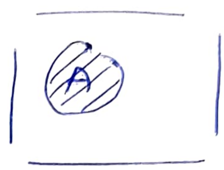
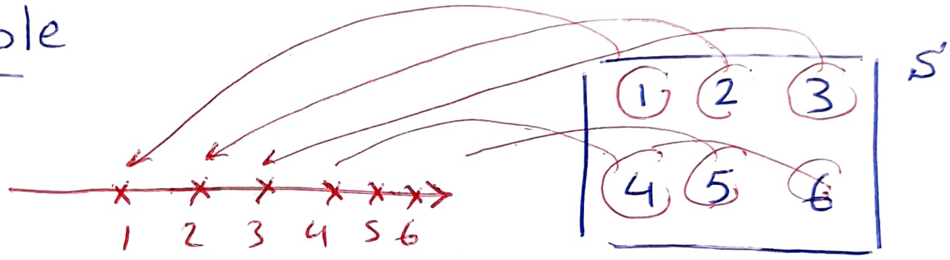
PASFML Mansoura AI45 20/11/2024 session 2.

- → Random Variable (Independence of R.V.s).
- → Discrete distributions (joint & marginal distributions)
- → Mean, Variance, covariance, correlation

$0 \le P(A) \le 1$ → Probability



Random Variable

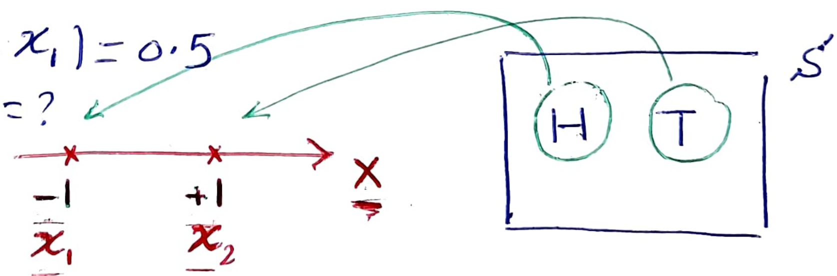


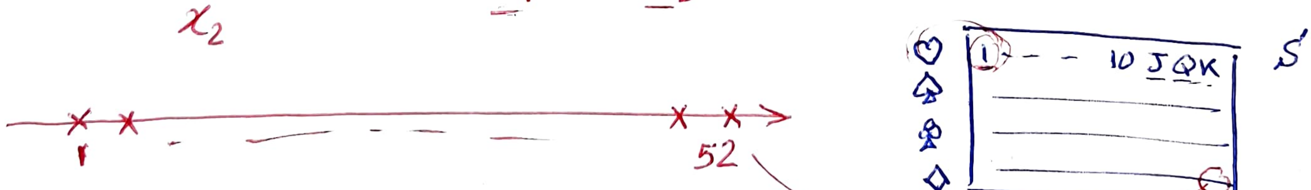
$P(H) = 0.5 \Leftrightarrow P(X = x_1) = 0.5$

