X,X)\mathrm{Cov}(Y,Y)$$

since $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$, $\mathrm{Cov}(Y,Y) = \mathrm{Var}(Y)$, taking the square root of both sides and rearranging gives the desired result. □

2 Alternative way to compute Covariance

Just as with the variance, we started with the definition but then expanded it to get an often easier object to compute.

THEOREM 2.1. *Let X and Y be random variables, with $\mathbb{E}[X] = \mu_X$ and $\mathbb{E}[Y] = \mu_Y$, then*

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

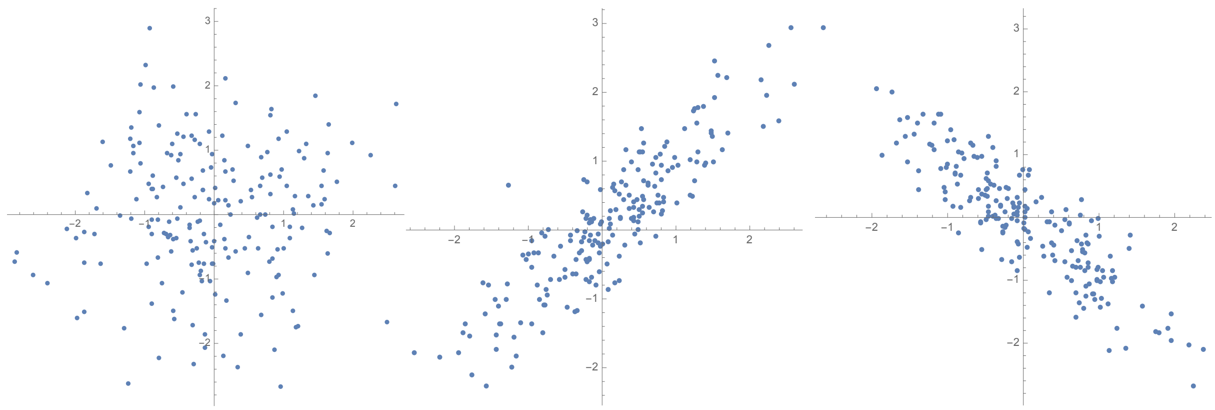
Proof.

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mu_X E[Y] - \mathbb{E}[X]\mu_Y + \mu_X\mu_Y = \mathbb{E}[XY] - \mu_X\mu_Y$$

where we have used that we can break up the expectation of sums and pull constants out of expectations. □

3 Meaning of the correlation coefficient

I will now plot the support of the joint pmf of several paris of random variables, to give a picture of what various correlation coefficients look like. I will center the pictures around the origin, but this is



(a) $\rho \approx 0$ (b) $\rho \approx 1$ (c) $\rho \approx -1$

not important, all the points could be shifted anywhere on the plane to give the same correlation, they just need to all be moved in the same way.

In each of the figures example I choose the listed ρ value and generated 200 pairs of variables. Each pair is represented by a dot, its x coordinate is one of the random variables and its y coordinate is the other.

Notice in the first picture, if you're told the x coordinate of a point is large and positive, you would not be able to guess if y value is positive or negative. On the other hand in the second picture, you would be fairly confident that the y value is positive and in the third picture that it is negative.

But it is important to remember not all the information about how random variables are related is included in ρ. For example, a point is chosen uniformly at random from the disk of radius 1, then its x and y coordinates will be uncorrelated but not independent. If I told you the x coordinate of such a randomly chosen point were large, then you would know it's y coordinate is near zero, but you wouldn't know if it was positive or negative. Since the y coordinate has expectation 0, being uncorrelated means you can't guess on which side of its expectation the y coordinate lands.

4 One more example

Let's do one more ugly example. We'll use a joint pdf we've looked at a little bit already.

EXAMPLE 4.1. *Let Y_1 and Y_2 be continuous random variables with jpdf:*

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \frac{3}{4}y_1^2, & \text{if } 0 \le y_1 \le y_2 \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

Compute $Cov(Y_1,Y_2)$.

Solution: Using the alternative formula above, we need to compute $\mathbb{E}[Y_1Y_2]$, $\mathbb{E}[Y_1]$, and $\mathbb{E}[Y_2]$.

We have previously computed the marginal distributions, so let's use them (if we hadn't computed the marginal distributions, we should probably just directly compute $\mathbb{E}[Y_1]$ from the jpdf like in the last set of notes.) Recall:

$$f_{Y_1}(y_1) = \begin{cases} \frac{3}{4}y_1^2(2-y_1), & \text{if } 0 \le y_1 \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

$$f_{Y_2}(y_2) = \begin{cases} \frac{1}{4}y_2^3, & \text{if } 0 \le y_2 \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

So

$$\mathbb{E}[Y_1] = \int y_1 f_{Y_1}(y_1)\ dy_1 = \int_0^2 y_1 \frac{3}{4} y_1^2 (2 - y_1)\ dy_1 = \frac{3}{4}\int_0^2 (2y_1^3 - y_1^4)\ dy_1 = \frac{3}{4}(8 - \frac{32}{5}) = 6/5$$

$$\mathbb{E}[Y_2] = \int y_2 f_{Y_2}(y_2)\ dy_2 = \int_0^2 y_2 \frac{1}{4} y_2^3\ dy_2 = \frac{1}{4}\int_0^2 y_2^4\ dy_2 = \frac{1}{4}\frac{32}{5} = 8/5$$

To compute $\mathbb{E}[Y_1 Y_2]$ we need to use the joint distribution

$$\mathbb{E}[Y_1 Y_2] = \int_0^2 \int_{y_1}^2 y_1 y_2 \frac{3}{4} y_1^2\ dy_2 dy_1 = \frac{3}{4}\int_0^2 y_1^3 \left(\frac{y_2^2}{2}\right)_{y_1}^2\ dy_1 = \frac{3}{4}\int_0^2 y_1^3 \left(\frac{4}{2} - \frac{y_1^2}{2}\right)\ dy_1 = \frac{3}{4}\left(2\frac{y_1^4}{4} - \frac{y_1^6}{12}\right)_0^2$$

So

$$\mathbb{E}[Y_1 Y_2] = \frac{3}{4}\left(2\frac{2^4}{4} - \frac{2^6}{12}\right) = 2$$

So

$$\mathrm{Cov}(Y_1, Y_2) = 2 - \frac{6}{5}\frac{8}{5} = 2/25$$