Lecture 19 - Covariance/Correlation

Tuesday, November 16, 2021 9:32 AM

Defn: Covariance

We define the covariance btwn two RVs X and Y as

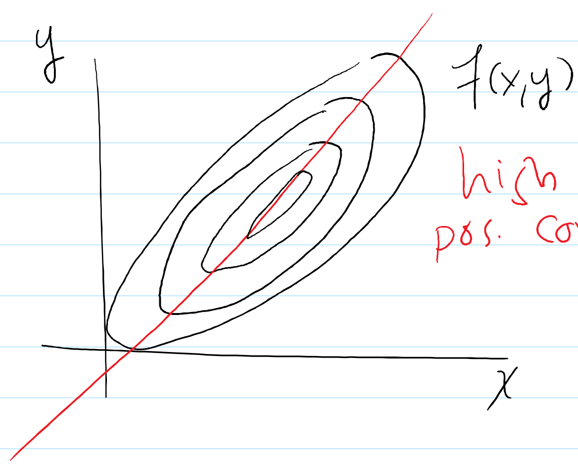
$$\mathrm{Cov}(X,Y) = E\left[(X-EX)(Y-EY)\right]$$

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

if $\mu_X = EX$ and $\mu_Y = EY$ then

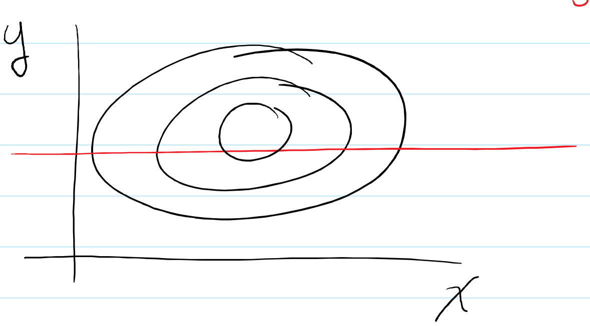
$$= E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$g(x,y) = (x-\mu_X)(y-\mu_Y)$



high pos. cov.

high neg. cov.



cov ≈ 0

Defn: Correlation

Re-scaled cov. to be between ± 1.

cor.
$+1$ = perfect lin. rel. (pos)
-1 = " neg. lin. rel
0 = no lin. rel.

we define

$$Cor(X, Y) = \frac{cov(X, Y)}{\sqrt{Var(X)\,Var(Y)}}$$

$$sd(X) = \sqrt{Var(X)}$$
$$sd(Y) = \sqrt{Var(Y)}$$

$$= \frac{cov(X, Y)}{sd(X)\,sd(Y)}.$$

Theorem If $a, b \in \mathbb{R}$ then

$$\boxed{Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X, Y)}$$

pf. $Z = aX + bY$

$$\mathrm{Var}(Z) = \mathbb{E}\left[\underline{(Z - E[Z])}^2\right]$$

$$Z - \mathbb{E}[Z] = aX + bY - \mathbb{E}[aX + bY]$$

$$= aX + bY - aEX - bEY$$

$$= \underline{a(X - EX) + b(Y - EY)}$$

$$\mathrm{Var}(Z) = \mathbb{E}\left[(Z - EZ)^2\right] = a^2 \underbrace{\mathbb{E}\left[(X - EX)^2\right]}_{\mathrm{Var}(X)} + b^2 \underbrace{\mathbb{E}\left[(Y - EY)^2\right]}_{\mathrm{Var}(Y)}$$

$$+ 2ab \underbrace{\mathbb{E}\left[(X - EX)(Y - EY)\right]}_{\mathrm{Cov}(X,Y)}$$

Theorem: if $a, b \in \mathbb{R}$

then $$\boxed{\mathrm{Cov}(aX + b, Y) = a\,\mathrm{Cov}(X, Y)}$$

pf.

$$\mathrm{Cov}(aX+b, Y) = \mathbb{E}\left[\left(aX \cancel{+b} - \overbrace{E[aX+b]}^{a\mathbb{E}X \cancel{+b}}\right)\left(Y - E[Y]\right)\right]$$

$$= \mathbb{E}\left[a(X - EX)(Y - EY)\right]$$

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

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

Note: $\mathrm{Cov}(X, aY+b) = a\,\mathrm{Cov}(X,Y)$

Further: $\boxed{\mathrm{Cov}(aX+b,\ cY+d) = ac\,\mathrm{Cov}(X,Y)}$

Theorem:

$$\boxed{\mathrm{Cor}(aX+b,\ cY+d) = \mathrm{Sign}(a)\,\mathrm{Sign}(c)\,\mathrm{Cor}(X,Y)}$$

$$\mathrm{Sign}(x) = \begin{cases} +1 & x>0 \\ 0 & x=0 \\ -1 & x<0 \end{cases}$$

if $X \neq 0$ then $\mathrm{sign}(x) = \dfrac{x}{|X|}$

Ex. $\mathrm{Cor}(-5X, Y) = -\mathrm{Cor}(X,Y)$

pf.

