Lec 18:

Last time:

Correlation coefficient:

$$\rho_{X,Y} = E\left[\left(\frac{X-E[X]}{\sigma_X}\right)\left(\frac{Y-E[Y]}{\sigma_Y}\right)\right]$$

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

Uncorrelated: $\mathrm{Cov}(X,Y) = 0$

i.e. $\rho_{X,Y} = 0$

Fact: If X, Y are independent r.v.s, then they are uncorrelated.

Proof: (in continuous r.v.s)

Suppose X, Y are independent.

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u \cdot v \, f_{X,Y}(u,v) \, du \, dv$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u \cdot v \cdot f_X(u) \, f_Y(v) \, du \, dv$$

(by independence)

$$= \int_{-\infty}^{\infty} v f_Y(v) \left(\underbrace{\int_{-\infty}^{\infty} u f_X(u)\, du}_{E[X]} \right) dv$$

$$= E[X] \cdot \underbrace{\int_{-\infty}^{\infty} v f_Y(v)\, dv}_{E[Y]}$$

$$= E[X] \cdot E[Y]$$

$$\therefore \quad Cov(X,Y) = E[XY] - E[X]E[Y]$$
$$= 0$$

$\therefore$ X, Y uncorrelated.

we showed:

$$\boxed{\text{independent} \implies \text{uncorrelated}}$$

Converse is generally false,

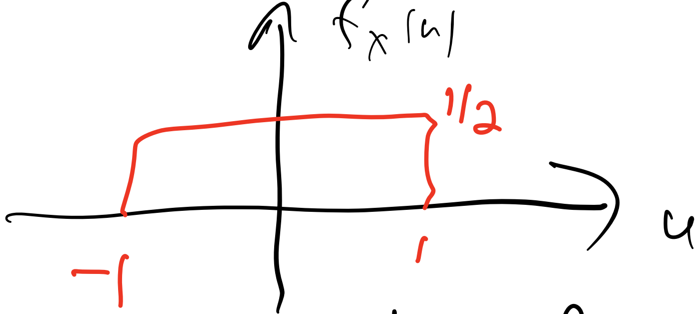
ie. It's possible for some r.v.s X, Y to be uncorrelated, but not independent.

$$\text{ie.} \quad \boxed{\text{uncorrelated} \not\Longrightarrow \text{independent}}$$

Counter example

Let X be uniform on $[-1, 1]$.

Let $Y = X^2$



Claim X, Y uncorrelated, but not indep.

$E[X] = 0$

$E[XY] = E[X \cdot X^2] = E[X^3]$

$$= \int_{-1}^{1} u^3 \cdot \frac{1}{2}\, du = \frac{u^4}{8}\Big|_{-1}^{1} = \frac{1}{8}(1-1) = 0$$

$$\therefore \ \mathrm{Cov}(X,Y) = \underbrace{E[XY]}_{0} - \underbrace{E[X]}_{0}E[Y] = 0$$

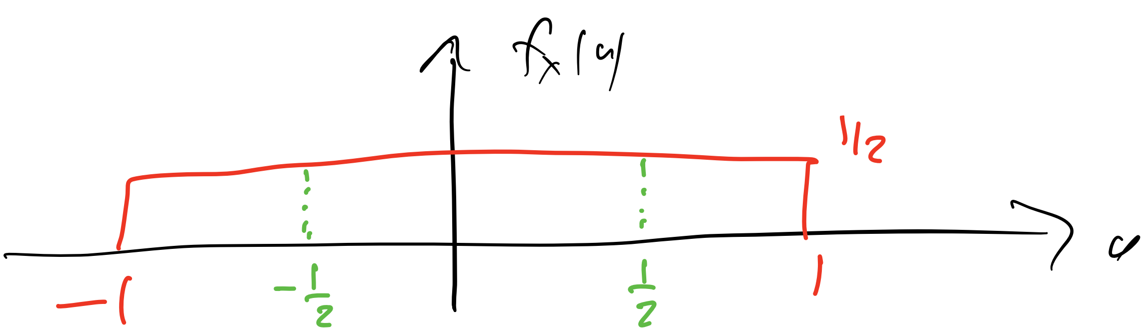
$\therefore$ X, Y uncorrelated.