$P(T)$ $\quad P_X(x_1) = ?$

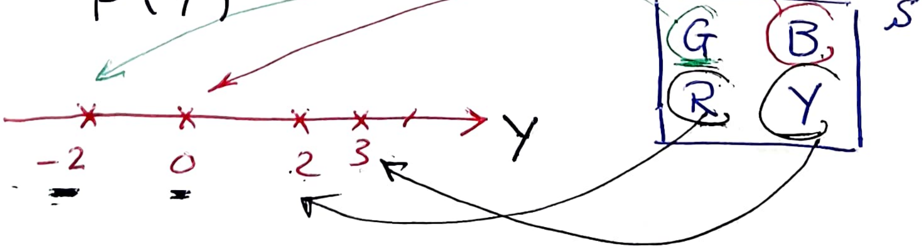
$P_X(-1) = P_X(+1) = 0.5$

$x_1 \quad x_2$





$P(G) \quad P(B) \quad P(Y)$



$P_Y(-2) = 1/4$

$P_Y(0) = 1/4$

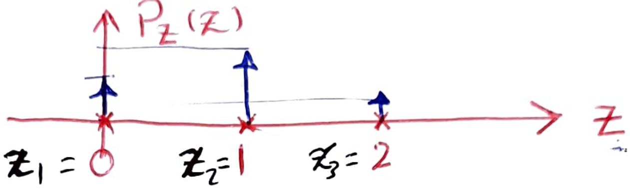
$\vdots$

ex Box : 3 Blue balls
5 Green balls
2 Yellow balls

$P(B) = 3/10$ $\quad P(Green) = \frac{5}{10}$ $\quad P(Y) = \frac{2}{10}$

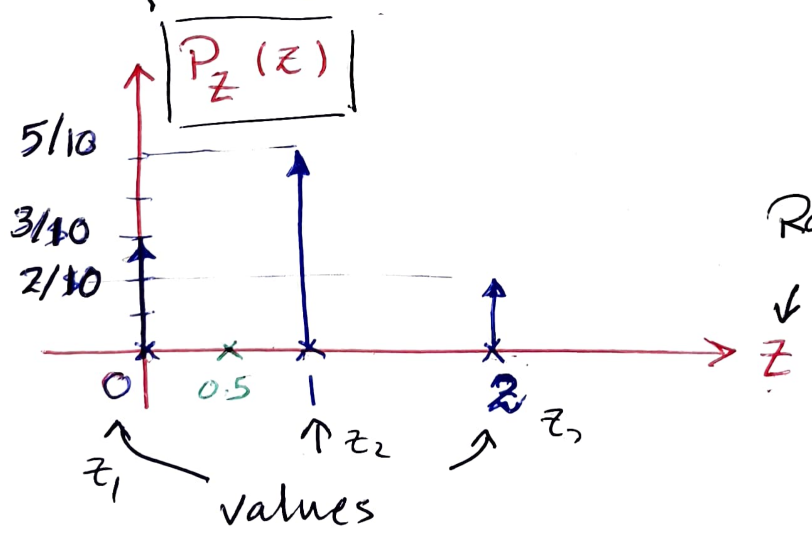
$P_Z(z_1) =$ $\quad P_Z(z_2) =$ $\quad P_Z(z_3) =$

$P_Z(0) = 3/10$ $\quad P_Z(1) = 5/10$ $\quad P_Z(2) = 2/10$



$Z = 0 \Leftrightarrow$ outcome is blue
$Z = 1 \Leftrightarrow$ " " Green
$Z = 2 \Leftrightarrow$ " " Yellow

Probability Mass Function (PMF)



