Covarience and Correlation

A way to measure the relationship or dependence of two random veriables.

Expectation for two RVs.

Let X, Y two RVs.

What is $E\,g(X,Y) = ?$

Defined in terms of the joint distribution.

Discrete Case. X, Y discrete RVs with joint distribution $p(x,y)$, then

$$E\,g(X,Y) = \sum_x \sum_y g(x,y)\,p(x,y)$$

Continuous Case. X, Y are continuous RVs with joint pdf $f(x,y)$, then

$$E\,g(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,f(x,y)\,dx\,dy$$

Properties

① $E[c] = c$
↳ constant

② $E[c\, g(X,Y)] = c\, E[g(X,Y)]$

③ $g_1(X,Y)$, $g_2(X,Y)$, then.

$$E[g_1(X,Y) + g_2(X,Y)] = E[g_1(X,Y)] + E[g_2(X,Y)]$$

↳ implies $\underline{E(X+Y) = E(X)} + E(Y)$

Ex Roll two dice, X, Y. $p(x,y) = 1/36$

$$E(XY) = \sum_x \sum_y xy\, p(x,y)$$

$$= \sum_{i=1}^{6} \sum_{j=1}^{6} ij \; 1/36 = \sum_{i=1}^{6} \frac{i}{36} \sum_{j=1}^{6} j \qquad \frac{6\cdot 7}{2} = 3\cdot 7$$

$$= \frac{3\cdot 7\cdot 3\cdot 7}{36} = \frac{49}{4}$$

Ex $f(y_1, y_2) = \begin{cases} 3y_1 & 0 \le y_2 \le y_1 \le 1 \\ 0 & \text{otherwise} \end{cases}$

$$E(Y_1 Y_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2 \, f(y_1, y_2)\, dy_2\, dy_1$$

$$= \int_0^1 \left(\int_0^{y_1} y_1 y_2 \, 3y_1 \, dy_2 \right) dy_1$$

$$= \int_0^1 3y_1^2 \left(\tfrac{1}{2} y_2^2 \Big|_0^{y_1} \right) dy_1$$

$$= \int_0^1 \tfrac{3}{2} y_1^4 \, dy_1 = \tfrac{3}{10}$$

Relation to independence

Let X, Y be two independent RVs

then $g(X)$, $h(Y)$

$$\boxed{E\, g(X) h(Y) = \left(E\, g(X)\right)\left(E\, h(Y)\right)}$$

Note: This is if and only if

Proof Continuous Case.

$$\mathbb{E}\, g(X)h(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)\, f(x,y)\, dx\, dy.$$

Since X, Y are independent then.

$$\boxed{f(x,y) = f_X(x)\, f_Y(y).}$$

$$E\, g(X)h(Y) = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} g(x)h(y)\, f_X(x)\, f_Y(y)\, dx\right) dy$$

$$= \int_{-\infty}^{\infty} h(y) f_Y(y) \underbrace{\left(\int_{-\infty}^{\infty} g(x) f_X(x)\, dx\right)}_{E(g(X))} dy$$

$$= E(g(X)) \underbrace{\int_{-\infty}^{\infty} h(y) f_Y(y)\, dy}_{E(h(Y))}$$

$$= E(g(X))\, E(h(Y)).$$

QED $\square$

Covariance

How can we measure how dependent to RVs are?

Def X, Y be two RVs with

$$\mu_X = E(X) \quad , \quad \mu_Y = E(Y).$$

$$Cov(X, Y) := E(X - \mu_X)(Y - \mu_Y)$$

Properties

① $Cov(X, X) = Var(X)$

② $a, b, c, d \in \mathbb{R}$

$$Cov(aX + b, cY + d) = ac\, Cov(X, Y)$$

③ $Cov(X, Y) = Cov(Y, X)$ - symmetry.

④ $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$

- Bilinearity.

⑤ $\boxed{Cov(X, Y) = E(XY) - \mu_X \mu_Y}$ Know the proof.

⑥ If X, Y are independent, then

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

WARNING: $Cov(X, Y) = 0$

Does Not imply that

X, Y are independent.

Proof ⑤ $Cov(X, Y) = E(XY) - \mu_X \mu_Y$.

$$Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y)$$

$$= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y)$$

$$= E(XY) - \mu_X \underbrace{E(Y)}_{= \mu_Y} - \mu_Y \underbrace{E(X)}_{= \mu_X} + \mu_X \mu_Y$$

$$= E(XY) - \mu_X \mu_Y.$$

QED. ▷

Proof ⑥ If X, Y are indep. then $Cov(X,Y) = 0$.

$$Cov(X,Y) = E(XY) - \mu_X \mu_Y$$

$$= \underset{\mu_X}{\underset{\parallel}{E(X)}}\ \underset{\mu_Y}{\underset{\parallel}{E(Y)}} - \mu_X \mu_Y = 0$$

QED □.

Can $Cov(X,Y) = 0$ but X, Y be dependent?