Show X, Y not indep:

Recall it suffices to find sets A, B of real numbers s.t.

$$P(X \in A, Y \in B) \neq P(X \in A)P(Y \in B)$$

Let's pick $A = [-\frac{1}{2}, \frac{1}{2}]$ $B = [\frac{1}{4}, 1]$



$$P(X\in A) = P\left(-\tfrac{1}{2} \le X \le \tfrac{1}{2}\right) = \tfrac{1}{2}$$

$$P(Y\in B) = P\left(\tfrac{1}{4} \le Y \le 1\right) = P\left(\tfrac{1}{4} \le X^2 \le 1\right)$$

$$= P\left(\tfrac{1}{2} \le X \le 1\right) + P\left(-1 \le X \le -\tfrac{1}{2}\right)$$

$$= \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2}$$

$$P(X\in A, Y\in B) = P\left(-\tfrac{1}{2} \le X \le \tfrac{1}{2}, \tfrac{1}{4} \le Y \le 1\right)$$

$$= P\left(-\tfrac{1}{2} \le X \le \tfrac{1}{2}, \tfrac{1}{4} \le X^2 \le 1\right) = 0$$

$$\neq \underbrace{P(X\in A)}_{1/2}\, \underbrace{P(Y\in B)}_{1/2} = \tfrac{1}{4}$$

$\therefore$ X, Y <u>not</u> independent.

Look at variances of sums:

$$\mathrm{Var}(X+Y) = E[(X+Y)^2] - (E[X+Y])^2$$

$$= E[X^2] + 2E[XY] + E[Y^2]$$

$$- (E[X] + E[Y])^2$$

$$= \underline{E[X^2] - (E[X])^2} + \underline{E[Y^2] - (E[Y])^2}$$

$$+ 2\left(\underline{E[XY] - E[X]E[Y]}\right)$$

$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\cdot \mathrm{Cov}(X,Y)$$

If we repeat this calculation with minus sign, get:

$$Var(X-Y) = Var(X) + Var(Y) - 2 \cdot Cov(X,Y)$$

In summary,

$$\boxed{Var(X \pm Y) = Var(X) + Var(Y) \pm 2\, Cov(X,Y)}$$

Special case: If X, Y uncorrelated,

then $Var(X \pm Y) = Var(X) + Var(Y)$

In particular, true if X, Y indep

Special case:

If $X_1, X_2, \ldots, X_n$ are indep,

then $Var(X_1 \pm X_2 \pm X_3 \pm \cdots \pm X_n)$

$= Var(X_1) + Var(X_2) + \cdots + Var(X_n)$

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

Proof: $0 \leq Var\left(\frac{X}{\sigma_X} \pm \frac{Y}{\sigma_Y}\right)$

$= Var\left(\frac{X}{\sigma_X}\right) + Var\left(\frac{Y}{\sigma_Y}\right) \pm 2 \cdot Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$

$= \frac{1}{\sigma_X^2} \cdot Var(X) + \frac{1}{\sigma_Y^2} Var(Y) \pm \frac{2}{\sigma_X \sigma_Y} \cdot Cov(X,Y)$

$= 1 + 1 \pm 2\rho_{X,Y}$

$$= 2 \pm 2\rho_{X,Y}$$

$$= 2(1 \pm \rho_{X,Y})$$

$$\therefore \quad 0 \leq 1 \pm \rho_{X,Y} \qquad \text{(2 inequalities)}$$

$$0 \leq 1 + \rho_{X,Y} \Rightarrow \rho_{X,Y} \geq -1$$

$$0 \leq 1 - \rho_{X,Y} \Rightarrow \rho_{X,Y} \leq 1$$

Combining,

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

$$\Rightarrow \quad |\rho_{X,Y}| \leq 1 \qquad \underline{\text{Q.E.D.}}$$

<u>Example of application</u>

Why is $\rho_{X,Y}$ useful?