Random Variable

$P_Z(0) = 0.3$

$P_Z(0.5) = 0$

$P_Z(z_1) = 3/10$ $\quad P_Z(z_2) = 5/10$ $\quad P_Z(z_3) = 2/10$

z	0	1	2
P(z)	0.3	0.5	0.2

$P_Z(z) =$ function of z ← session 3

Joint Probability

Joint PMF

Joint Distribution

$P(H_1, H_2)$

$P_{X,Y}(x_i, y_j)$

Joint ≡ تقاطع = and = $\cap$

X: -1, $+1$

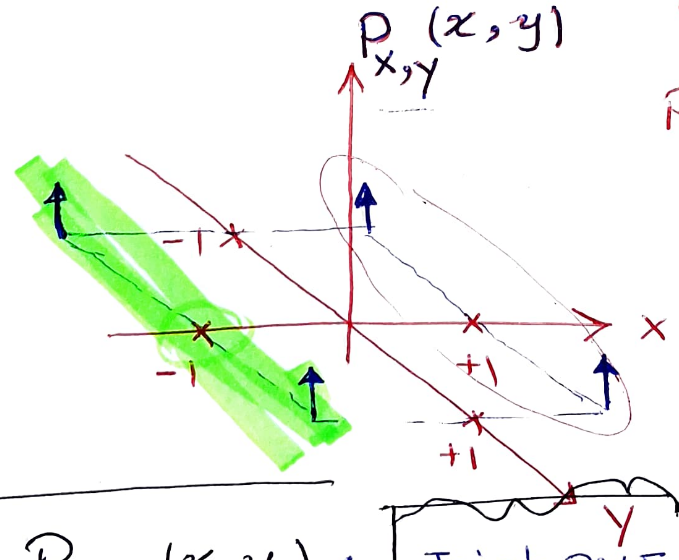
Y: -1, $+1$

$P_{X,Y}(-1,-1) = 1/4$

$P_{X,Y}(-1,+1) = 1/4$

$P_{X,Y}(+1,-1) = 1/4$

$P_{X,Y}(\cancel{+}-1,-1) = 1/4$



Y \ X	-1	+1
-1	1/4	1/4
+1	1/4	1/4

$P_{X,Y}(x,y)$

given $P_{X,Y}(x,y)$ ← Joint PMF

can we find $P_X(x)$ ← ? marginal PMF's

and $P_Y(y)$

$P_X(-1) = P_{X,Y}(-1,-1) + P_{X,Y}(-1,+1)$

or

$P_Y(y_1) = ?$

$P_Y(y_2) = ?$

$P_X(x_1) =$

$P_X(+1)$

given Joint PMF we can find marginal PMF's as follows:

$$\sum_j P_{X,Y}(x_i, y_j) = P_X(x_i)$$

$$\sum_i P_{X,Y}(x_i, y_j) = P_Y(y_j)$$

$$\sum_i \underbrace{\sum_j P_{X,Y}(x_i, y_j)}_{} = \underbrace{\sum_i P_X(x_i)}_{} = 1$$

Review Independence of events.

A, B are independent if

$$P(A/B) = P(A) \quad \text{or} \quad P(B/A) = P(B)$$

$$\boxed{\text{Indep.} \quad \Leftrightarrow \quad P(AB) = P(A) \cdot P(B)}$$

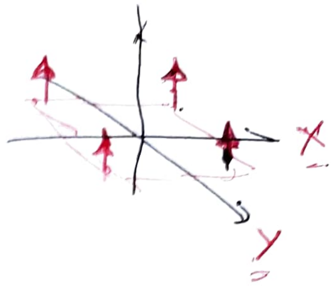
$$\text{Not Indep.} \quad \Leftrightarrow \quad P(AB) \neq P(A) \cdot P(B)$$

Two R.V.'s X, Y are independent if

$$\boxed{P_{X,Y}(x, y) = P_X(x) \cdot P_Y(y)}$$

ex. 1

$Y \backslash X$	-1	$+1$
-1	$1/4$	$1/4$
$+1$	$1/4$	$1/4$



$P_X(-1) = 1/2$ $\qquad$ $P_Y(-1) = 1/2$

$P_X(+1) = 1/2$ $\qquad$ $P_Y(+1) = 1/2$

are X, Y indep. or not ?

$$P_{X,Y}(x_i, y_j) \stackrel{?}{=} P_X(x_i) \cdot P_Y(y_j)$$

$$1/4 = P_{X,Y}(-1, -1) = 1/2 \times 1/2 . \checkmark$$

$\vdots$ $\qquad\qquad \checkmark$

$\Rightarrow$ X, Y are indep. ※

ex. 2

$Y \backslash X$	-1	$+1$
-1	$1/8$	$1/4$
$+1$	$1/8$	$1/2$

$P_{X,Y}(x_i, y_j)$

$P_X(x_i)$;

$P_X(-1) = 1/4$

$P_X(+1) = 3/4$

$P_Y(y_j)$;

$P_Y(-1) = 3/8$

$P_Y(+1) = 5/8$

are X, Y indep. or not ?

