СЕМ
лек. 7

Ковариация и корелация

17/11/22

Деф. Нека (X, Y) са 2 сл. вел. във вер. пр-во V.
(ков.) Тогава $cov(X, Y) := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$
се нарича ковариация на X и Y.

(X, Y)
$X = f(Y)$ някаква f
$X = aY + b$
$X \approx aY + b$

⊕ $cov(X, Y) = \sum_{i,j} (x_i - \mathbb{E}X)(y_j - \mathbb{E}Y)\, p_{ij}$

$\tilde{X} = X - \mathbb{E}X$ центриране на X
$\tilde{Y} = Y - \mathbb{E}Y$ –"– Y

Тв: За (X, Y) имаме, че $cov(X, Y) = \mathbb{E}XY - \mathbb{E}X \cdot \mathbb{E}Y$ и ако $X \perp\!\!\!\perp Y$, то $cov(X, Y) = 0$

Д-во $cov(X; Y) = \mathbb{E}(X \cdot Y - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}X \cdot \mathbb{E}Y)$
$= \mathbb{E}XY - \cancel{\mathbb{E}Y}\mathbb{E}X - \mathbb{E}X\mathbb{E}Y + \cancel{\mathbb{E}X}\mathbb{E}Y$

Ако $X \perp\!\!\!\perp Y$, то $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$

Не е вярно, че $cov(X, Y) = 0 \Rightarrow X \perp\!\!\!\perp Y$

⊕ $X \to 10X$ $\quad cov(10X, 10Y) = 100\, cov(X, Y)$
$Y \to 10Y$

Деф (X, Y) са 2 сл. вел., такива че $DX < \infty$, $DY < \infty$.
(корел.) Тогава $\rho(X, Y) = \dfrac{cov(X, Y)}{\sqrt{DX}\sqrt{DY}}$. (коеф. за корелация)

⊕ $\rho(10X, 10Y) = \dfrac{100\, cov(X, Y)}{\sqrt{D10X}\sqrt{D10Y}} = \dfrac{100\, cov(X, Y)}{10\sqrt{DX}\, 10\sqrt{DY}} = \rho(X, Y)$

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

Тв: Ако X и Y са сл. вел. такива, че D...
Тогава с $\bar{X} = \dfrac{X - \mathbb{E}X}{\sqrt{DX}}$ и $\bar{Y} = \dfrac{Y - \mathbb{E}Y}{\sqrt{DY}}$ имаме, че
3) $\rho(X, Y) = \mathbb{E}\bar{X}\bar{Y}$ и 1) $\mathbb{E}\bar{X} = \mathbb{E}\bar{Y} \overset{?}{=} 0 \to$ центриране
2) $D\bar{X} = D\bar{Y} = 1$
} центриране и нормиране

1) $\mathbb{E}\bar{X} = \mathbb{E}\left(\dfrac{X - \mathbb{E}X}{\sqrt{DX}}\right) = \dfrac{1}{\sqrt{DX}}\mathbb{E}(X-$ ... $= 0$
–"– за $\bar{Y}$
2) $1 = D\bar{X} = D\left(\dfrac{X - \mathbb{E}X}{\sqrt{DX}}\right) = \dfrac{1}{DX} D(X - \mathbb{E}$ ... $= \dfrac{DX}{DX} = 1$
! $D(X + c) = DX$

Д-во 3) $\rho(X, Y) = \dfrac{\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)}{\sqrt{DX}\sqrt{DY}} = \mathbb{E}\bar{X}\bar{Y}$

Scanned with CamScanner

$D(X+c) = DX,\ c \in \mathbb{R}$

Теор. Нека X и Y са 2 сл. вел., такива че $DX < \infty$ и $DY < \infty$.
Тогава а) $|\rho(X, Y)| \leq 1$
б) $|\rho(X, Y)| = 1 \iff Y = aX + b,\ \exists a, b \in \mathbb{R}$

Д-во а) $\rho(X, Y) = \mathbb{E}\bar{X}\bar{Y}$

Вярно е, че $0 \leq \mathbb{E}(\bar{X} + \bar{Y})^2 = \mathbb{E}\bar{X}^2 + \ldots + \mathbb{E}\bar{Y}^2$

$\ldots \bar{Y}^2$

$0 \leq \mathbb{E}(\bar{X} + \bar{Y})^2 = \mathbb{E}X^2 + \mathbb{E}Y^2 + 2\mathbb{E}\bar{X}\bar{Y}$
$= 2 + 2\mathbb{E}\bar{X}\bar{Y}$
$\Rightarrow \mathbb{E}\bar{X}\bar{Y} \geq -1$

