10/20/22

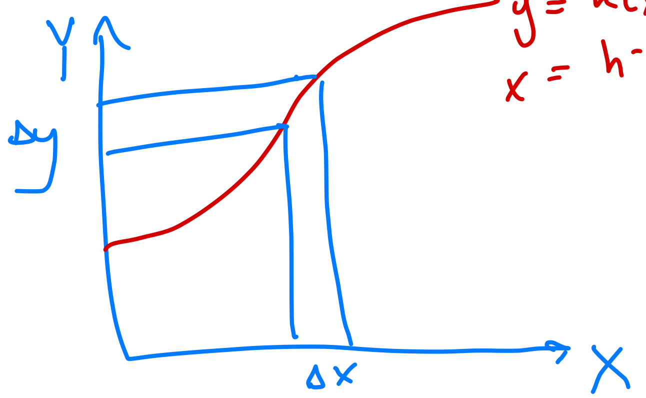
Last time:

Change of variable

substitution rule:

$$\int_a^b f(g(x))\, g'(x)\,dx = \int_c^d f(u)\,du$$

where $u = g(x)$



$$f_X(x)\,dx = f_Y(y)\,dy$$

Symbolically

$$X \sim f_X(x)\,dx = f_X(g(y))\,dg(y) = f_X(g(y))\,|g'(y)|\,dy$$

$$= f_Y(y)\,dy \sim Y$$ density

last time:

$$R \sim \underbrace{e^{-\frac{r^2}{2}} r\,dr}_{f_R(r)} = e^{-\frac{r^2}{2}}\, d\left(\frac{r^2}{2}\right) = \underbrace{e^{-t}\,dt}_{f_T(t)} = d\,e^{-t} = \underbrace{1\,du}_{f_U(u)}$$

($t = \frac{r^2}{2}$, $u = e^{-t}$)

$U \sim \mathrm{Unif}[0,1]$

$T = -\log U_1$

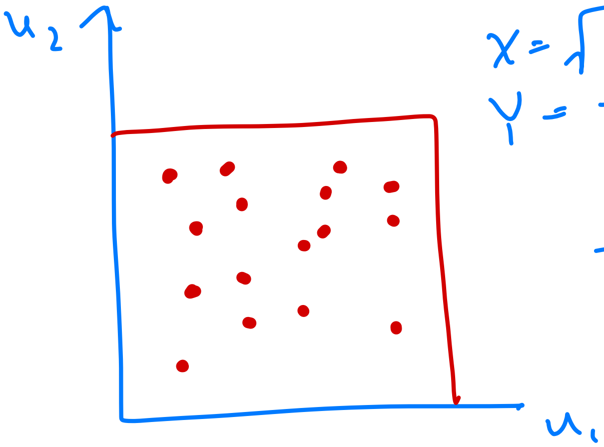
$R = \sqrt{-2T} = \sqrt{-2\log U_1}$

$\Theta \sim \frac{1}{2\pi}\,d\Theta = d\,\frac{\Theta}{2\pi} = du \quad ; \quad \Theta = 2\pi U_2$

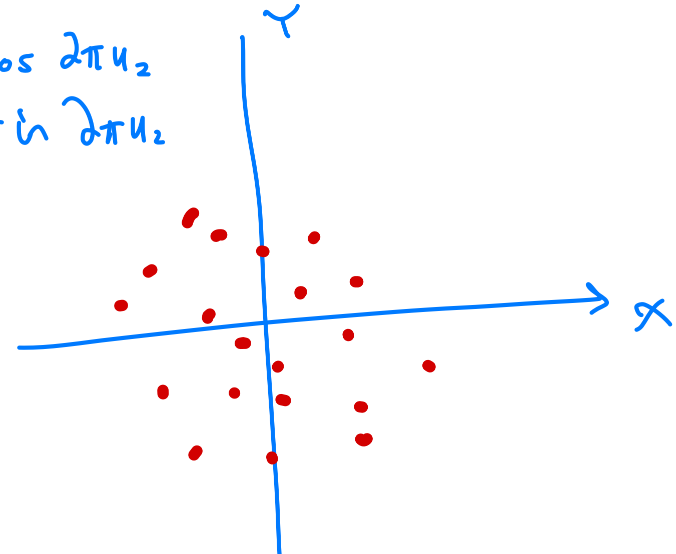
$X = R\cos\Theta$

$Y = R\sin\Theta$

$(X, Y) \sim N(0,1)$ ind.

