4/15/22

Covariance

- HW 10 (Apr 22), HW 11 (Apr 29)
- Quiz 6 Apr 27 (Wed) (3 ~ 4 Q w/ subquestions)
- Final
  - 30 Take-home E posted on May 6, Due May 12
  - 70 May 13, PL exam.

    Time window : 8am – 11am

    Time Limit : 90 min. (10 ~ 12 problem)

    TF / MC / Free response

Next Monday. { give back Exam; go through the questions.

<u>Recall</u> $A_1, A_2, A_3, \cdots, A_n$: Events.

$$X = \mathbb{1}_{A_1} + \mathbb{1}_{A_2} + \cdots + \mathbb{1}_{A_n}$$

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[(X)^2] - \mathbb{E}[X]^2 \\
&= \sum_{i=1}^{n} \underbrace{\mathbb{P}(A_i)(1-\mathbb{P}(A_i))}_{=\mathrm{Var}(\mathbb{1}_{A_i})} + 2\sum_{i<j}\left(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\cdot\mathbb{P}(A_j)\right) \\
&= \sum_{i=1}^{n} \mathrm{Var}(\mathbb{1}_{A_i}) + \boxed{?}
\end{aligned}$$

What is this.

In general, $X = X_1 + X_2$

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

$$\mathbb{E}[X^2] = \mathbb{E}[(X_1+X_2)^2] = \mathbb{E}[X_1^2 + X_2^2 + 2X_1\cdot X_2].$$

$$\mathrm{Var}(X) = \mathbb{E} X_1^2 + \mathbb{E} X_2^2 + 2\,\mathbb{E}[X_1\cdot X_2] - \left(\mathbb{E}[X_1] + \mathbb{E}[X_2]\right)^2$$

$$= \left(\mathbb{E} X_1^2 - (\mathbb{E} X_1)^2\right) + \left(\mathbb{E} X_2^2 - (\mathbb{E} X_2)^2\right)$$

$$+ 2\left(\mathbb{E}[X_1\cdot X_2] - (\mathbb{E} X_1)(\mathbb{E} X_2)\right)$$

$X_1 = X_2 = Y$

$X = X_1 + X_2 = 2Y$

$$= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\underbrace{\left(\mathbb{E} X_1\cdot X_2 - \mathbb{E} X_1\cdot \mathbb{E} X_2\right)}_{\overset{\text{def}}{=}\ \mathrm{Cov}(X_1, X_2)}$$

$$\mathrm{Var}(2Y) = \mathrm{Var}(Y) + \mathrm{Var}(Y) + 2\cdot \mathrm{Var}(Y)$$
$$= 4\cdot \mathrm{Var}(Y)$$

Def The covariance of X_1, X_2 is

$$\mathrm{Cov}(X_1, X_2) = \underbrace{\mathbb{E}[X_1\cdot X_2] - \mathbb{E} X_1\cdot \mathbb{E} X_2}$$
$$= \mathbb{E}\left[(X_1 - \mathbb{E} X_1)\cdot(X_2 - \mathbb{E} X_2)\right]$$

$$\mathbb{E}\left[(X_1 - \mathbb{E} X_1)\cdot(X_2 - \mathbb{E} X_2)\right] = \mathbb{E}\big[X_1\cdot X_2 - X_1\cdot(\mathbb{E}X_2) - X_2\cdot(\mathbb{E} X_1)$$
$$+ \mathbb{E}X_1\cdot \mathbb{E}X_2\big]$$

$$= \boxed{\mathbb{E}[X_1\cdot X_2] - \mathbb{E}X_2\ \mathbb{E}[X_1]} - \mathbb{E}[\cancel{X_2\cdot}](\mathbb{E}X_1)$$
$$+ \mathbb{E}[\cancel{X_1]\cdot \mathbb{E}X_2}]$$

Properties

1. If X, Y independent, $\mathbb{E}[X\cdot Y] = \mathbb{E}X\cdot \mathbb{E}Y$

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

Q: $\mathrm{Cov}(X,Y) = 0$ does this imply X, Y indep. ?