$P_{X,Y}(x_i, y_j) \overset{?}{=} P_X(x_i) \cdot P_Y(y_j)$

$1/8 = P_{X,Y}(-1,-1) \neq P_X(-1) \cdot P_Y(-1) = 1/4 \times 3/8 = 3/32$

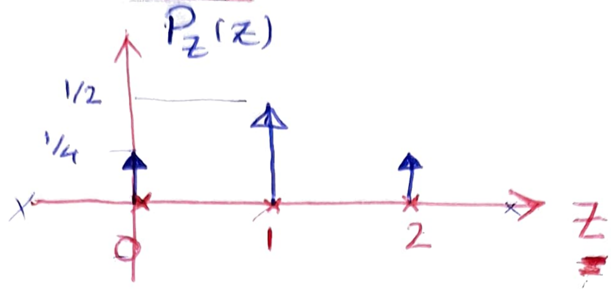
$\Rightarrow$ X, Y are not independent. *

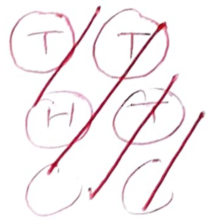
session 5 $\rightsquigarrow$ $P_{X/Y}(x_i / y_j) \neq P_X(x_i)$ $\qquad$ $P_{Y/X}(y_j / x_i) \neq P_Y(y_j)$

$$\underbrace{P_{X/Y}(x_i / y_j)}_{\text{Conditional PMF}} = \underbrace{\boxed{P_{X,Y}(x_i, y_j)}}_{\text{Joint PMF}} \div \underbrace{P_Y(y_j)}_{\text{Marginal PMF}}$$

ex

→ tossing two coins observing total number of heads.





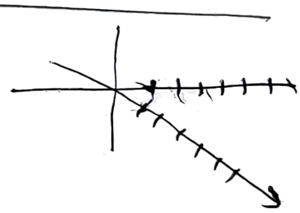
$\cancel{P\binom{n}{k} = \binom{2}{k}}$ $\quad P_Z(k) = \binom{n}{k}(1/2)^k(1/2)^{n-k}$

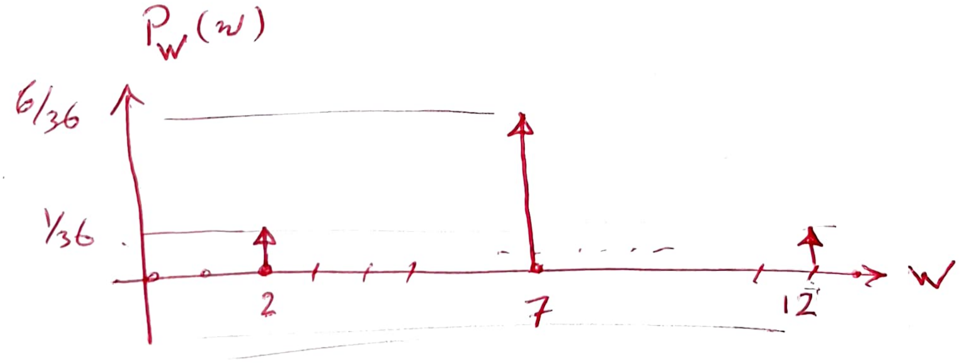
Binomial dist.

$$\begin{cases} P_Z(0) = \binom{2}{0}(1/2)^0(1/2)^2 = 1/4 \\ P_Z(1) = \binom{2}{1}(1/2)^1(1/2)^1 = 1/2 \\ P_Z(2) = \binom{2}{2}(1/2)^2(1/2)^0 = 1/4 \end{cases}$$

---

ex two dice are thrown observing sum.





→ Statistical tools

→ Moments

→ mean, variance, correlation, covariance, standard deviation, correlation coefficient

covariance matrix

→ mean, average, expected value, expectation, estimated value;

ex.; X : 3, 4, 4, 5, 5

$$m_X \equiv \mu_X \equiv \bar{X} \equiv E[X] \equiv \langle X \rangle$$

$$\bar{X} = \frac{3+4+4+5+5}{5} = 3 \times \frac{1}{5} + 4 \times \frac{2}{5} + 5 \times \frac{2}{5}$$

$$\bar{X} = \sum_i x_i \boxed{\frac{n(x_i)}{N}} \equiv P_X(x_i)$$

$$\boxed{\bar{X} = \sum_i x_i P_X(x_i)}$$

1^{st} moment of R.V. "X"

