Lecture 5: Collections of Random Variables

Professor Ilias Bilionis

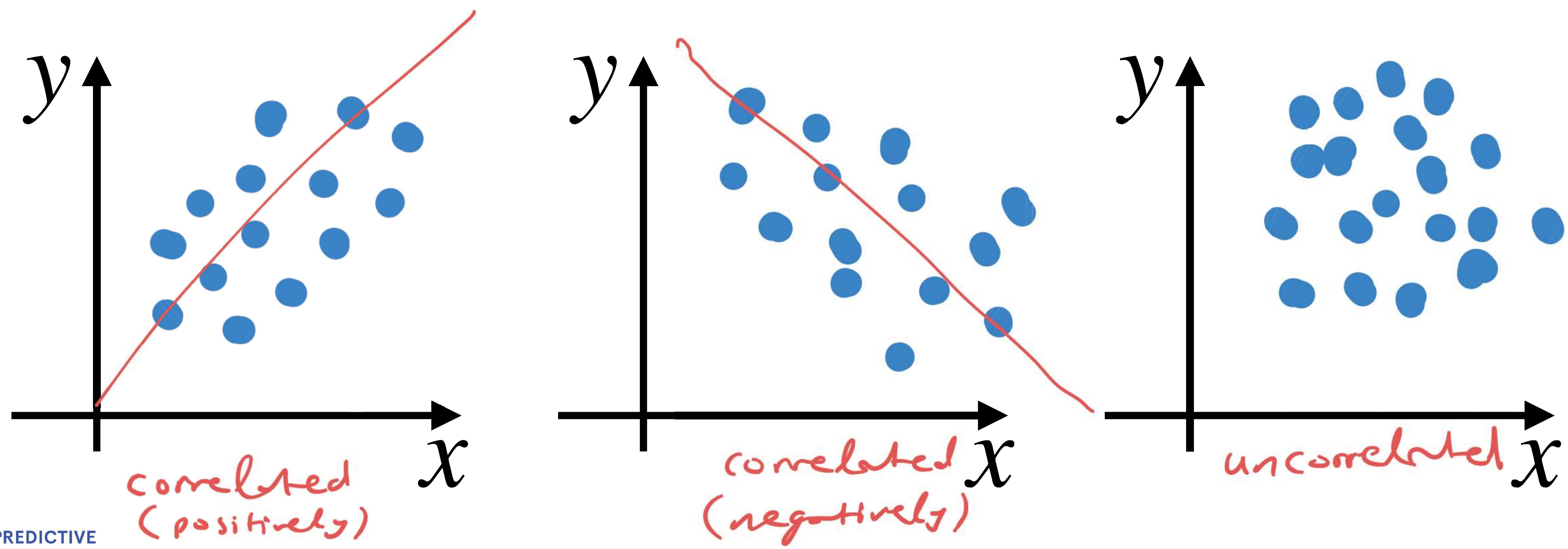
Correlated random variables

PREDICTIVE
SCIENCE LABORATORY

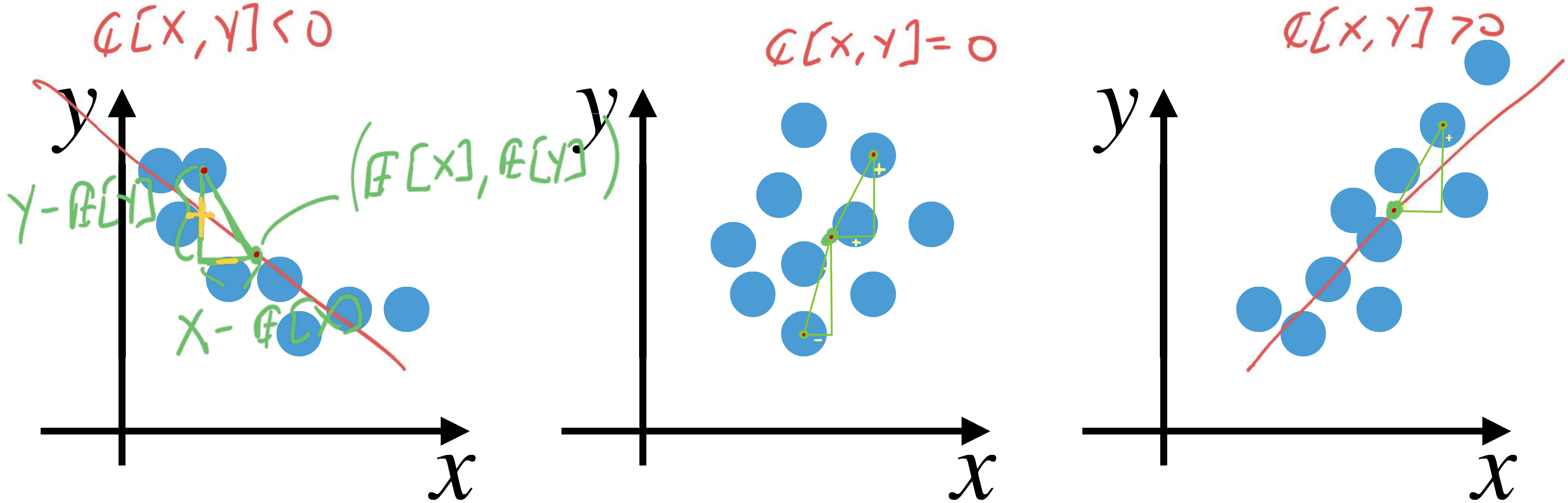
Correlated random variables

- Consider two random variables X and Y.
- Let's take samples from them and visualize the various possibilities.

scatter



The covariance of two random variables

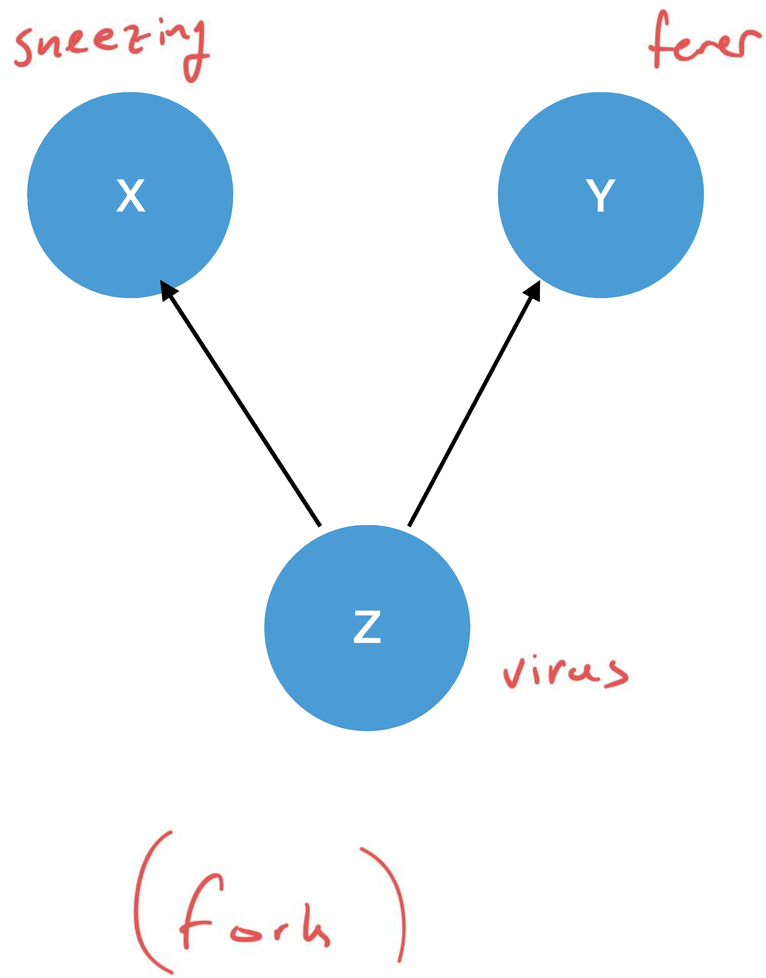


- The covariance operator measures how correlated two random variables X and Y are:

$$\boxed{\mathbb{C}[X,Y] = \mathbb{E}\left[(X - \mathbb{E}[X])\,(Y - \mathbb{E}[Y])\right].}$$

PREDICTIVE SCIENCE LABORATORY

Correlation is not causation



PREDICTIVE
SCIENCE LABORATORY

Properties of the covariance

- Let X be a random variable.
- Then:

self-covariance

$$\mathbb{C}[X,X] = \mathbb{V}[X]$$

Proof:

$$\mathbb{C}[X,X] = \mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])]$$
$$= \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{V}[X]$$

PREDICTIVE
SCIENCE LABORATORY

Properties of the covariance

- Let X be a random variable.
- Then for any constant λ:

$$\mathbb{C}[X,\lambda] = 0$$

Proof: $\mathbb{C}[X,\lambda] = \mathbb{E}[(X-\mathbb{E}[X])\cdot(\underbrace{\lambda - \overset{\lambda}{\mathbb{E}[\lambda]}}_{0})] = 0$