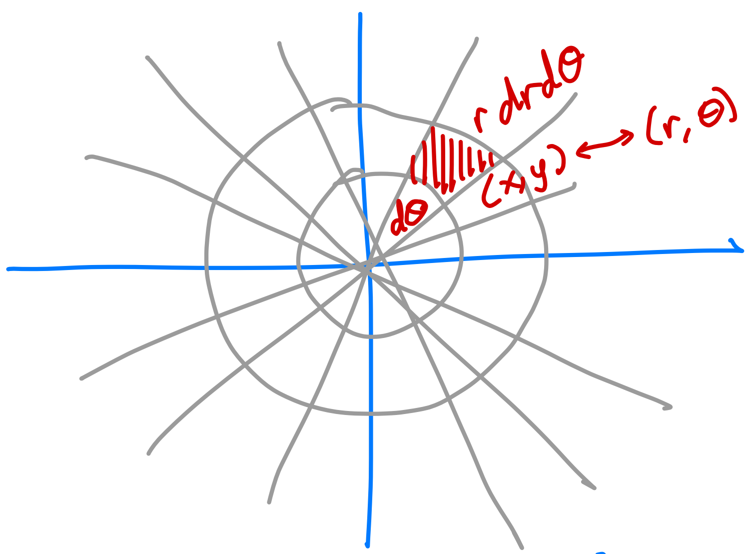


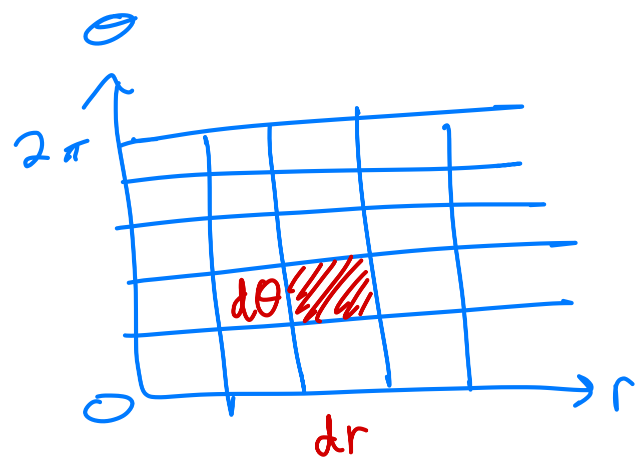
$$X = \sqrt{-2\log u_1}\, \cos 2\pi u_2$$
$$Y = \sqrt{-2\log u_1}\, \sin 2\pi u_2$$



2D transformation

Polar Method



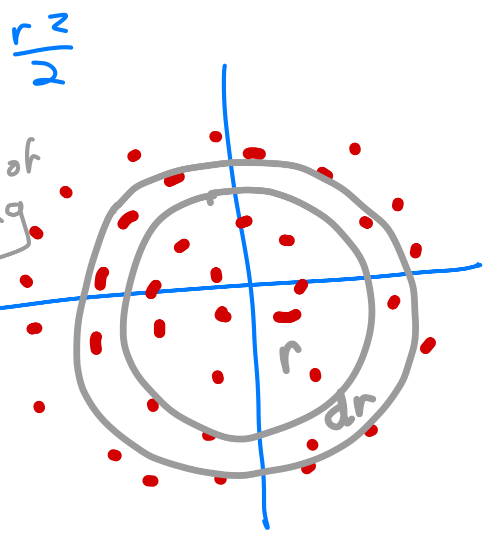


$$f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} = \frac{1}{2\pi} e^{-\frac{r^2}{2}}$$