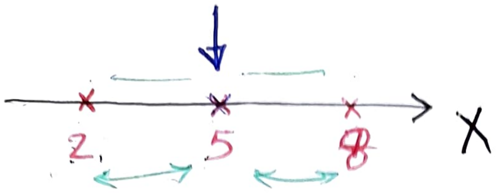
"weighted average"

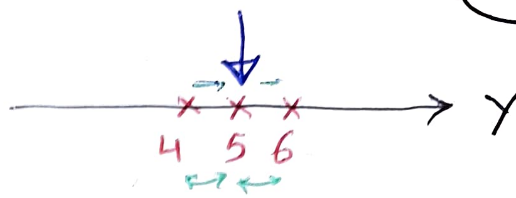
→ $\overline{g(X)} = \sum_i g(x_i) P_X(x_i)$ ← ✶

function

ex. X; 1, ..., 6

$Y = X^2$





$\bar{X} = \dfrac{2+5+8}{3} = 5$

$\bar{Y} = \dfrac{4+5+6}{3} = 5$

→ $\bar{X}$: measure of centrality.

we need other measures to describe distribution.

→ e.g., measures of dispertion around $\bar{X}$.

2nd central moment.

variance : σ_x^2

$\sqrt{\sigma_x^2} = \sigma_x$: standard deviation.

$$\sigma_x^2 = \overline{(x_i - \bar{x})^2}$$

$$\sigma_x^2 = \frac{(2-5)^2 + (5-5)^2 + (8-5)^2}{n}$$

$$= \frac{9+0+9}{3}$$

$$= 6$$

$$\sigma_x = \sqrt{6}$$

$$\sigma_y^2 = \overline{(y_i - \bar{x})^2}$$

$$\sigma_y^2 = \frac{(4-5)^2 + (5-5)^2 + (6-5)^2}{3}$$

$$= \frac{1+0+1}{3}$$

$$= 2/3 = 0.66$$

$$\sigma_y = \sqrt{2/3}$$

$\rightarrow \quad \sigma_x^2 = \sum_i (x_i - \bar{x})^2 P_x(x_i)$ ← ✶

$\overline{a+b+c} = \bar{a} + \bar{b} + \bar{c}$

$\rightarrow \quad \sigma_x^2 = \overline{(x_i - \bar{x})^2}$ ← ✶

$= \overline{(x_i^2 - 2x_i\bar{x} + \bar{x}^2)}$

$= \overline{x_i^2} - \overline{2x_i\bar{x}} + \overline{\bar{x}^2}$

$\overline{\bar{x}^2} = \mu_x^2 = \bar{x}^2$

$= \overline{x^2} - 2\bar{x}^2 + \bar{x}^2$

$2\bar{x}\bar{x} = 2\bar{x}^2$

mean square value

$\rightarrow \quad \sigma_x^2 = \overline{x^2} - \bar{x}^2$

✶

$$\begin{array}{c} x \\ y \end{array}\begin{bmatrix}10\\9\end{bmatrix} \quad \begin{bmatrix}8\\9\end{bmatrix} \quad \begin{bmatrix}7\\8\end{bmatrix}$$

$$\begin{bmatrix}150\\4\end{bmatrix} \quad \begin{bmatrix}120\\3\end{bmatrix} \quad \begin{bmatrix}80\\2\end{bmatrix}$$

$$\begin{bmatrix}175\\9\end{bmatrix} \quad \begin{bmatrix}160\\9\end{bmatrix} \quad \begin{bmatrix}180\\7\end{bmatrix}$$