$$\mathrm{Cor}(aX+b,\ cY+d) = \frac{\mathrm{Cov}(aX+b,\ cY+d)}{\sqrt{\mathrm{Var}(aX+b)}\,\sqrt{\mathrm{Var}(cY+d)}}$$

$\sqrt{a^2} = |a|$

$$= \frac{ac\,\mathrm{Cov}(X,Y)}{\sqrt{a^2 \ldots}\ \sqrt{c^2 \mathrm{Var}(Y)}}$$

$\sqrt{a^2} = |a|$

$$= \frac{}{\sqrt{a^2 Var(X)}\ \sqrt{c^2 Var(Y)}}$$

$$= \underbrace{\frac{a}{|a|}}_{sign(a)} \underbrace{\frac{c}{|c|}}_{sign(b)} \underbrace{\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}}_{Cor(X,Y)}$$

Claim: $\boxed{-1 \le Cor(X,Y) \le 1}$

linear transf

$$\tilde{X} = \frac{X - E[X]}{sd(X)} \rightsquigarrow E\tilde{X} = 0$$

$$Var(\tilde{X}) = 1$$

$$\tilde{Y} = \frac{Y - EY}{sd(Y)}$$

① $Cor(X,Y) = Cor(\tilde{X}, \tilde{Y})$

② $Var(\tilde{X} \pm \tilde{Y}) = \underbrace{Var(\tilde{X})}_{1} + \underbrace{Var(\tilde{Y})}_{1} \pm 2Cov(\tilde{X}, \tilde{Y})$

$$= 2 \pm 2\,Cov(\tilde{X}, \tilde{Y})$$

③ $\text{Cor}(\tilde{X}, \tilde{Y}) = \dfrac{\text{Cov}(\tilde{X}, \tilde{Y})}{1 \cdot 1} = \text{Cov}(\tilde{X}, \tilde{Y})$

$$= \underline{2 \pm 2\,\text{Cor}(\tilde{X}, \tilde{Y})}$$

var ≥ 0

$$= \underline{2 \pm 2\,\text{Cor}(X, Y)} \geq 0$$

so

$$1 \pm \text{Cor}(X, Y) \geq 0$$

either

$1 + \text{Cor} \geq 0$ and $1 - \text{Cor} \geq 0$

$\boxed{\text{Cor} \geq -1}$ and $\boxed{\text{Cor} \leq 1}$

Theorem: Covariance Short-cut

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - (\mathbb{E}X)(\mathbb{E}Y).$$

Ex, continue from last lect.

$$f(x,y) = 1 \quad \text{for} \quad \begin{array}{l} 0 < x < 1 \\ x < y < x+1 \end{array}$$

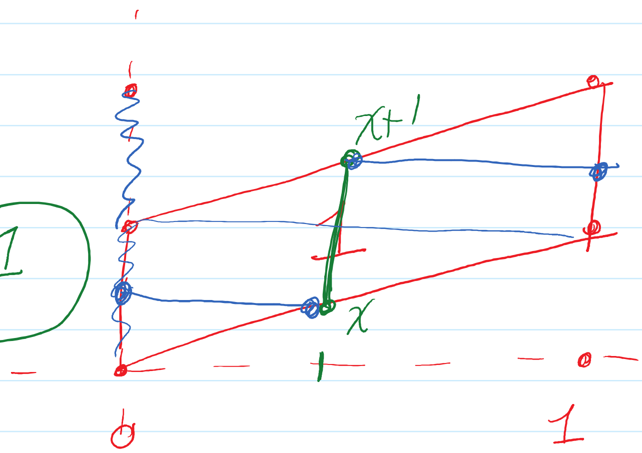
Showed: $E[XY] = 7/12$

Want: $\boxed{Cov(X,Y) = EXY - EXEY}$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{x}^{x+1} 1\,dy$$