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

$$= \mathbb{E}\left[(X - \mathbb{E}X)\cdot(X - \mathbb{E}X)\right] = \mathrm{Var}(X)$$

3. $\mathrm{Cov}(aX + b, Y) = \mathbb{E}[\underbrace{(aX+b)\cdot Y}_{aXY + bY}] - \overbrace{\mathbb{E}[aX + b]}^{(a\mathbb{E}X + b)}\cdot \mathbb{E}[Y].$

$a, b \in \mathbb{R}$

$$= a\,\mathbb{E}[XY] + \cancel{b\,\mathbb{E}Y} - \left(a\,\mathbb{E}X\cdot \mathbb{E}Y + \cancel{b\cdot \mathbb{E}Y}\right)$$

$$= a\,(\mathbb{E}XY - \mathbb{E}X\cdot\mathbb{E}Y)$$

$$= a\cdot \mathrm{Cov}(X,Y)$$

4. $\mathrm{Cov}(X_1+X_2+\cdots+X_n,\ Y_1+Y_2+\cdots+Y_m)$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{cov}(X_i, Y_j)$$

In particular,

$$\mathrm{Var}(X_1+X_2+\cdots+X_n) = \mathrm{Cov}(X_1+X_2+\cdots+X_n,\ X_1+X_2+\cdots+X_n)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

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

<u>Def</u> Correlation Coefficient

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

$$\mathrm{Cov}(X,Y) = \mathbb{E}[\underbrace{(X-\mathbb{E}X)}_{=\overline{X}}\cdot\underbrace{(Y-\mathbb{E}Y)}_{=:\overline{Y}}]$$

$$= \mathbb{E}[\overline{X}\cdot\overline{Y}]$$

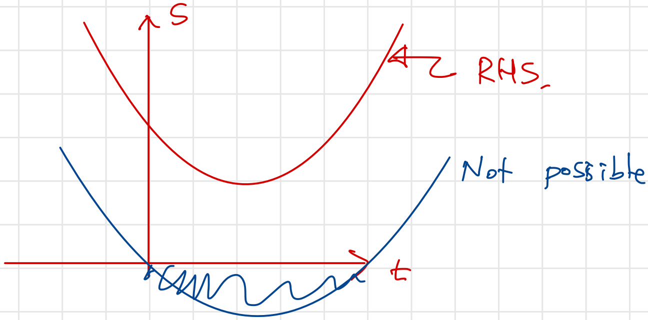
For $t\in\mathbb{R}$,

$$0 \le \mathbb{E}[(\overline{X}-t\cdot\overline{Y})^2] = \mathbb{E}[\overline{X}^2 - 2t\cdot\overline{X}\cdot\overline{Y} + t^2\overline{Y}^2]$$

Not depend on t.

$$= \mathbb{E}[\overline{X}^2] - 2t\cdot\mathbb{E}[\overline{X}\,\overline{Y}] + t^2\cdot\mathbb{E}[\overline{Y}^2]$$

$$= \underbrace{\mathrm{Var}(X) - 2t\cdot\mathrm{Cov}(X,Y) + t^2\cdot\mathrm{Var}(Y)}_{2^{nd}\text{ degree polynomial.}} = f.$$



No distinct. real solution

$$\underbrace{\qquad}_{} t^2 \cdot Var(Y) - 2t \cdot Cov(X,Y) + Var(X) = 0$$

$$\Rightarrow \quad Cov(X,Y)^2 - Var(X) \cdot Var(Y) \leq 0 .$$

$$\Rightarrow \quad Cov(X,Y)^2 \leq Var(X) \cdot Var(Y) .$$

$$\Bigg(\frac{Cov(X,Y)}{\underbrace{\sqrt{Var(X) \cdot Var(Y)}}_{=:\, \rho(X,Y)}} \Bigg)^2 \leq 1$$

$$\Rightarrow \quad -1 \leq \rho(X,Y) \leq 1$$

Correlation Coefficient.