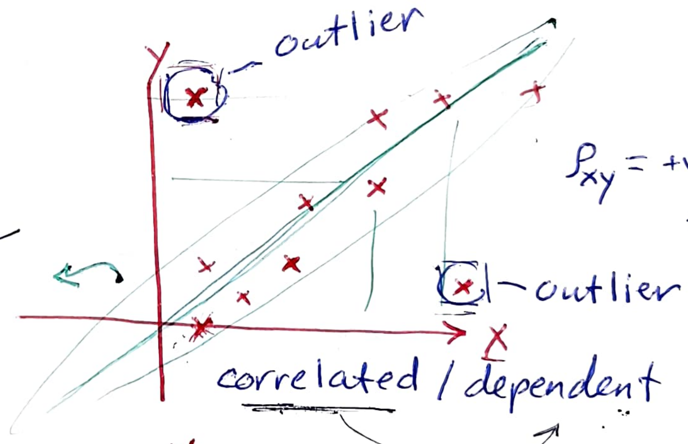
Correlation ≠ Dependence

l_1 l_2

↙

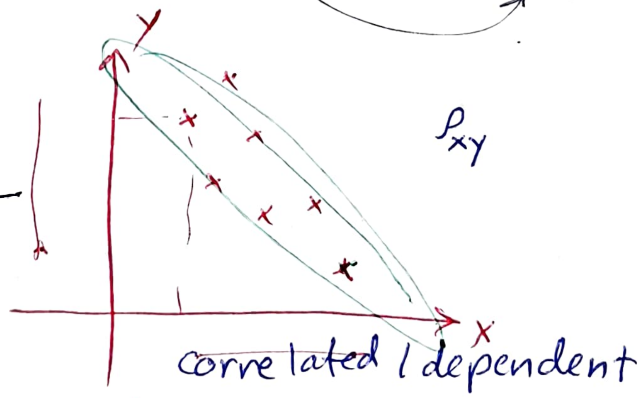
Linear correlation

positive correlation ←



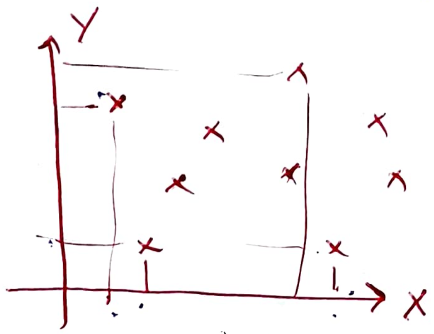
$\rho_{xy} = +ve \uparrow$
< 1

negative correlation ←



ρ_{xy}

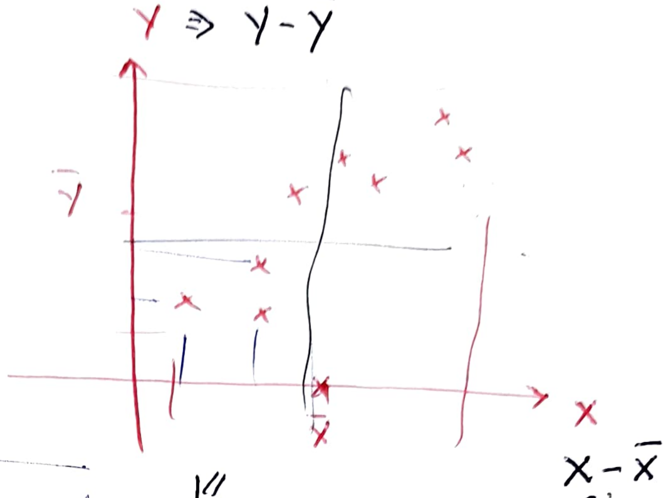
no correlation ←



$\rho_{xy} \approx 0$

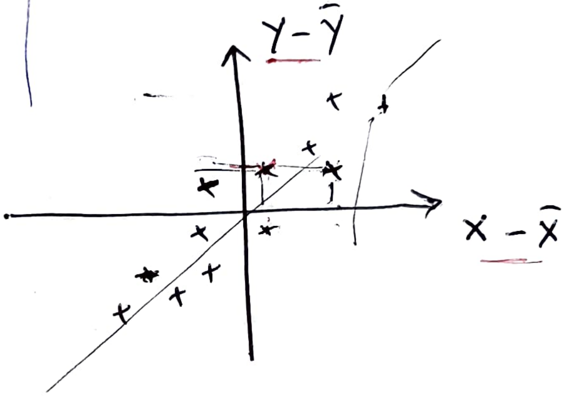
← Independend and uncorrelated

Covariance σ_{xy}



$$\sigma_{xy} = \frac{\sum_i (x_i-\bar{x})\cdot(y_i-\bar{y})}{n}$$

$$= \overline{(x_i-\bar{x})(y_i-\bar{y})}$$



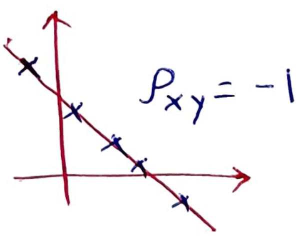
Variance $\sigma_x^2 \equiv \sigma_{xx}$

