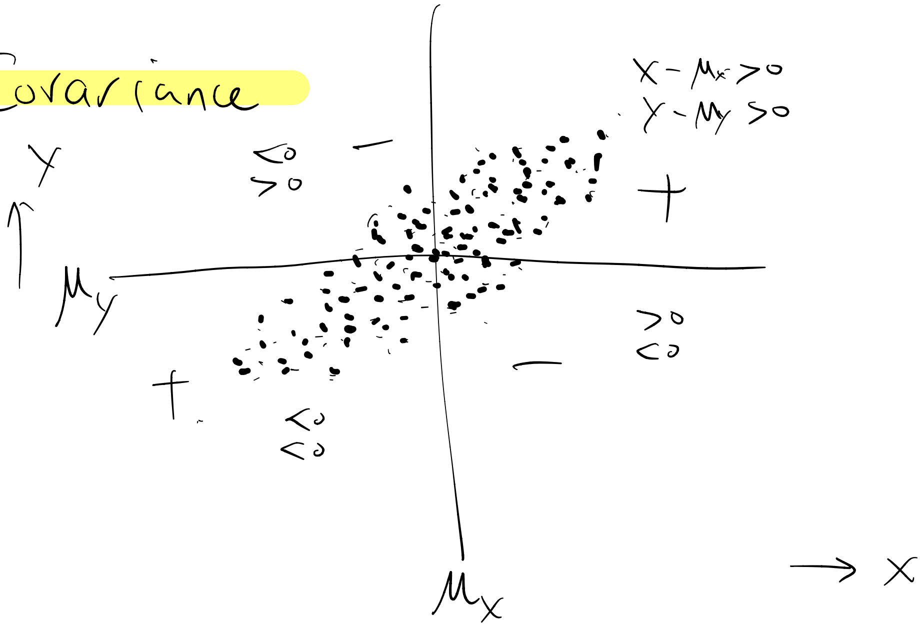
Covariance



$$\mu_X = E(X) \qquad \mu_Y = E(Y)$$

$$Cov(X,Y) = E\big((X - E(X))(Y - E(Y))\big)$$

$$Cov(X,X) = Var(X) = \sigma_X^2$$
$$Cov(Y,Y) = Var(Y) = \sigma_Y^2$$

$$Corr(X,Y) = Cov\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right)$$

$$= \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \rho$$ standardize

Covariance & Correlation

	X	Y	$\tilde{X}$	$\tilde{Y}$
1	x_1	y_1		
2	x_2	y_2		
$\vdots$	$\vdots$	$\vdots$		
i	x_i	y_i	$\tilde{x}_i = x_i - \mu_x$	$\tilde{y}_i = y_i - \mu_y$
$\vdots$	$\vdots$	$\vdots$		
n	x_n	y_n		

$$\mathrm{Cov}(X,Y) \simeq \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)$$

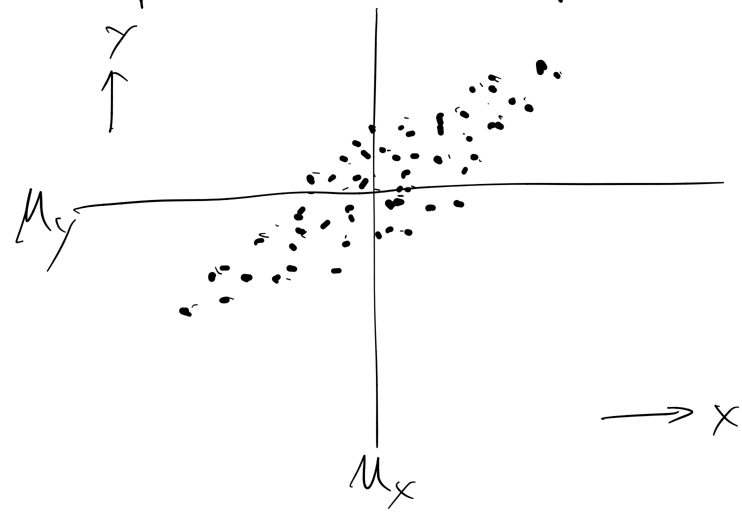
$$= \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\tilde{y}_i$$