$0 \leq \mathbb{E}(\bar{X} - \bar{Y})^2 = \mathbb{E}X^2 + \mathbb{E}Y^2 - 2\mathbb{E}\bar{X}\bar{Y}$
$= 2 - 2\mathbb{E}\bar{X}\bar{Y}$
$\Rightarrow \mathbb{E}\bar{X}\bar{Y} \leq 1$

$\Rightarrow |\rho(X, Y)| \leq 1$

$\rho(X, Y) = 0 \Rightarrow$ линейна независимост
$\not\Rightarrow$ независимост

$\Rightarrow \rho(X \ldots$
$\rho(X, Y) \leq 1$

б) "$\Leftarrow$" $Y = aX + b$ за някакви a, b

$Y - \mathbb{E}Y = a(X - \mathbb{E}X + \mathbb{E}X) + b - \mathbb{E}Y$
$= a(X - \mathbb{E}X) + a\mathbb{E}X + b - \mathbb{E}Y \quad / \sqrt{DY}$

$$\frac{Y - \mathbb{E}Y}{\sqrt{DY}} = \underbrace{\frac{\sqrt{DX}\, a}{\sqrt{DY}}}_{v} \frac{(X - \mathbb{E}X)}{\sqrt{DX}} + \underbrace{\frac{a\mathbb{E}X + b - \mathbb{E}Y}{\sqrt{DY}}}_{w}$$

$\bar{X}\bar{Y} = v\bar{X}^2 + w\bar{X} \quad / \mathbb{E}$

$\rho(X, Y) = \mathbb{E}\bar{X}\bar{Y} = v\,\mathbb{E}\bar{X}^2 = v$

$\rho(X, Y) = v$

как да намерим v?

$\bar{Y} = v\bar{X} + w \quad / D$

$1 = D\bar{Y} = D(v\bar{X} + w)$
$= v^2 \cdot 1 \qquad Dw = 0,\ w\text{-const}$

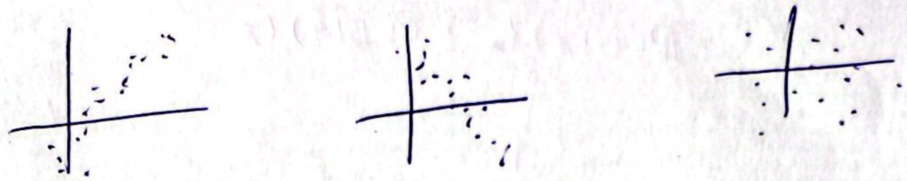
"$\Rightarrow$" $\rho(X, Y) = \mathbb{E}\bar{X}\bar{Y} = 1$

$0 \leq \mathbb{E}(\bar{X} - \bar{Y})^2 = \mathbb{E}\bar{X}^2 - 2\mathbb{E}\bar{X}\bar{Y} + \mathbb{E}\bar{Y}^2$
$= 2 - 2 \cdot 1 = 0$
$\Rightarrow \mathbb{E}(\bar{X} - \bar{Y})^2 = 0$
$\Rightarrow \bar{X} - \bar{Y} = 0$
$\Rightarrow \bar{X} = \bar{Y}$

$$\frac{Y - \mathbb{E}Y}{\sqrt{DY}} = \frac{X - \mathbb{E}X}{\sqrt{DX}}$$

$$\Rightarrow Y = \underbrace{\frac{1 \cdot X}{\sqrt{DX}}}_{a} + \underbrace{\frac{\mathbb{E}Y}{\sqrt{DY}} - \frac{\mathbb{E}X}{\sqrt{DX}}}_{b}$$

$\rho(X,Y)\approx 1$ $\qquad$ $\rho(X,Y)\approx -1$ $\qquad$ $\rho(X,Y)\approx 0$



! ако $\rho = 0$, това не означава че са $\perp\!\!\!\perp$, а линейно $\perp\!\!\!\perp$

$\rho(X,Y)\le 1$

$cov(X,Y)\le \sqrt{DX}\sqrt{DY}$

$|\langle a,b\rangle| \le \|a\|\,\|b\|$ .. Шварц

Условно мат. очакване

X, то $\mathbb{E}X$ се характ. с това, че $\min_{a\in\mathbb{R}} \mathbb{E}(X-a)^2$

$= DX = \mathbb{E}(X-\mathbb{E}X)^2$

$$\min_{a\in\mathbb{R}} \sum_j (x_j-a)^2 p_j$$

$\rho(X,Y)=\pm 1$ $\qquad$ $X = aY+b$

елем. събития

$\{Y=1\}=A$; $\{Y=0\}=\bar{A}$

⊕ $X = Ber(p)$; $Y = \begin{cases} 1 & p \\ 0 & 1-p=q \end{cases}$

$$\min_G \mathbb{E}(X-G(Y))^2 = \min_{a,b} \mathbb{E}(X-a1_A-b1_{\bar A})^2$$

$$= \min_{a,b\in\mathbb{R}} f(a,b)$$

как да намерим G?