Properties of the covariance

- Let X and Y be two random variables.
- Then:

$$\mathbb{C}[Y, X] = \mathbb{C}[X, Y]$$

Proof:

$$\mathbb{C}[X,Y] = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$$

$$= \mathbb{E}[(Y-\mathbb{E}[Y])(X-\mathbb{E}[X])] = \mathbb{C}[Y,X]$$

Properties of the covariance

- Let X and Y be two random variables.
- Then for any λ and μ:

$$\mathbb{C}[\lambda X, \mu Y] = \lambda\mu\mathbb{C}[X, Y]$$

Proof: $\mathbb{C}[\lambda X, \mu Y] = \mathbb{E}[(\lambda X - \underbrace{\mathbb{E}[\lambda X]}_{\lambda \mathbb{E}[X]}) \cdot (\mu Y - \underbrace{\mathbb{E}[\mu Y]}_{\mu \mathbb{E}[Y]})]$

can then factor
out constants

Properties of the covariance

- Let X and Y be two random variables.

- Then for any λ and μ:

$$\mathbb{C}[X+\lambda, Y+\mu] = \mathbb{C}[X, Y]$$

Proof:

$$\mathbb{C}[X+\lambda, Y+\mu] = \mathbb{E}\left[\left(X+\lambda - \underbrace{\mathbb{E}[X+\lambda]}_{\mathbb{E}[X]+\lambda}\right)\cdot\left(Y+\mu - \mathbb{E}[Y+\mu]\right)\right]$$

constants cancel out

Properties of the covariance

- Let X, Y, and Z be random variables.
- Then:

$$\mathbb{C}[X, Y + Z] = \mathbb{C}[X, Y] + \mathbb{C}[X, Z]$$

Proof:

$$\mathbb{C}[X, Y+Z] = \mathbb{E}\left[(X - \mathbb{E}[X]) \cdot (Y + Z - \underbrace{\mathbb{E}[Y+Z]}_{\mathbb{E}[Y] + \mathbb{E}[Z]})\right]$$

$$= \mathbb{E}\left[\underbrace{(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])}_{\mathbb{C}[X,Y]} + \underbrace{(X - \mathbb{E}[X]) \cdot (Z - \mathbb{E}[Z])}_{\mathbb{C}[X,Z]}\right]$$

$$= \dots$$

Properties of the covariance

- Let X and Y be two random variables.

- Then:

$$
\boxed{\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\mathbb{C}[X,Y]}
$$

Proof:

$$
\mathbb{V}[X+Y] = \mathbb{E}\left[\left(X+Y-\mathbb{E}[X+Y]\right)^2\right] =
$$

$$
= \mathbb{E}\left[\left(\underline{X-\mathbb{E}[X]} + \underline{Y-\mathbb{E}[Y]}\right)^2\right] = \text{FOIL}
$$

$$
= \mathbb{E}\left[\left(X-\mathbb{E}[X]\right)^2 + \left(Y-\mathbb{E}[Y]\right)^2 + 2\left(X-\mathbb{E}[X]\right)\cdot\left(Y-\mathbb{E}[Y]\right)\right]
$$

$$
= \underbrace{\mathbb{E}\left[\left(X-\mathbb{E}[X]\right)^2\right]}_{\mathbb{V}[X]} + \underbrace{\mathbb{E}\left[\left(Y-\mathbb{E}[Y]\right)^2\right]}_{\mathbb{V}[Y]} + 2\underbrace{\mathbb{E}[\ \cdots\ ]}_{\mathbb{C}[X,Y]}
$$

$$C[X,Y] = E[(X-E[X])(Y-E[Y])]$$
$$= E[(X-0.7)(Y-0.7)]$$
$$= \sum_{xy} (x-0.7)(y-0.7)\,p(x,y)$$

if given pdf, can always find expectation using summation or integration

$$= (0-0.7)(0-0.7)(0.1) + (1-0.7)(0-0.7)(0.2)$$
$$+ (0-0.7)(1-0.7)(0.2) + (1-0.7)(1-0.7)(0.5) = 0.01$$

$$C[X,X] = E[(X-E[X])^2]$$
$$= E[(X-0.7)^2]$$
$$= \sum_{X} (x-0.7)^2\,p(x)$$
$$= (0-0.7)^2(0.3) + (1-0.7)^2(0.7) = 0.21$$

$C[Y, 0.31] = 0$ by the properties

$$C[3X, Y] = 3C[X,Y] = 0.03$$

$$V[X+Y] = V[X] + V[Y] + 2C[X,Y] =$$

$$C[X, Y+Z] = C[X,Y] + C[X,Z]$$
$$= C[X,Y] + 0.1$$
$$= 0.01 + 0.1 = 0.11$$

from Q1