$$= \frac{1}{n}\langle \tilde{X}, \tilde{Y} \rangle$$

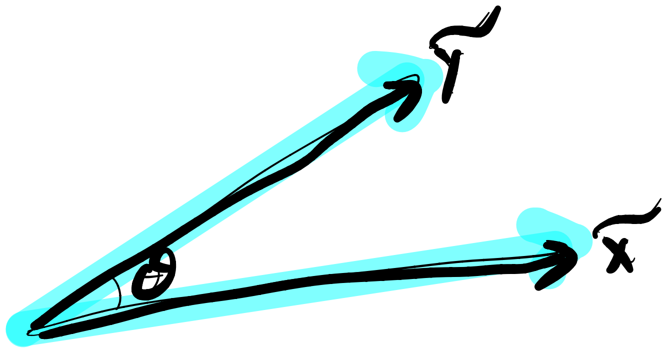
$$\mathrm{Var}(X) = \mathrm{Cov}(X,X) = \frac{1}{n}\langle \tilde{X}, \tilde{X} \rangle = \frac{1}{n}|\tilde{X}|^2$$

$$\mathrm{Var}(Y) = \quad \text{. - - - -} \quad = \frac{1}{n}|\tilde{Y}|^2$$

Scatter plot & vector plot



2-dimensional (row)



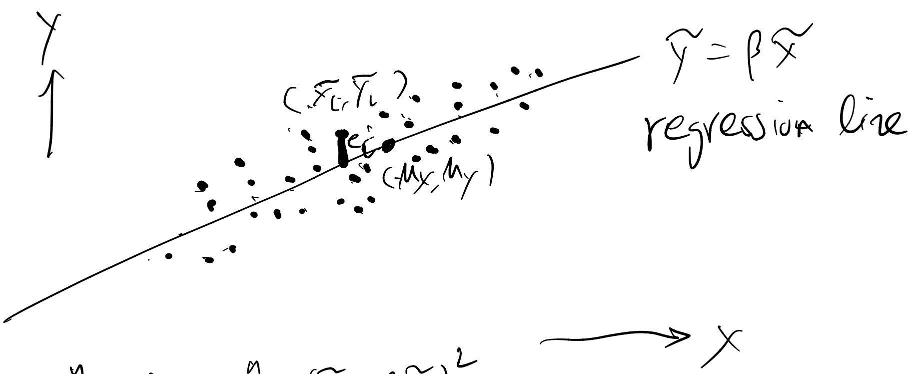
n-dim (column)

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} = \frac{\langle \tilde{X}, \tilde{Y} \rangle}{|\tilde{X}|\,|\tilde{Y}|}$$

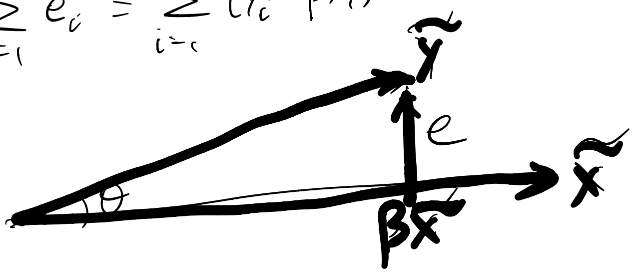
$$= \cos\theta$$

Correlation & regression

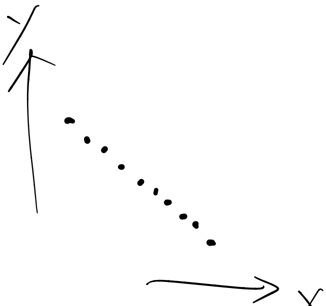
	$\tilde{x}$	$\tilde{y}$	$\beta\tilde{x}$	$e = \tilde{y} - \beta\tilde{x}$
1				
2				
⋮				
i	$\tilde{x}_i = x_i - \mu_x$	$\tilde{y}_i = y_i - \mu_y$	$\beta\tilde{x}_i$	$e_i = \tilde{y}_i - \beta\tilde{x}_i$
⋮				
n				



