

5.2: Expected Values, Covariance and Correlation

Taylor

University of Virginia

Proposition

Let X and Y be jointly distributed random variables with pmf $p(x, y)$ or pdf $f(x, y)$ (according to whether the rvs are discrete or continuous). Then

$$E[h(X, Y)] = \sum_x \sum_y h(x, y)p(x, y)$$

if they're jointly discrete, or

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)f(x, y)dx dy$$

if they are continuous.

Note: even though $h(X, Y)$ is a random variable itself, and it has a new density, we don't have to use that to find it's expected value. (LOTUS)

Proposition

Let X_1, X_2, \dots, X_n be independent random variables and assume that all the expected values we write down exist. Then

$$E[h_1(X_1)h_2(X_2)\cdots h_n(X_n)] = E[h_1(X_1)] \times \cdots \times E[h_n(X_n)]$$

Definition

The **covariance** between two rvs X and Y is

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)]$$

if these are jointly cts, then

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dx dy$$

if discrete then

$$\text{Cov}(X, Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) p(x, y)$$

If $\text{Cov}(X, Y) > 0$ that means $(X - \mu_X)$ and $(Y - \mu_Y)$ tend to be the same sign (both negative, or both positive).

If $\text{Cov}(X, Y) < 0$ that means $(X - \mu_X)$ and $(Y - \mu_Y)$ tend to be opposite signs (ones positive and the other is negative, and vice versa)

Remember that this is a probability weighted average. If a covariance is positive, for example, that tells you they have the same sign on *average*.

Proposition

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

Here's a proof:

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - EX)(Y - EY)] \\ &= E[XY - X(EY) - (EX)Y + (EX)(EY)] \\ &= E[XY] - E[X(EY)] - E[(EX)Y] + E[(EX)(EY)] \\ &= E[XY] - E[X](EY) - (EX)E[Y] + (EX)(EY) \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

Covariance formula

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

It is sometimes easier to use this formula! For example, if you have the first two derivatives of a moment generating function.

Proposition

The following is called *bilinearity*. If a and b are constants, X , Y and Z are three rvs...

$$\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$$

$$\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

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

Proposition

What's $\text{Cov}(X, X)$?

What's $\text{Cov}(X, X)$?

$$\begin{aligned}\text{Cov}(X, X) &= E[(X - EX)(X - EX)] \\ &= E[(X - EX)^2] \\ &= \text{Var}[X]\end{aligned}$$

Motivation

"It would appear that the relationship in the insurance example is quite strong since $\text{Cov}(X, Y) = 1875$, whereas in the nut example $\text{Cov}(X, Y) = -\frac{2}{75}$ would seem to imply quite a weak relationship."

This isn't true, though. You can't compare covariances. The number depends on the scale. For instance, if you compared the correlation between prices of one item with another item, it totally depends on whether you're talking about dollars or cents.

You can see this formally just using bilinearity. For any X , Y , and $a, b \neq 1$, $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y) \neq \text{Cov}(X, Y)$.

Definition

That's why we talk about **correlation**

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

It might help to think about it like this:

$$\text{Corr}(X, Y) = \text{Cov} \left[\frac{X}{\sqrt{\text{Var}(X)}}, \frac{Y}{\sqrt{\text{Var}(Y)}} \right]$$

where $\text{Var} \left[\frac{X}{\sqrt{\text{Var}(X)}} \right] = \frac{1}{\text{Var}(X)} \text{Var}(X) = 1$

Correlation

A few more things (prove these on your own)

- ① if a and c have the same sign, $\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$
- ② $-1 \leq \text{Corr}(X, Y) \leq 1$
- ③ if X and Y are independent, then $\text{Corr}(X, Y) = 0$
- ④ if $\text{Corr}(X, Y) = 0$, X and Y can still be dependent
- ⑤ $|\text{Corr}| = 1$ if and only if $Y = aX + b$ for some scalars a, b with $a \neq 0$

Note: to prove (2) and one half of (5), you need something called the Cauchy-Schwarz inequality

Note: also recall the old adage, “correlation is not causation,” so this one number summary says nothing about causality

Proof of 3

Prove that “independence” is stronger than “uncorrelated.” That is prove that independence implies $\text{Corr}(X, Y) = 0$.