area of ring

$$P(R \in (r, r+dr)) = \frac{1}{2\pi} e^{-\frac{r^2}{2}} \, 2\pi r\, dr$$

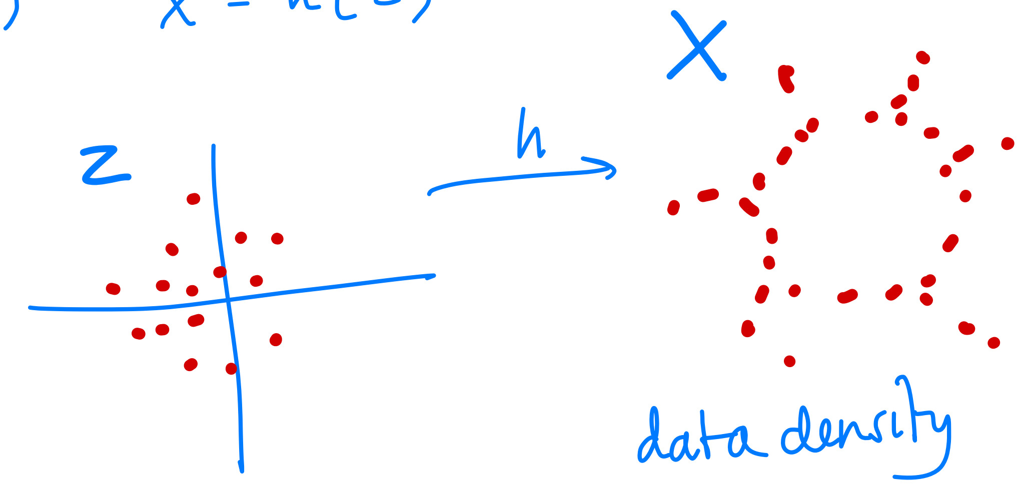
$$= e^{-\frac{r^2}{2}} r\, dr$$



Generative Modeling

$z \sim N(0,1)$ $\quad$ $x = h(z)$

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \\ \vdots \\ z_d \end{bmatrix} \sim N(0,1)$$



Multivariate Statistics

$X \sim f(x)$

could be vector of many elements

Properties of expectation

- $E(h(X)) = \int h(x) f(x)\, dx$
- $E(h(x) + g(x)) = E(h(x)) + E(g(x))$
- $E(a\, h(x)) = a\, E(h(x))$

$X_{m \times n}$ matrix

$$\begin{array}{c|c} & j \qquad n \\ \hline i & \begin{bmatrix} & & \\ & x_{ij} & \\ & & \end{bmatrix} \\ m & \end{array}$$

$$E(X) = \begin{bmatrix} E(x_{ij}) \end{bmatrix}$$

$$E(\underset{\text{constant matrix}}{A}X) = A\,E(X)$$

$$\begin{array}{c} \\ i \end{array} \underset{A}{\begin{bmatrix} a_{ik} \end{bmatrix}} \begin{array}{c} j \\ \begin{bmatrix} x_{kj} \end{bmatrix} \end{array} = \begin{array}{c} \\ i \end{array} \begin{array}{c} j \\ \begin{bmatrix} \sum_k a_{ik} x_{kj} \end{bmatrix} \end{array}$$

$$E\left(\sum_k a_{ik} x_{kj}\right) = \sum_k a_{ik} E(x_{kj})$$

expectation as long run average

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

$$\frac{1}{n}\sum_{i=1}^{n} A X_i = A \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E(AX) = A\,E(X)$$

$$E(XB) = E(X)B$$

In univariate case,

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

$$\mu = E(X)$$

In multivariate case,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_d \end{bmatrix} \qquad \mu = E(X) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_d \end{bmatrix}$$

$$Var(X) = E\left(\underbrace{(X-\mu)}_{d \times 1} \underbrace{(X-\mu)^T}_{1 \times d} \right) \quad = d \times d$$

$$= \begin{array}{c|c} & j \\ \hline k & E((X_k-\mu_k)^2) = Var(X_k) = Cov(X_k, X_k) \\ i & E((X_i-\mu_i)(X_j-\mu_j)) = Cov(X_i, X_j) \end{array}$$

