Indep. random variables

X, Y are independent if

$\forall a,b \in \mathbb{R} \quad \mathbb{P}(X \le a, Y \le b) = \mathbb{P}(X \le a) \cdot \mathbb{P}(Y \le b)$

i.e. if $\{X \le a\}$ and $\{Y \le b\}$ are indep

i.e. if $F_{X,Y}(a,b) \le F_X(a) \cdot F_Y(b)$

Suppose X, Y are indep. Then

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

joint pdf

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

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

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

In particular, $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Def: The <u>covariance</u> between X and Y is

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

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

Formula:

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[XY - (\mathbb{E}X)Y - X\cdot\mathbb{E}(Y) + (\mathbb{E}X)\cdot(\mathbb{E}Y)\right]$$

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

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

Ex:

joint pmf

$Y \backslash X$	0	1
-1	0	1/4
1	1/2	1/4

$\mathbb{E}X = 1/2$

$\mathbb{E}Y = -1/4 + 3/4 = 1/2$

$\mathbb{E}[XY] = -1\cdot\frac{1}{4} + 1\cdot\frac{1}{4} = 0.$

$$\longrightarrow \quad \mathrm{Cov}(X,Y) = 0 - \frac{1}{2}\cdot\frac{1}{2} = \boxed{-\frac{1}{4}}.$$

[redacted]: (*) X, Y indep $\Rightarrow$ $\mathrm{Cov}(X,Y) = 0$

But not necessarily the other way:

Ex: $X \sim \mathrm{Unif}(-1,1)$, $Y = X^2$

$\mathbb{E}X = 0$

$\mathbb{E}(XY) = \mathbb{E}(X^3) = 0$

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

Interpretation: e.g. poker hands

(*) $X = \#$ spades, $Y = \#$ black cards — $\mathrm{Cov}(X,Y) > 0$ — X, Y tend to be large together or small together

(*) $X = \#$ spades, $Y = \#$ red cards — $\mathrm{Cov}(X,Y) < 0$ — X tends to be large when Y is small, and vice versa

Def: The correlation coefficient of X and Y is

rho $$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}$$ ← unitless

Thm (Cauchy-Schwarz inequality)

$$|\mathbb{E}(XY)|^2 \leq \mathbb{E}(X^2) \cdot \mathbb{E}(Y^2)$$

proof: Define $Z = X + \lambda \cdot Y$ for $\lambda \in \mathbb{R}$.

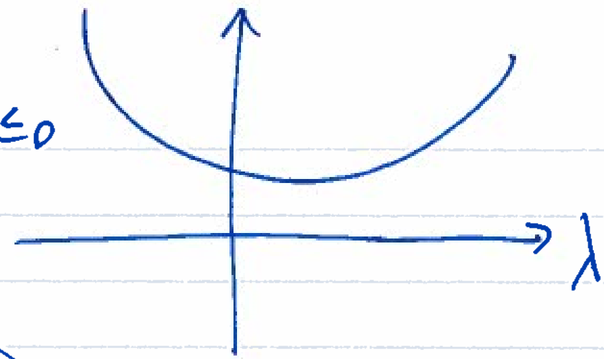
$$0 \leq \mathbb{E}(Z^2) = \mathbb{E}(X^2) + 2\lambda \cdot \mathbb{E}[XY] + \lambda^2 \cdot \mathbb{E}[Y^2]$$

Minimum when $\lambda = -\frac{2\mathbb{E}[XY]}{2\mathbb{E}(Y^2)}$

$$\left(2\mathbb{E}[XY]\right)^2 - 4\cdot\mathbb{E}(X^2)\cdot\mathbb{E}(Y^2) \le 0$$

$$\downarrow$$

$$\left(\mathbb{E}[XY]\right)^2 \le \mathbb{E}(X^2)\cdot\mathbb{E}(Y^2)$$



$$b^2 - 4ac \le 0$$

$$a\lambda^2 + b\lambda + c \ge 0$$

Apply this to correlation coefficient:

$$|\rho(X,Y)| = \frac{\left|\mathbb{E}\left[(X-\mathbb{E}X)\cdot(Y-\mathbb{E}Y)\right]\right|}{\sqrt{\mathbb{E}(X-\mathbb{E}X)^2\cdot\mathbb{E}(Y-\mathbb{E}Y)^2}} \le 1.$$

So $$-1 \le \rho(X,Y) \le 1.$$

Can show $$\rho(X,Y) = \begin{cases} 1 & \text{if and only if } Y = aX+b \text{ with } a>0 \\ -1 & \text{if and only if } Y = aX+b \text{ with } a<0. \end{cases}$$

Variance and independence:

(*) If X, Y are indep, then

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

<u>proof</u>:

$$Var(X+Y) = \mathbb{E}(X+Y)^2 - (\mathbb{E}(X+Y))^2$$

$$= \mathbb{E}X^2 + 2\mathbb{E}[XY] + \mathbb{E}Y^2 - (\mathbb{E}X)^2 - 2(\mathbb{E}X)\cdot(\mathbb{E}Y) - (\mathbb{E}Y)^2$$

extends to many variables

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

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

So if X, Y are indep $\Rightarrow$ $Cov(X,Y) = 0$ $\Rightarrow$

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

$\square$

<u>Ex</u>: $X \sim Bin(n,p)$ # successes in n <u>indep</u> Bernoulli trials with param p

$$X = X_1 + X_2 + \dots + X_n$$

Since $X_1, \dots, X_n$ are indep,

$$Var(X) = Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n) = n \cdot Var(X_1) = \boxed{np(1-p)}$$