Covariance and Correlation

Two RVs X, Y independent.

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}[h(X)]\,\mathbb{E}[g(Y)].$$

Continuous $f_{XY}(x,y) = f_X(x)\, f_Y(y)$.

$$"\mathbb{E}[h(X)g(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)g(y)\, f_X(x)\, f_Y(y)\, dx\, dy$$

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

$$= \mathbb{E}h(X)\;\mathbb{E}g(Y)."$$

Law of iterated expectation.

$$E[h(X)g(Y)] = \mathbb{E}\{\mathbb{E}[h(X)g(Y) \mid Y]\}$$

$$= \mathbb{E}\{g(Y)\,\underbrace{\mathbb{E}[h(X) \mid Y]}_{= \mathbb{E}[h(X)] \text{ by independence.}}\}$$

$$= \mathbb{E}\{g(Y)\;\mathbb{E}[h(X)]\}$$

$$= \mathbb{E}[h(X)]\;\mathbb{E}[g(Y)].$$

Definition Let X, Y be two RVs.

Then the covariance is

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]$$

$\mu_X = \mathbb{E}X$, $\mu_Y = \mathbb{E}Y$.

Formula

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

Proof:

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

$$= \mathbb{E}\left[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y\right]$$

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

$$= \mathbb{E}XY - \mu_X\mu_Y.$$

Properties

① $Cov(X,X) = Var(X)$

② if X and Y are independent $Cov(X,Y) = 0$.

③ $Cov(X,Y) = Cov(Y,X)$ - symmetric

④ $Cov(aX+b, Y) = a\,Cov(X,Y) + \cancel{Cov(b,Y)}^{0}$

⑤ $Cov(X+Y, Z) = Cov(X,Z) + Cov(Y,Z)$

Ex X, Y jointly continuous

$$f_{XY}(x,y) = \begin{cases} 2 & x+y \le 1,\ x \ge 0,\ y \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

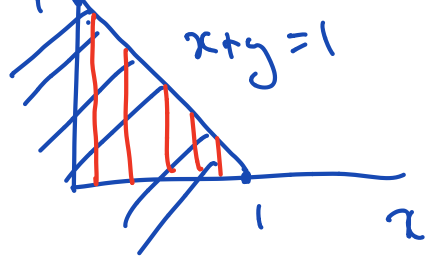
$Cov(X,Y)$?

$0 \le y \le 1-x$

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

$$\mu_X = \mathbb{E}X = \iint_R x f_{XY}(x,y)\, dy$$

$$= \int_0 \int_0 x\,2\,dy\,dx$$



$$= \int_0^1 x\,2(1-x)\,dx$$

$$= 1/3$$

$$\mathbb{E}Y = \iint_R y\, f_{XY}(x,y)\,dA$$

$$= \int_0^1 \int_0^{1-x} 2y\,dy\,dx = \int_0^1 (1-x)^2\,dx = 1/3.$$

$$\mathbb{E}XY = \iint_R xy\, f_{XY}(x,y)\,dA$$

$$= 2\int_0^1 x \int_0^{1-x} y\,dy\,dx = \int_0^1 x(1-x)^2\,dx$$

$$= \frac{1}{12}$$

$$\mathrm{Cov}(X,Y) = 1/12 - 1/3\,(1/3) = -\frac{1}{36}.$$

Warning If X, Y are independent.

$$\text{Cov}(X,Y) = 0.$$

The reverse is not true!!

Ex

X	-1,	0,	1
$P_X(x)$	$1/3$	$1/3$	$1/3$

$Y = X^2$

$$\text{Cov}(X,Y) = 0.$$

Ex

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

then $\text{Cov}(X,Y) = 0$.

- You get odd integrals.

Recall X, Y are independent. then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$$

Sum of two RVs

Let X, Y be two RVs

$$\boxed{Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)}$$

More generally

$X_1, X_2, \ldots, X_n$ RVs.

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{\substack{i=1,\, j=1 \\ i \neq j}}^{n} Cov(X_i, X_j)$$

Correlation.

Covariance depends on units!

$$Cov(aX, bY) = ab\, Cov(X,Y)$$

Remove units.

Def Correlation of X, Y.

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

rho.

Property that

$$Cor(aX, bY) = Cor(X,Y).$$

Properties.

① $-1 \le \rho_{XY} \le 1$

X, Y linearly dependent.

② $\rho_{XY} = 1$ iff $Y = aX + b$ $a > 0$

③ $\rho_{XY} = -1$ iff $X = aY + b$ $a > 0$.

Correlation only measure linear dependence.