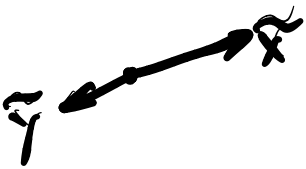
$$|e|^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (\tilde{y}_i - \beta\tilde{x}_i)^2$$



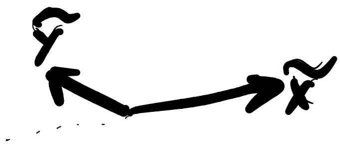
Scatter plot & Vector plot

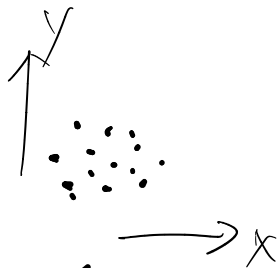


$\rho = -1$

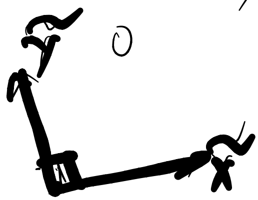


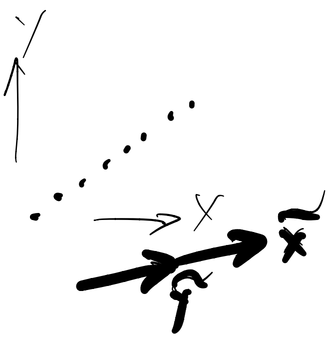
$-.5$



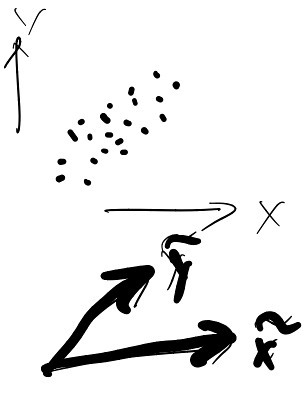


0



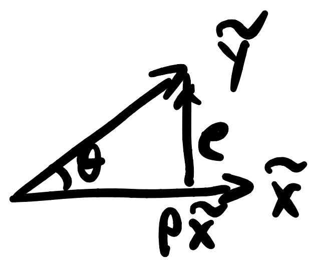


$\rho = 1$



$.5$

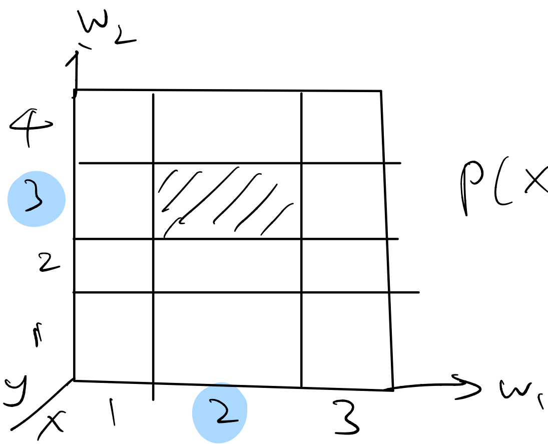
$$|\tilde{x}|^2 = |\tilde{y}|^2$$



$$\rho = \cos\theta < 1$$

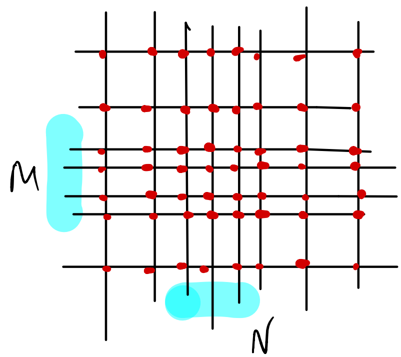
Independence & Covariance

discrete



$$p(x,y) = p(x)\,p(y)$$

MxN points



Continuous

$$f(x,y) = f(x)\,f(y)$$

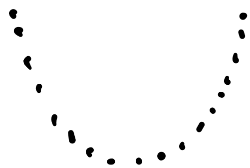
$$\mathrm{Cov}(X,Y) = E\big((X-\mu_x)(Y-\mu_y)\big)$$