$$= (x+1) - x = 1$$

for $0 < x < 1$



$X \sim U(0,1) \longrightarrow EX = 1/2,\ Var(X) = \frac{1}{12}$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \begin{cases} \int_0^y 1\,dx & , 0 < y < 1 \\ \int_{y-1}^{2} 1\,dx & , 1 < y < 2 \end{cases}$$

$$\boxed{= \begin{cases} y, & 0 < y < 1 \\ 2-y, & 1 < y < 2 \end{cases}}$$

Can show: $E[Y] = 1$ and $Var(Y) = 1/6$

So $\boxed{\begin{aligned} &\mathrm{Cov}(X,Y) = \left(7/12\right) - \left(\tfrac{1}{2}\right)(1) \\ &\mathrm{Cor}(X,Y) = \frac{(7/12) - (1/2)(1)}{\sqrt{1/12}\,\sqrt{1/6}} \end{aligned}}$

Recall: Conditional probability

e.g. $$P(A|B) = \frac{P(AB)}{P(B)}.$$

If X and Y are discrete let

$A = \{X = x\}$ and $B = \{Y = y\}$ then

$$\underbrace{P(X = x \mid Y = y)}_{\text{conditional dist}} = P(A|B) = \frac{P(AB)}{P(B)}$$

$$= \frac{\overbrace{P(X = x, Y = y)}^{\text{joint dist}}}{\underbrace{P(Y = y)}_{\text{marginal of } Y}}$$

Defn: Conditional PMF (discrete)

Defn: Conditional PMF (discrete)

The conditional PMF of X given $Y=y$ is defined as

$$f_{X|Y=y}(x) = f(x|y) = \frac{f(x,y)}{f_Y(y)}$$

Ex, Joint PMF $f_X(0) = 4/18$

y \ x	0	1	2
30	0	3/18	4/18
20	2/18	4/18	0
10	2/18	3/18	0

what is the dist. of Y given $X=0$

$$f_{Y|X=0}(y) = f(y|0) = \frac{f(0,y)}{f_X(0)}$$

$$f(y|x) = \begin{cases} \dfrac{f(0,10)}{f_X(0)} = \dfrac{2/18}{4/18} = 1/2 & y=10 \\ \dfrac{f(0,20)}{\ldots} = \dfrac{2/18}{\ldots} = 1/2 & y=20 \end{cases}$$

$$f(y|0) = \begin{cases} \dfrac{f(0,20)}{f_X(0)} = \dfrac{2/18}{4/18} = \tfrac{1}{2} & y = 20 \\[2ex] \dfrac{f(0,30)}{f_X(0)} = \dfrac{0}{4/18} = 0 & y = 30 \end{cases}$$

<u>what about continuous?</u>

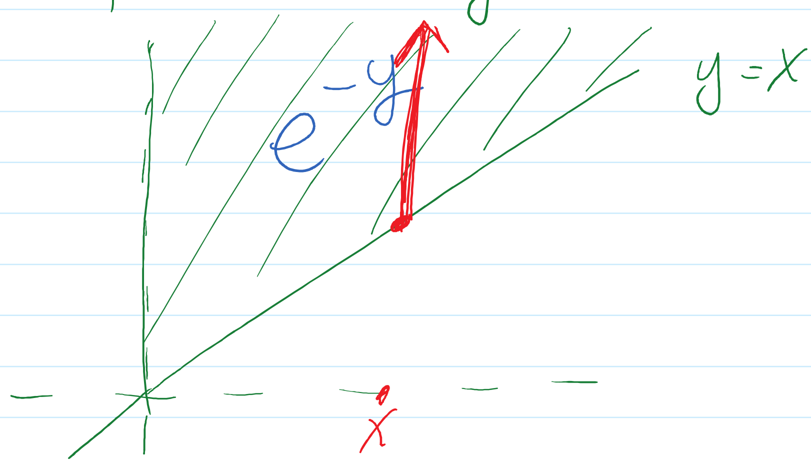
If X and Y are cts then the conditional PDF of X given $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}$$

<u>Ex,</u> $f(x,y) = e^{-y}$ for $0 < x < y$

<u>Q:</u> what is

$$f(y|x) = \frac{f(x,y)}{f_X(x)}$$



$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{x}^{\infty} e^{-y}\,dy = e^{-x}$$

Hence

Hence

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \boxed{\frac{e^{-y}}{e^{-x}} \quad 0 < x < y}$$