$\cdot\ G(Y) = a1_A + b1_{\bar A} \overset{*}{=} \underbrace{\frac{\mathbb{E}X1_A}{\mathbb{P}(A)}1_A}_{\text{ако } Y=1,\text{ това число характ. } X} + \underbrace{\frac{\mathbb{E}1_{\bar A}}{\mathbb{P}(\bar A)}1_{\bar A}}_{\text{ако } Y=0,\text{ това число характ. } X}$

$a = G(1)$ | защото Y
$b = G(0)$ | приема само 1 и 0

$\cdot\ f(a,b) = \mathbb{E}(X^2 + a^2 1_A + b^2 1_{\bar A} - 2a1_AX - 2b1_{\bar A}X)$

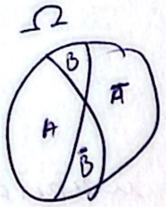
$= \mathbb{E}X^2 + a^2\mathbb{P}(A) + b^2\mathbb{P}(\bar A) - 2a\mathbb{E}1_AX - 2b\mathbb{E}1_{\bar A}X$

$$\left.\begin{aligned} 0 &= \frac{\partial f}{\partial a} = 2a\mathbb{P}(A) - 2\mathbb{E}X1_A \Rightarrow a = \frac{\mathbb{E}X1_A}{\mathbb{P}(A)} \\ 0 &= \frac{\partial f}{\partial b} = 2b\mathbb{P}(\bar A) - 2\mathbb{E}X1_{\bar A} \Rightarrow b = \frac{\mathbb{E}X1_{\bar A}}{\mathbb{P}(\bar A)} \end{aligned}\right\} *$$

ако $X = 1_B$

$$G(Y) = \frac{\mathbb{E} 1_B 1_A}{\mathbb{P}(A)} 1_A + \frac{\mathbb{E} 1_B 1_{\bar{A}}}{\mathbb{P}(\bar{A})} 1_{\bar{A}}$$

$$= \mathbb{P}(B|A) 1_A + \mathbb{P}(B|\bar{A}) 1_{\bar{A}}$$



ако X също е бинарно => издираме условна вероятност

Деф. (УМО) / Нека X, Y са 2 условни сл. величини.

X, Y - сл. вел - $G(Y)$
G - ф-я
G^* - сл. вел

Тогава $\mathbb{E}(X|Y) = \underbrace{G^*(Y)}_{\text{сл. вел}}$, която минимизира

$$\min_G \mathbb{E}(X - G(Y))^2 = \mathbb{E}(X - \mathbb{E}(X|Y))^2 = \mathbb{E}(X - G^*(Y))^2$$

⊕

Y	y_1	y_2	...	y_k	...
$\mathbb{P}$	p_1	p_2	...	p_k	...

$\mathbb{P}(Y = y_k) = p_k > 0$

$$\mathbb{E}(X|Y) = \sum_k \frac{\mathbb{E}(X 1_{A_k})}{p_k} 1_{A_k} \quad ; \quad A_k = \{Y = y_k\}, \quad p_k = \mathbb{P}(A_k)$$

в част. случай:

⊕ $X = 1_B$ — бинарна сл. вел.

$$\mathbb{E}(X|Y) = \sum_k \frac{\mathbb{E} 1_B 1_{A_k}}{\mathbb{P}(A_k)} 1_{A_k}$$

$$= \sum_k \mathbb{P}(B|A_k) 1_{A_k}$$

⊕ $X = \sum_i x_i 1_{B_i}$ — дискретна сл. вел.

$$\mathbb{E}(X|Y) = \sum_k \frac{\sum_i x_i \mathbb{P}(X = x_i;\ Y = y_k)}{\mathbb{P}(Y = y_k)} 1_{A_k}$$

$$= \sum_k \underbrace{\left(\sum_i x_i \mathbb{P}(X = x_i | Y = y_k)\right)}_{\mathbb{E}(X|Y = y_k)} 1_{A_k}$$

$$\mathbb{E}(X 1_{A_k}) = \mathbb{E}\left(\sum_i x_i 1_{B_i}\right) 1_{A_k}$$
$$= \sum_i x_i \mathbb{E} 1_{B_i} 1_{A_k}$$
$$= \sum_i \mathbb{P}(B_i \cap A_k)$$
$$= \sum_i x_i \mathbb{P}(X = x_i;\ Y = y_k)$$

$$\mathbb{E}(X|Y) = \sum_k \mathbb{E}(X|Y = y_k) 1_{A_k}$$

условно очакване на X, при фиксирана Y

$\mathbb{E}(X\|Y)$	$\mathbb{E}(X\|Y = y_1)$	...	$\mathbb{E}(X\|Y = y_k)$	...
$\mathbb{P}$	p_1		p_k	

Деф. Нека X и Y са сл. величини, където Y е дискретна. Тогава условно очакване на X при зададена стойност на Y (да кажем $Y=y_k$) се разбира

$$\mathbb{E}(X \mid Y=y_k) = \sum x_i \mathbb{P}(X=x_i \mid Y=y_k),$$

когато X също е дискретна!