$$= \iint (x-\mu_x)(y-\mu_y)\,f(x)\,f(y)\,dx\,dy$$

$$= \int (x-\mu_x)\,f(x)\,dx \int (y-\mu_y)\,f(y)\,dy$$

$$= E(X-\mu_x)\,E(Y-\mu_y) = 0$$

Conversely



$X \sim \mathrm{Unif}[-1, 1]$

$Y = X^2$

$$E((X-\mu_X)(Y-\mu_Y))$$

$$= E(X(Y-\mu_Y))$$

$$= E(X(X^2-\mu_Y))$$

$$= E(X^3) - E(X)\mu_Y = 0$$

independence $\longrightarrow$ cov $= 0$

X

Sum & average

$$E(X+Y) = \int (x+y) f(x,y)\,dx\,dy$$

$$= \int x f(x,y)\,dx\,dy + \int y f(x,y)\,dx\,dy$$

$$= E(X) + E(Y)$$

$$Var(X+Y) = E\left(\left((X+Y) - E(X+Y)\right)^2\right)$$

$$= E\left(\left((X+Y) - (E(X) + E(Y))\right)^2\right)$$

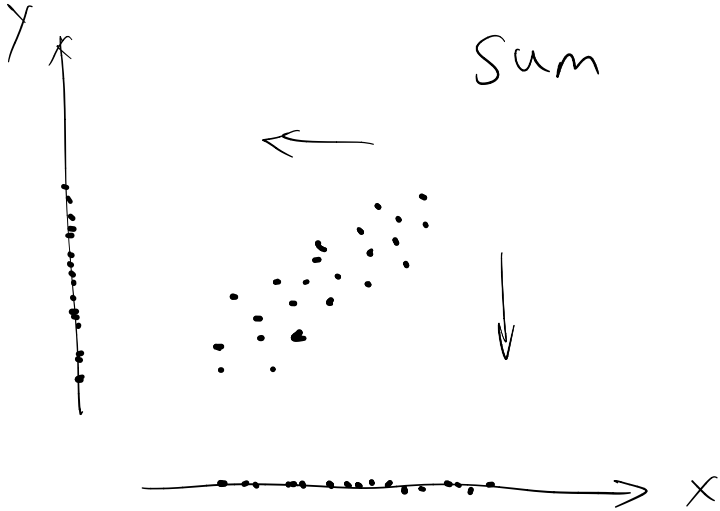
$$= E\left(\left((X - E(X)) + (Y - E(Y))\right)^2\right)$$