diagonals are variance

$$Var(\underbrace{AX}_{d \times 1}) = E\left((AX - E(AX))(AX - E(AX))^T \right)$$

$$= E\left((AX - A\,E(X))(AX - A\,E(X))^T \right)$$

$$= E\left(A(X - E(X))(X - E(X))^T A^T \right)$$

$$= A \; Var(X) \; A^T$$

Agrees w/ univariate case: $Var(aX) = a^2 \, Var(X)$

2D

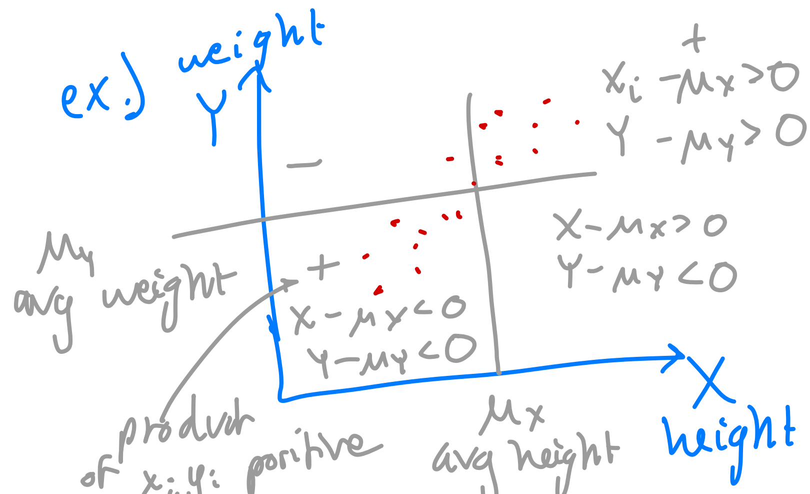
$$(X, Y) \sim f(x, y)$$

$$\begin{aligned}
\mathrm{cov}(X,Y) &= E((X - E(X))(Y - E(Y))) \\
&= E((X - \mu_X)(Y - \mu_Y)) \\
&= E(XY - \mu_X Y - X\mu_Y + \mu_X \mu_Y) \\
&= E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

parallel result:

$$\mathrm{Var}(X) = E(X^2) - E(X)^2$$

ex.)



scatterplot

cov > 0

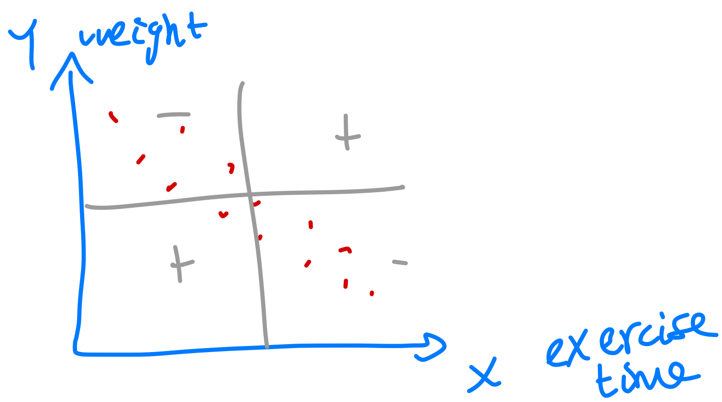
Data

	height X	weight Y
1		
2		
⋮	x_i	y_i
n		
⋮		

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

converge to pop. avg.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)(Y_i - \mu_Y)$$



cov < 0 $\rightarrow$ because more points in $-$ regions

$$\mathrm{Cov}(aX+b,\ cY+d)$$

$$= E\big((aX+b - E(aX+b)),\ (Y+d - E(cY+d))\big)$$

$$= E\big((aX+b - aE(X) - b),\ (cY+d - cE(Y) - d)\big)$$

$$= E\big(a(X-E(X))\ c\cdot(Y-E(Y))\big)$$

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

$$\mu_X = E(X) \qquad \sigma_X^2 = \mathrm{Var}(X)$$

$$\frac{X-\mu_X}{\sigma_X}$$

$$\mu_Y,\ \sigma_Y,\ \frac{y-\mu_Y}{\sigma_y}$$

$$\mathrm{Cov}\left(\frac{X-\mu_X}{\sigma_X},\ \frac{y-\mu_y}{\sigma_y}\right) = \frac{\mathrm{Cov}(x,y)}{\sigma_x\sigma_y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