Лема Нека X и Y са сл. величини, където Y е дискретна. Тогава

а) ~~∃~~ ако Z е сл. вел., то $\mathbb{E}(aX+bZ \mid Y) = a\,\mathbb{E}(X \mid Y) + b\,\mathbb{E}(Z \mid Y)$

б) $X \perp\!\!\!\perp Y$, то $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$, защото ~~$Y$~~ не носи информация

в) $X = f(Y)$, то $\mathbb{E}(X \mid Y) = f(Y) = X$

г) $\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}X$

д) $\mathbb{E}(f(X,Y) \mid Y=y_k) = \mathbb{E} f(X, y_k)$ ако $X \perp\!\!\!\perp Y$

Д-во а) заместваме:

$$\mathbb{E}(aX+bZ \mid Y) = \sum_k \overbrace{\frac{\mathbb{E}(aX+bZ)1_{A_k}}{\mathbb{P}(A_k)}}^{\text{линейност}} 1_{A_k}$$

$$= \sum_k \frac{a\,\mathbb{E}X1_{A_k} + b\,\mathbb{E}Z1_{A_k}}{\mathbb{P}(A_k)} 1_{A_k}$$

$$= a\,\mathbb{E}(X \mid Y) + b\,\mathbb{E}(Z \mid Y)$$

б) допускаме, че X е дискретна

$$\mathbb{E}(X \mid Y=y_k) = \sum_i x_i \mathbb{P}(X=x_i \mid \cancel{Y=y_k}) \overset{X \perp\!\!\!\perp Y}{=} \mathbb{E}X$$

$$\mathbb{E}(X \mid Y) = \sum_k \mathbb{E}X 1_{A_k} = \mathbb{E}X \underbrace{\sum_k 1_{A_k}}_{=1} = \mathbb{E}X$$

1 ,ако $Y=y_k$
0, иначе

пълна група от събития

в) $$\mathbb{E}(X \mid Y) = \sum_k \frac{\mathbb{E}(f(Y)1_{A_k})}{\mathbb{P}(A_k)} 1_{A_k}$$

$$= \sum_k \frac{f(y_k)\cancel{\mathbb{E}1_{A_k}}}{\cancel{\mathbb{P}(A_k)}} 1_{A_k} = \sum_k f(y_k) 1_{A_k} = f(Y)$$

+) $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}\Big(\sum_k \underbrace{\frac{\mathbb{E} X 1_{A_k}}{\mathbb{P}(A_k)}}_{const} \underbrace{1_{A_k}}_{индикатор} \Big)$

очакването на инд-икатор е вероятността му

$$= \sum_k \frac{\mathbb{E} X 1_{A_k}}{\mathbb{P}(A_k)} \mathbb{E} 1_{A_k}$$

$$= \sum_k \mathbb{E}(X 1_{A_k})$$

$$= \mathbb{E} \sum_k X 1_{A_k}$$

$$= \mathbb{E} X \Big(\underbrace{\sum_k 1_{A_k}}_{=1} \Big)$$

$$= \cancel{\mathbb{E}(X)} = \mathbb{E} X$$

$$\mathbb{E}(X | Y = y_k) = \sum_i x_i \, \mathbb{P}(X = x_i | Y = y_k)$$

$X \mid Y=y_k$	x_1	...	x_i	...
$\mathbb{P}$	p_1	...	p_i	...

$$q_i = \mathbb{P}(X = x_i | Y = y_k)$$

Тв. $\sum_i \mathbb{P}(X = x_i | Y = y_k) = 1$

Д-во $\sum_i \frac{\mathbb{P}(X = x_i; Y = y_k)}{\mathbb{P}(Y = y_k)} \overset{?}{=} 1$

+)

Y \ X	0	1	2	Y
0	16/36	8/36	1/36	25/36
1	8/36	2/36	0	
2	1/36	0	0	
X	25/36	10/36	1/36	

$X \mid Y=0$	0	1	2
$\mathbb{P}$	16/25	8/25	1/25

+) $\frac{16}{36} : \frac{25}{36} = \frac{16}{\cancel{36}} \cdot \frac{\cancel{36}}{25} = 16/25$

+) $X \sim Ber(Y)$

$$Y = \begin{cases} 1/3, & p = 1/2 \\ 2/3, & 1-p = 1/2 \end{cases} \qquad X \perp\!\!\!\perp Y$$