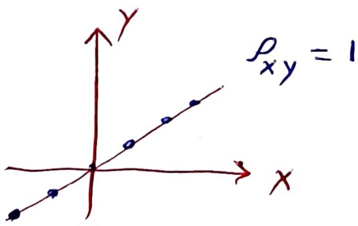
$$\sigma_{xx} = \overline{(x-\bar{x})(x-\bar{x})} \equiv \overline{(x-\bar{x})^2} = \sigma_x^2$$

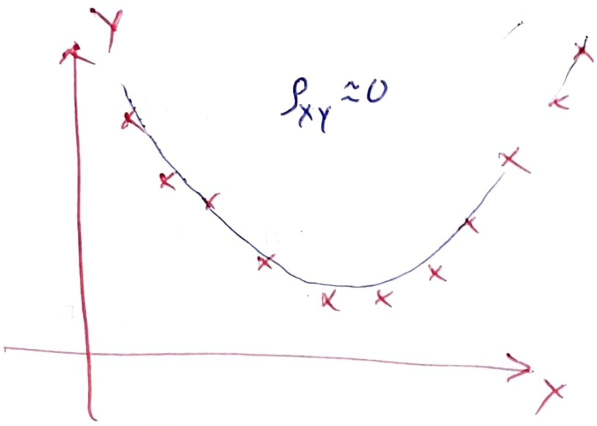
Correlation coefficient

(standardization of covariance)

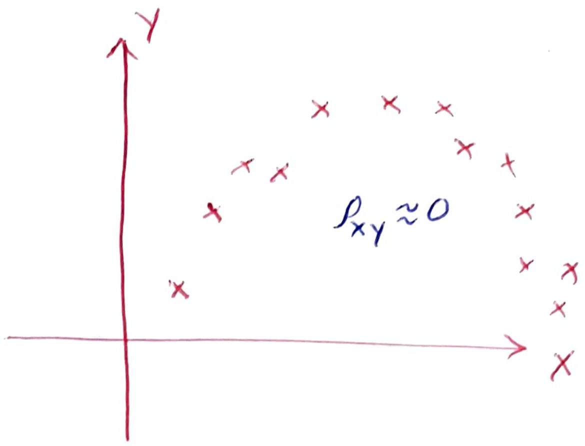
$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad -1 \le \rho_{xy} \le 1$$







e.g. = $Y = (X+b)^2$



X, Y are not independent
are ~~in~~ dependent
but not linearly correlated

X, Y are uncorrelated
but dependent

if X, Y are independent $\Rightarrow$ uncorrelated

if X, Y are uncorrelated → but not indep.
linearly
or → indep.

$$\sigma_{xy} = \overline{(x-\bar{x})(y-\bar{y})}$$

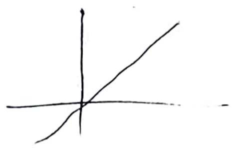
$$= \overline{xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y}}$$

$$= \overline{xy} - \overline{x\bar{y}} - \overline{\bar{x}y} + \overline{\bar{x}\bar{y}}$$

$$= \overline{xy} - \bar{x}\bar{y} - \cancel{\bar{x}\bar{y}} + \cancel{\bar{x}\bar{y}}$$

$$\sigma_{xy} = \overline{xy} - \bar{x}\bar{y}$$

if x, y are uncorrelated



$$\Rightarrow \quad \sigma_{xy} = 0 = \overline{xy} - \bar{x}\bar{y}$$

$$\Rightarrow \quad \boxed{\overline{xy} = \bar{x}\,\bar{y}}$$

if x, y are independent

$$\boxed{\underbrace{P_{xy}(x,y)}_{\text{Joint}} = P_x(x)\cdot P_y(y)}$$