$$= \mathrm{Corr}(X,Y) = \rho$$

measures strength of relationship between X&Y

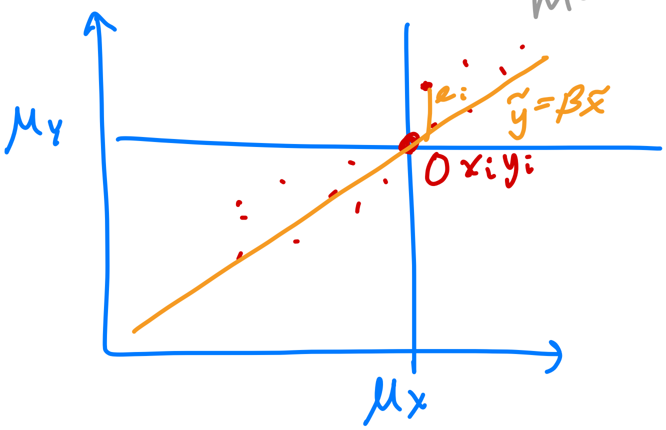
Data

	X	Y
1		
⋮		
i	x_i	y_i
⋮		
n		

$\xrightarrow{\text{centralize}}$

	$\tilde{X}$	$\tilde{Y}$
1		
⋮		
i	$\tilde{x}_i = x_i - \mu_X$	$\tilde{y}_i = y_i - \mu_Y$
⋮		
n		

2D scatterplot



move origin / center

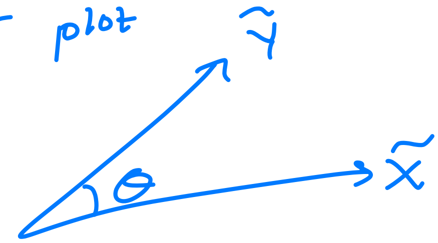
each point is a row

	$\tilde{X}$	$\tilde{Y}$
1		
⋮		
i	$\tilde{X}_i$	$\tilde{y}_i$
⋮		
n		

change point of view

columns are n-dim vector

vector plot



n-dim

$$\rho = \cos\theta$$

$$\downarrow$$

$$\text{corr}(X,Y) = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sqrt{\text{var}(X)}\sqrt{\text{Var}(Y)}}$$

$$E((X-\mu_X)(Y-\mu_Y)) = \frac{1}{n}\sum_{i=1}^{n}\tilde{X}_i\cdot\tilde{Y}_i = \frac{1}{n}\langle\tilde{X},\tilde{Y}\rangle$$

dot product

$$\text{var}(X) = E((X-\mu_X)^2) = \frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i^2 = \frac{1}{n}|\tilde{X}|^2 = \frac{1}{n}\langle\tilde{X},\tilde{X}\rangle$$

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

$$= \frac{\cancel{\frac{1}{n}}\langle\tilde{X},\tilde{Y}\rangle}{\cancel{\frac{1}{\sqrt{n}}}|\tilde{X}|\cancel{\frac{1}{\sqrt{n}}}|\tilde{Y}|} = \cos\theta$$

Regression

see scatterplot on prev page

$$\tilde{y}_i = \beta \tilde{x}_i + e_i$$

$$e = \tilde{y} - \beta \tilde{x}$$

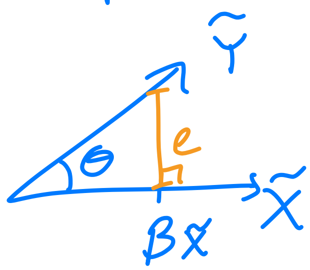
	$\tilde{X}$	$\tilde{Y}$	e_i
1			
⋮			
i	$\tilde{x}_i$	$\tilde{y}_i$	e_i
⋮			
n			

