Lecture 30

04/16

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E[\overline{X}] = \mu$$

$\overline{X}$ is unbiased estimator of μ.

$E(X+Y) = E(X) + E(Y)$ for any X, Y ← no need to make any assumptions.

$Var(X+Y) = Var(X) + Var(Y)$? ← no, not in general.

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$, where

$Cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right]$

- If $X < E(X)$ and $Y < E(Y)$, then the $Cov(X,Y)$ will be positive and possibly large.

$= E\left[XY - E(X)Y - XE(Y) + E(X)E(Y)\right]$

$= E[XY] - E(X)E(Y) - \cancel{E(X)E(Y)} + \cancel{E(X)E(Y)}$

$$\boxed{= E(XY) - E(X)E(Y)}$$

- Covariance of X & Y, $Cov(X,Y)$, characterizes possible dependence correlation between X & Y.
- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,X) = Var(X)$
- $Cov(aX+b, cY+d) = ac \cdot Cov(X,Y)$

$= E([aX - E(aX)][cY - E(cY)]) = ac \cdot Cov(X,Y)$

$Var(X+Y) = E\left[(X+Y)^2\right] - \left[E(X+Y)\right]^2 = E\left[(X+Y) - (E(X) + E(Y)\right]$

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

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

$= Var(X) + 2Cov(X,Y) + Var(Y)$

$\cdot\ V(X-Y) = E[X-Y-(E(X)-E(Y))^2] = E[(X-E(X)-(Y-E(Y)))]$

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

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

· If X and Y are indep, then

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

· So, in this case,

$$Var(X+Y) = Var(X) + Var(Y)$$

$$Var(X-Y) = Var(X) + Var(Y)$$

However, the converse is not necessarily true, i.e.,

$Cov(X,Y) = 0$ does not imply $X \perp\!\!\!\perp Y$

<u>Ex</u>: $X = \begin{cases} 1, & w/\ p\ \frac{1}{2} \\ -1, & w/\ p\ \frac{1}{2} \end{cases}$

$Y = 0$, if $X = -1$

$Y = \begin{cases} -1 & w/p\ \frac{1}{2} \\ 1 & w/p\ \frac{1}{2} \end{cases}$, if $X = 1$

$E(XY) = 0 \cdot \frac{1}{2} + 1 \times \frac{1}{4} - 1 \times \frac{1}{4} = 0$

$E(X) = 0$, $E(Y) = 0$, $Cov(X,Y) = 0$

$P(X=1, Y=1) = \frac{1}{4}$

$P(X=1) = \frac{1}{2}$, $P(Y=1) = \frac{1}{4}$

$\Rightarrow P(X=1)P(Y=1) \neq P(X=1, Y=1)$

Thus, X and Y are not indep.

$X_1, \dots, X_n \overset{i.i.d.}{\sim} (\mu, \sigma^2)$

$\bar{X}$ can raise the accuracy of estimation.

$Var(\bar{X}) = \frac{1}{n}\sigma^2$

$Var(X_1) = \sigma^2$

$Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{1}{n}\sigma^2$

$X_1, \dots, X_n \overset{i.i.d.}{\sim} Bernoulli(p)$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{p}$ (← sample proportion)

$E(\bar{X}) = p$, $E(\hat{p}) = p$ 、

$Var(\hat{p}) = Var(\bar{X}) = \frac{1}{n} p(1-p)$

· Things will be different if we don't have independence

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

· $X \perp\!\!\!\perp Y$ independence

· $X \not\perp\!\!\!\perp Y$ dependence

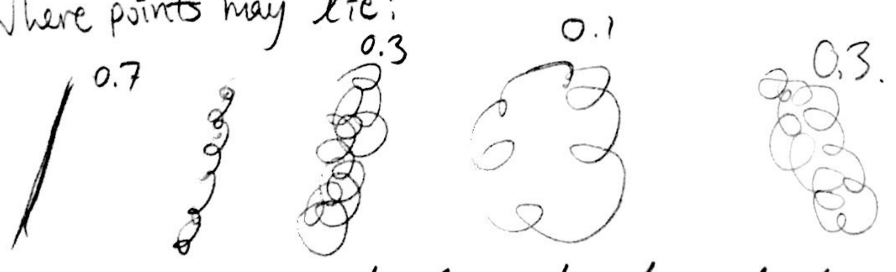
↓

linear dependence

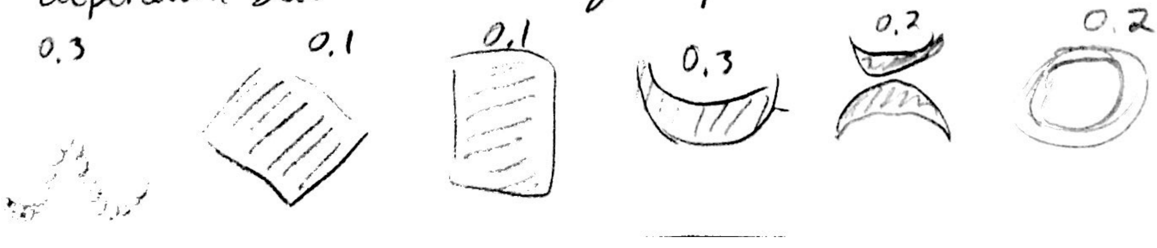
↑

could be quantified w/ covariance.

Where points may lie:



dependent but not linearly dependent.



Peason's Correlation ρ:

$Corr(X,Y)$ or $\rho(x,y)$

$$Corr(X,Y) = \frac{Cov(x,y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{x,Y}}{\sigma_x \sigma_Y}$$

Cauchy-Schwarty Inequality: $|Cov(X,Y)| \le \sqrt{Var(X)Var(Y)}$

$\rho_{x,y} \in [-1,1]$

if $X \perp\!\!\!\perp Y \Rightarrow \rho_{x,Y} = 0$

$\rightarrow$ the reverse is not true, i.e.,

$\rho_{x,Y} = 0 \not\Rightarrow X \perp\!\!\!\perp Y$

$\rho_{x,Y} = \pm 1$ iff $Y = aX + b$, for some $a \neq 0$ and b

for which case X and Y are said to be perfectly correlated.

Ex:

		Y = 0	Y = 1
X	-1	$\frac{1}{3}$	0
	0	0	$\frac{1}{3}$
	1	$\frac{1}{3}$	0

$E(XY) = 0$

$E(X) = 0$

$E(Y) = \frac{1}{3}$

$Cov(X,Y) = 0 \Rightarrow \rho_{x,Y} = 0$

$P(X=0, Y=1) = \frac{1}{3}$

$P(X=0) = \frac{1}{3}$

$P(Y=1) = \frac{1}{3}$

$\Rightarrow P(X=0)P(Y=1) \neq P(X=0, Y=1)$

Ex: $X \sim U(-1,1)$, $Y = X^2$. find $\rho_{x,Y}$

$E(X) = 0$

$E(X^3) = 0$ $\quad \int_{-1}^{1} \frac{x^3}{2}\, dx = 0$

$Cov(X,Y) = Cov(X,X^2) = \underbrace{E(X^3)}_{0} - \underbrace{E(X)}_{0}\underbrace{E(X^2)}_{0} = 0 \Rightarrow \rho_{x,Y} = 0.$

Y is completely determined by X.
$Y = X^2$, but it's not a linear relation.