One answer: For "prediction" (or "estimation")

Suppose X, Y have a joint pdf. or joint pmf.

We can only observe Y, but we want to know (or estimate) X.

How can we estimate X from Y?

Form an "estimate" of X as a

deterministic function of Y.

$$\hat{X} = g(Y)$$

estimate of X

function

observe

One way (not nec. best) is a "linear estimate".

$\hat{X} = aY$ where "a" is a constant.

What's the best choice of a?

Let's define one type of "best".

Want $\hat{X} \approx X$

what does this mean.

One popular method is "minimum mean squared error"

error = true value − estimate

$$e = X - \hat{X}$$

Make e small by making e^2 small on average.

e is a r.v.

Try to minimize $E[e^2]$.

$$E[e^2] = E[(X-\hat{X})^2]$$
$$= E[(X-aY)^2]$$
$$= E[X^2] - 2aE[XY] + a^2E[Y^2]$$

Minimize this w.r.t. "a". Find best a.

Set $0 = \frac{d}{da}E[e^2]$
$$= -2E[XY] + 2aE[Y^2]$$

$$\Rightarrow \quad a = \frac{E[XY]}{E[Y^2]} = \text{best choice of } a.$$

Special case:
$$\sigma_X = \sigma_Y = 1$$
$$E[X] = E[Y] = 0$$

$\Rightarrow E[Y^2] = 1$ since $\sigma_Y^2 = E[Y^2] - (E[Y])^2$

Then $a = E[XY] = \rho_{X,Y}$ in this case.

i.e. best linear estimate of X from Y is

$$\boxed{\hat{X} = \rho_{X,Y} \cdot Y}$$

Recall, binomial pmf:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$k = 0, 1, \ldots, n$$

X = # Heads in n flips of biased coin.

we derived mean of X is

$$E[X] = np$$

Alternative derivation

Let $Y_i = \begin{cases} 1 & \text{if Heads on } i^{th} \text{ flip} \\ 0 & \text{if Tails " "} \end{cases}$

we have rvs $Y_1, Y_2, \ldots, Y_n$ iid

we can write

$$X = Y_1 + Y_2 + \cdots + Y_n$$

$$\begin{aligned} E[X] &= E[Y_1 + \cdots + Y_n] \\ &= E[Y_1] + \cdots + E[Y_n] \\ &= n \cdot E[Y_1] \\ &= n(1 \cdot P(\text{Heads}) + 0 \cdot P(\text{Tails})) \\ &= n(1 \cdot p + 0(1-p)) \\ &= np \end{aligned}$$

same answer

Def: Random variables X, Y are called jointly Gaussian if their joint pdf is:

$$f_{X,Y}(u,v) = \frac{1}{2\pi\,\sigma_X \sigma_Y \sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{u-m_X}{\sigma_X}\right)^2 + \left(\frac{v-m_Y}{\sigma_Y}\right)^2 - 2\rho\left(\frac{u-m_X}{\sigma_X}\right)\left(\frac{v-m_Y}{\sigma_Y}\right)\right]}$$

It turns out ρ is indeed the correlation coefficient of X, Y.

Special Case: $m_X = m_Y = 0$
$\sigma_X = \sigma_Y = 1$

$$f_{X,Y}(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)}$$

If $\rho = 0$, get:

$$f_{X,Y}(u,v) = \underbrace{\frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{u-m_X}{\sigma_X}\right)^2}}_{f_X(u)} \cdot \underbrace{\frac{1}{\sigma_Y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{v-m_Y}{\sigma_Y}\right)^2}}_{f_Y(v)}$$

$N(m_X, \sigma_X^2)$ $\qquad$ $N(m_Y, \sigma_Y^2)$

So for joint Gaussians,

X, Y uncorrelated $\Longrightarrow$ X, Y independent

I.e. for Gaussians

X, Y uncorrelated $\Longleftrightarrow$ X, Y independent

Can generalize to n jointly Gaussian r.v.s. Used in statistics.