Why Covariance and Correlation Measure Linearity

Lecture XIX Supplement

Covariance and Correlation

Covariance

$$\mathrm{Cov}(X, Y) \overset{\text{Definition}}{=} \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \overset{\text{Shortcut}}{=} \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Covariance and Correlation

Covariance

$$\mathrm{Cov}(X, Y) \overset{\text{Definition}}{=} \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \overset{\text{Shortcut}}{=} \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

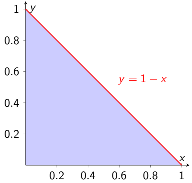
Correlation

The correlation between X and Y is

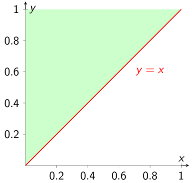
$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y},$$

where σ_X, σ_Y are the standard deviations.
Note that $|\rho_{X,Y}| \leq 1$.

Two Examples from Last Class



(a) $f(x, y) = 2, (x, y) \in T$: Negative-covariance.



(b) $f(x, y) = 2, (x, y) \in T$: Positive-covariance.

$$\mathrm{Cov}(X, Y) = -\frac{1}{36}$$

$$\sigma_X = \sigma_Y = \frac{1}{\sqrt{18}}$$

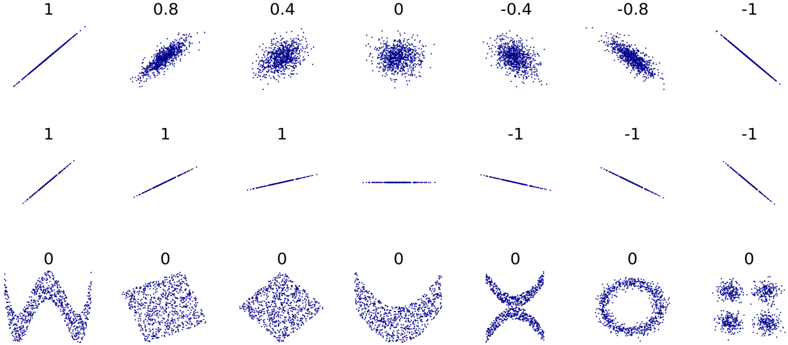
$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = -\frac{1}{2}$$

$$\mathrm{Cov}(X, Y) = \frac{1}{36}$$

$$\sigma_X = \sigma_Y = \frac{1}{\sqrt{18}}$$

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1}{2}$$

Some Example Distributions and Their Correlation Values



Why focus on covariance and correlation?

Much of what we did with expectation and variance for a single variable extends naturally to pairs of random variables.

Why focus on covariance and correlation?

Much of what we did with expectation and variance for a single variable extends naturally to pairs of random variables.

But one question remains: *what exactly does covariance (and correlation) tell us?*

Why focus on covariance and correlation?

Much of what we did with expectation and variance for a single variable extends naturally to pairs of random variables.

But one question remains: *what exactly does covariance (and correlation) tell us?*

- They are easy enough to compute once a joint distribution is known.

Why focus on covariance and correlation?

Much of what we did with expectation and variance for a single variable extends naturally to pairs of random variables.

But one question remains: *what exactly does covariance (and correlation) tell us?*

- They are easy enough to compute once a joint distribution is known.
- Yet, without interpretation, they risk being just another formula.

Why focus on covariance and correlation?

Much of what we did with expectation and variance for a single variable extends naturally to pairs of random variables.

But one question remains: *what exactly does covariance (and correlation) tell us?*

- They are easy enough to compute once a joint distribution is known.
- Yet, without interpretation, they risk being just another formula.

The key insight is that covariance appears naturally when we ask:

Why focus on covariance and correlation?

Much of what we did with expectation and variance for a single variable extends naturally to pairs of random variables.

But one question remains: *what exactly does covariance (and correlation) tell us?*

- They are easy enough to compute once a joint distribution is known.
- Yet, without interpretation, they risk being just another formula.

The key insight is that covariance appears naturally when we ask:

"What is the best linear predictor of Y in terms of X?"

Why focus on covariance and correlation?

Much of what we did with expectation and variance for a single variable extends naturally to pairs of random variables.

But one question remains: *what exactly does covariance (and correlation) tell us?*

- They are easy enough to compute once a joint distribution is known.
- Yet, without interpretation, they risk being just another formula.

The key insight is that covariance appears naturally when we ask:

"What is the best linear predictor of Y in terms of X?"

This least–squares perspective shows that covariance and correlation quantify *linear dependence* — no more, no less.

Motivation

We know covariance appears in formulas, but *why* does it measure **linear** dependence (and it also fails to measure nonlinearity)?

Idea: Ask for the best linear predictor of Y in terms of X.

We try to fit a line:

$$\hat{Y} = aX + b$$

that minimizes the mean squared error, defined by

$$L(a, b) := \mathbb{E}\big[(Y - aX - b)^2\big].$$

Step 1: Differentiate with respect to b

$$L(a, b) = \mathbb{E}\big[(Y - aX - b)^2\big].$$

Differentiate:

$$\frac{\partial}{\partial b} L(a, b) = \mathbb{E}\big[2(Y - aX - b)(-1)\big].$$

So

$$\frac{\partial}{\partial b} L(a, b) = -2\big(\mathbb{E}[Y] - a\mathbb{E}[X] - b\big).$$

Set to zero:

$$b = \mu_Y - a\mu_X.$$

Step 2: Differentiate with respect to a

$$L(a, b) = \mathbb{E}\left[(Y - aX - b)^2\right].$$

Differentiate:

$$\frac{\partial}{\partial a} L(a, b) = \mathbb{E}\left[2(Y - aX - b)(-X)\right].$$

So

$$\frac{\partial}{\partial a} L(a, b) = -2\,\mathbb{E}[X(Y - aX - b)].$$

Set to zero:

$$\mathbb{E}[XY] - a\mathbb{E}[X^2] - b\mu_X = 0.$$

Step 3: Substitute $b = \mu_Y - a\mu_X$

$$\mathbb{E}[XY] - a\mathbb{E}[X^2] - (\mu_Y - a\mu_X)\mu_X = 0.$$

Simplify:

$$\mathbb{E}[XY] - \mu_X\mu_Y - a\big(\mathbb{E}[X^2] - \mu_X^2\big) = 0.$$

Recognize covariance and variance:

$$\mathrm{Cov}(X, Y) - a\,\mathrm{Var}(X) = 0.$$

Step 4: The optimal slope and intercept

So the optimal coefficients are

$$a^* = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}, \qquad b^* = \mu_Y - a^* \mu_X.$$

Interpretation:

- If $\mathrm{Cov}(X, Y) = 0$, then $a^* = 0$: the best linear predictor is $\hat{Y} = \mu_Y$.
- Otherwise, the sign of covariance determines the direction of the slope.

Step 5: Connection to correlation

Rewrite slope as

$$a^* = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}.$$

- $\rho_{X,Y}$ (correlation) tells us the *strength and direction* of the linear relation.
- The ratio σ_Y/σ_X rescales units.

By Cauchy–Schwarz inequality (you will learn this at one point in your life):

$$|\rho_{X,Y}| \leq 1,$$

with equality iff Y is an exact linear function of X.

Takeaway

Key Point

Covariance and correlation arise naturally from the *population least-squares problem*.

- Covariance = raw numerator in the optimal slope.
- Correlation = standardized measure of linearity, bounded between -1 and 1.
- If correlation is 0, best linear predictor is a flat line $\hat{Y} = \mu_Y$.

Therefore: Covariance and correlation quantify *linear dependence*, not arbitrary dependence.