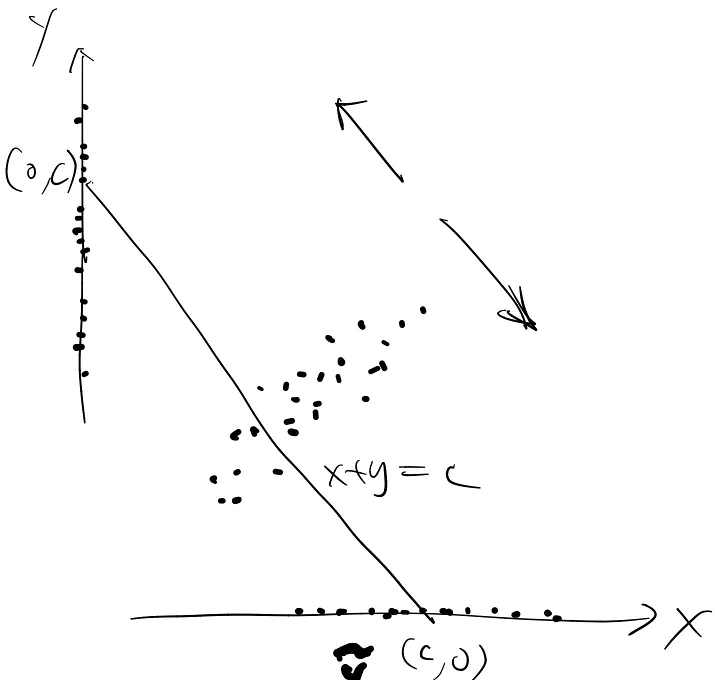
$$= E\Big((X-E(X))^2 + E(Y - E(Y))^2$$

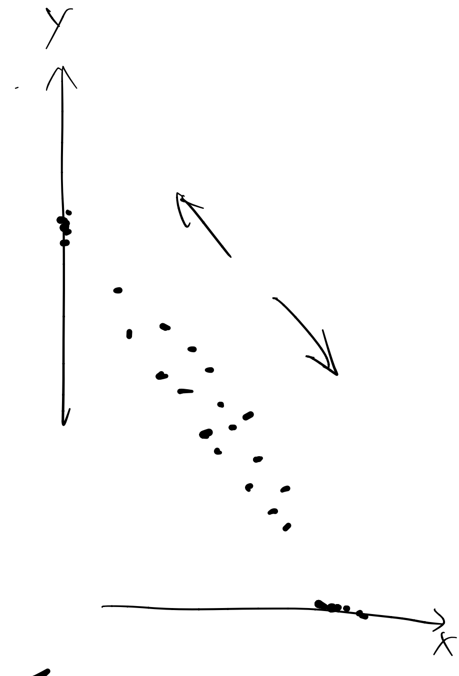
$$+ 2E((X-E(X))(Y-E(Y)))$$

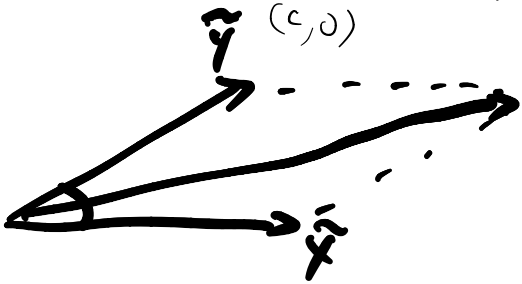
$$= Var(X) + Var(Y) + 2\,Cov(X,Y)$$

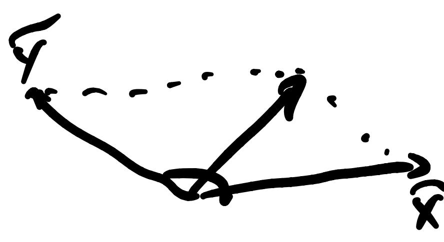
Sum





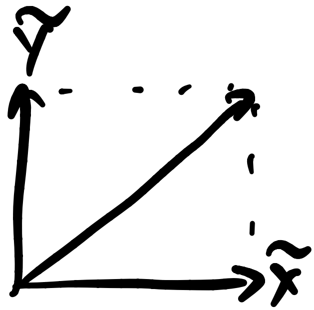






Independent $\rightarrow$ $Cov = 0$

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



$$X_1 \quad X_2 \quad \cdots \quad X_i \quad \cdots \quad X_n \overset{iid}{\sim} f(x)$$

iid : independent & identically distributed

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad E(X_i) = \mu$$

$$Var(X_i) = \sigma^2$$

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$$

$$Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Law of large number

$$P(|\bar{X} - \mu| > \varepsilon) \longrightarrow 0$$

Special case

prob = proportion

"wholesale"