Proof of 3

Prove that “independence” is stronger than “uncorrelated.” That is prove that independence implies $\text{Corr}(X, Y) = 0$.

Using the previous slides:

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0$$

So $\text{Corr}(X, Y) = 0$.

Another Example

What is the correlation between X and $Y = aX + b$? a and b are real-valued constants.

Another Example

What is the correlation between X and $Y = aX + b$? a and b are real-valued constants.

$$\begin{aligned}\text{Corr}(X, Y) &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \\&= \frac{\text{Cov}(X, aX + b)}{\sqrt{\text{Var}(X) \text{Var}(aX + b)}} \\&= \frac{\text{Cov}(X, aX)}{\sqrt{\text{Var}(X) \text{Var}(aX)}} \\&= \frac{a \text{Cov}(X, X)}{\sqrt{\text{Var}(X) a^2 \text{Var}(X)}} \\&= \frac{a}{|a|}\end{aligned}$$

Another Example

What is the variance of $w_1X + w_2Y$? What if they are independent? Hint: use bilinearity!

Another Example

What is the variance of $w_1X + w_2Y$? What if they are independent? Hint: use bilinearity!

$$\begin{aligned}\text{Var}[w_1X + w_2Y] &= \text{Cov}(w_1X + w_2Y, w_1X + w_2Y) \\ &= \text{Cov}(w_1X + w_2Y, w_1X) + \text{Cov}(w_1X + w_2Y, w_2Y) \\ &= \text{Cov}(w_1X, w_1X) + \text{Cov}(w_2Y, w_1X) + \\ &\quad \text{Cov}(w_1X, w_2Y) + \text{Cov}(w_2Y, w_2Y) \\ &= w_1^2 \text{Var}[X] + 2w_1w_2 \text{Cov}(X, Y) + w_2^2 \text{Var}[Y]\end{aligned}$$

If they're independent, the covariance is 0. Can you use this result to find the variance of $X - Y$?

Joint Entropy

We don't focus that much on this sort of stuff, but Joint Entropy is another cool example and it is defined for both continuous and discrete rvs.

Let $h(X, Y) = -\log f(X, Y)$ or $h(X, Y) = -\log p(X, Y)$, the “surprise.”

$$E[-\log f(X, Y)] = \iint -\log f(x, y)f(x, y)dx dy$$

or

$$E[-\log p(X, Y)] = \sum_x \sum_y -\log p(x, y)p(x, y)$$

Another Example

Multivariate LOTUS is a special case for univariate LOTUS. Define $h(X, Y) = h(X)$. WLOG assume X and Y are both continuous:

$$E[h(X, Y)] = \iint h(x)f(x, y)dxdy = \int h(x)f(x)dx = E[h(X)]$$

Example

“Kullback-Leibler (KL) Divergence:”

$$KL(p||q) = E_p[\log \left(\frac{p(x)}{q(x)} \right)]$$

Continuous:

$$\int \log \left(\frac{p(x)}{q(x)} \right) p(x) dx$$

or discrete:

$$\sum_x \log \left(\frac{p(x)}{q(x)} \right) p(x)$$

It compares two distributions for the same random variable X .

Example (continued)

“Kullback-Leibler (KL) Divergence”

We could extend this to the multivariate case as follows:

$$KL(f_{X,Y}^1 || f_{X,Y}^2) = E_1 \left[\log \left(\frac{f^1(X, Y)}{f^2(X, Y)} \right) \right]$$

- continuous: $\iint \log \left(\frac{f^1(x,y)}{f^2(x,y)} \right) f_{X,Y}^1(x, y) dx dy$
- discrete : $\sum_x \sum_y \log \left(\frac{f^1(x,y)}{f^2(x,y)} \right) f_{X,Y}^1(x, y)$