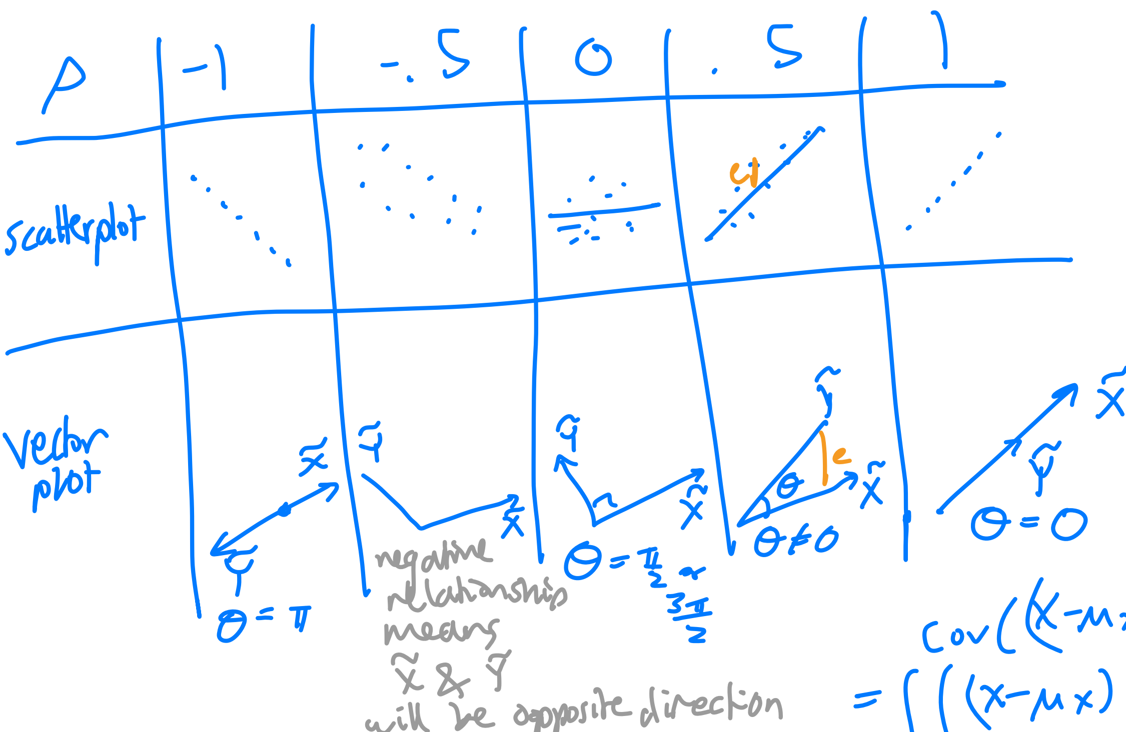
$$\frac{|e|^2}{|\tilde{Y}|^2} = \frac{\sum e_i^2}{\sum \tilde{y}_i^2} = \sin^2\theta$$

$$= 1 - \cos^2\theta$$

$$= 1 - \rho^2$$

if ρ^2 close to 1, then e_i^2 is close to 0





$$X \perp Y$$

$$f(x,y) = f_X(x) f_Y(y)$$

$$\operatorname{cov}((X-\mu_x)(Y-\mu_Y))$$

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

$$= \int (x-\mu_x) f_x(x)\, dx \int (y-\mu_Y) f_Y\, dy$$

$$= E(\underbrace{X - \mu_y}_{0})\, E(\underbrace{Y - \mu_x}_{0}) \rightarrow 0$$

If $X \perp Y \Rightarrow Cov = 0$

But if $Cov = 0 \not\Rightarrow X \perp Y$

$X \sim Unif[-1, 1]$

$Y = X^2$

$Cov(X, Y) = E(XY) - E(X)E(Y)$

$= E(X^3) - E(X)E(X^2) = 0$