Example

X	-2	-1	0	1	2
$p_X(x)$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$

$Y = X^2$

$Y \backslash X$	-2	-1	0	1	2	$p_Y(y)$
0	0	0	$1/5$	0	0	$1/5$
1	0	$1/5$	0	$1/5$	0	$2/5$
2	$1/5$	0	0	0	$1/5$	$2/5$
$p_X(x)$	$1/5$	$1/5$	$1/5$	$1/5$	$1/5$	1

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \underset{\overset{\shortparallel}{0}}{\mu_X}\,\mu_Y$$

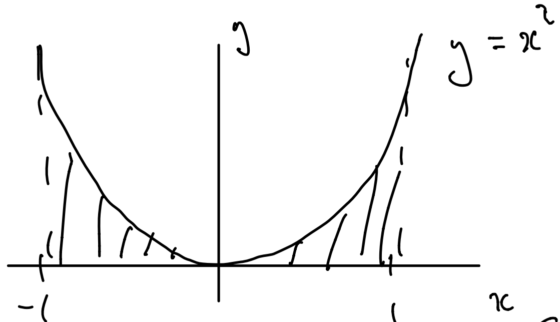
$$\mu_X = 0$$

$$= \mathbb{E}(X,Y) = \sum_x \sum_y xy\, p(x,y)$$

$$= \tfrac{1}{5}(-8 - 1 + 1 + 8) = 0$$

<u>Example</u> X, Y have joint pdf.

$$f(x,y) = \begin{cases} 2/3 & , -1 \le x \le 1, \quad 0 \le y \le x^2 \\ 0 & \text{otherwise} \end{cases}$$



$$\mu_X = \mathbb{E}(X) = \int_{-1}^{\overset{x}{1}} \left(\int_0^{x^2} x \tfrac{2}{3}\, dy \right) dx$$

$$= \int_{-1}^{1} \underbrace{\tfrac{2}{3} x^3}_{\text{odd}}\, dx = 0.$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \underbrace{\mu_X \mu_Y}_{=0}$$

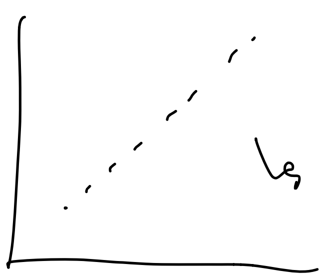
$$= \mathbb{E}(XY) = \int_{-1}^{1}\left(\int_0^{x^2} xy\,\tfrac{2}{3}\,dy\right)dx$$

$$= \int_{-1}^{1} \frac{2}{3}x\left(\tfrac{1}{2}y^2\right)\Big|_0^{x^2} dx$$

$$= \int_{-1}^{1} \underbrace{\tfrac{1}{3}x^5}_{\hookrightarrow \text{ odd}}\,dx = 0$$

Take away:

Covariance only measures linear dependence between RVs.



↳ strong correlation

→ zero covariance

Correlation

Since

$$\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X, Y).$$

- Covariance has units!!

Def

$$\mathrm{Cor}(X, Y) = \rho := \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

$$\sigma_X = \sqrt{\mathrm{Var}(X)}, \quad \sigma_Y = \sqrt{\mathrm{Var}(Y)}.$$

$$\mathrm{Cor}(aX, bY) = \frac{\mathrm{Cov}(aX, bY)}{\sqrt{\mathrm{Var}(aX)}\sqrt{\mathrm{Var}(bY)}}$$

$$= \frac{\cancel{ab}\,\mathrm{Cov}(X, Y)}{\cancel{ab}\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

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

Cor(X, Y) has no units!!

Properties

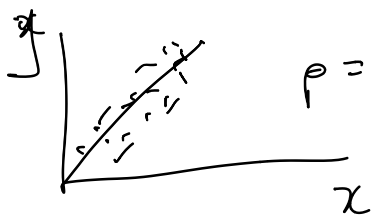
① $Cor(X, Y)$ is unitless.

② $-1 \le Cor(X, Y) \le 1$

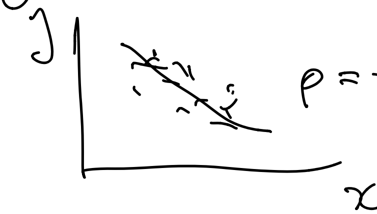
linear relationship

③ $Cor(X, Y) = 1$ if and only if $Y = aX + b$, $\underline{a > 0}$

$Cor(X, Y) = -1$ if and only if $Y = aX + b$, $a < 0$



$\rho = 1$



$\rho = -1$

Proof of ② (optional).

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

$$0 \le Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = Var\left(\frac{X}{\sigma_X}\right) + Var\left(\frac{Y}{\sigma_Y}\right) - 2\,Cov\left(X/\sigma_X, Y/\sigma_Y\right)$$

$$\Rightarrow 0 \le \underbrace{\frac{Var(X)}{\sigma_X^2}}_{=1} + \underbrace{\frac{Var(Y)}{\sigma_Y^2}}_{=1} - 2\underbrace{\frac{Cov(X,Y)}{\sigma_X, \sigma_Y}}_{=\rho}.$$

$$0 \le 2(1-\rho). \Rightarrow \boxed{\rho \le 1.}$$

$$0 \le Var\left(\frac{X}{\sigma_X} + Y/\sigma_Y\right) = 2 + 2\rho$$

$$\Rightarrow \boxed{\rho \ge -1.}$$