(1) $X_i \sim \text{Bernoulli}(\frac{1}{2})$

$\mu = E(X_i) = \frac{1}{2}$ $\sigma^2 = Var(X_i) = \frac{1}{4}$

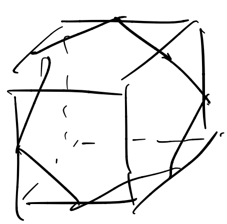
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\#\text{ of heads}}{n} = \text{freq of heads}$$

Among all 2^n seqs, proportion of "bad" ones $\longrightarrow 0$

(2) $X_i \sim \text{Unif}[0,1]$

$\mu = E(X_i) = \frac{1}{2}$ $\sigma^2 = Var(X_i) = \frac{1}{12}$

$$(X_1, X_2 \ldots X_n) \sim \text{Unif}[0,1]^n$$



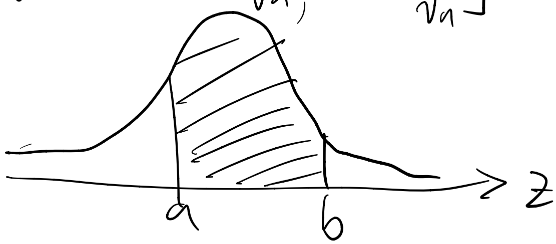
Volume of off-diagonal pieces $\longrightarrow 0$

Central limit theorem

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{var(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)$$

$$P(Z \in [a,b]) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$X \in [\mu + a\frac{\sigma}{\sqrt{n}}, \mu + b\frac{\sigma}{\sqrt{n}}]$$

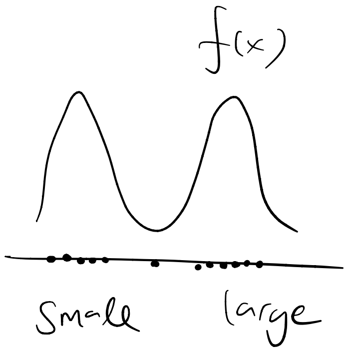


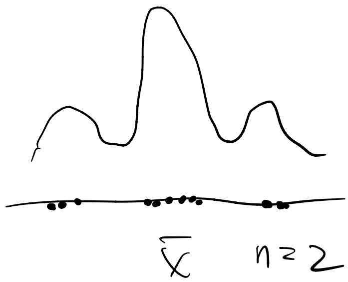
or $\quad \bar{X} \longrightarrow N(\mu, \frac{\sigma^2}{n})$

μ ↓ CLT

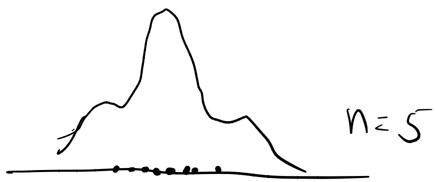
$\frac{\sigma^2}{n}$ ↘ LLN

Example



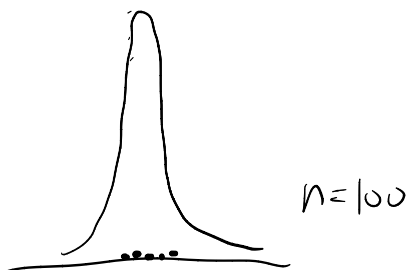


$$X_1 \quad X_2 \overset{iid}{\sim} f(x)$$

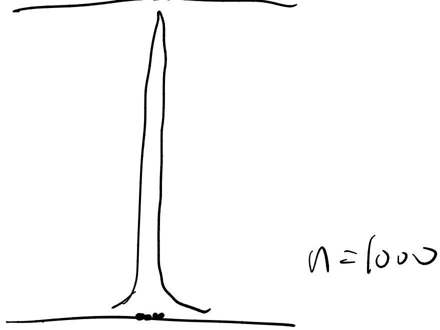


$$\bar{X} = \frac{X_1 + X_2}{2}$$

X_1 \ X_2	small	large
small	small	medium
large	medium	large



$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$



Summary

Basic concepts

1 random variable

2 random variables

Limiting theorems

Stochastic processes