28.11.2022

YZV 231E

Probability Theory & Stats

Week 10

Gü

Recap: Transform r.v.s to derive new distributions:

$$X \text{ r.v. } \xrightarrow{g(\cdot)} Y = g(X)$$

g: g^{-1} exists

one-to-one

many-to-one

$$2 \text{ r.v.s } (X, Y) \xrightarrow{g, h} (W, Z)$$

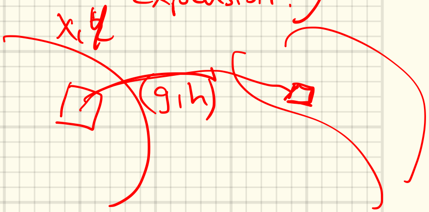
$$W = g(X, Y)$$
$$Z = h(X, Y)$$

$$P_{W,Z}(w,z) = P_{X,Y}\Big(\underline{g^{-1}(w,z)}, \underline{h^{-1}(w,z)}\Big) \left| \underbrace{\det\left(\frac{\partial(x,y)}{\partial(w,z)}\right)}_{\text{Jacobian of the xformation}} \right|$$

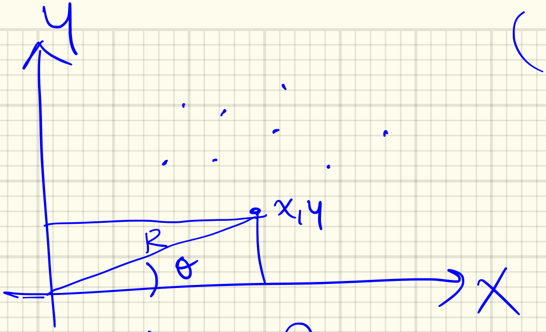
Change of Variables way

contraction

expansion.



Ex:



(X, Y) : Received signal coordinates in radar/sonar.

$$\left.\begin{aligned} X &\sim \mathcal{N}(0, \sigma^2) \\ Y &\sim \mathcal{N}(0, \sigma^2) \end{aligned}\right\} \; X \perp\!\!\!\perp Y \text{ are independent r.v.s.}$$

$$P_{R,\Theta}(r,\theta) = ?$$

R : magnitude

θ : angle

2 new r.v.s.

$$\begin{cases} R = \sqrt{X^2+Y^2}, & R \geq 0 \\ \theta = \operatorname{atan}\left(\frac{Y}{X}\right), & 0 \leq \theta \leq 2\pi \end{cases}$$

$X = R\cos\theta$

$Y = R\sin\theta$

$$P_{R,\theta}(r,\theta) = P_{X,Y}\left(g^{-1}(R,\theta), h^{-1}(R,\theta)\right) \underbrace{|\det J|}_{R}$$

$$J = \frac{\partial(X,Y)}{\partial(R,\theta)} = \underbrace{\begin{bmatrix} \frac{\partial X}{\partial R} & \frac{\partial X}{\partial \theta} \\ \frac{\partial Y}{\partial R} & \frac{\partial Y}{\partial \theta} \end{bmatrix}}_{} = \begin{bmatrix} \cos\theta & -R\sin\theta \\ \sin\theta & R\cos\theta \end{bmatrix} \rightarrow \det J = R \underset{?}{\geq} 0$$

exercise:

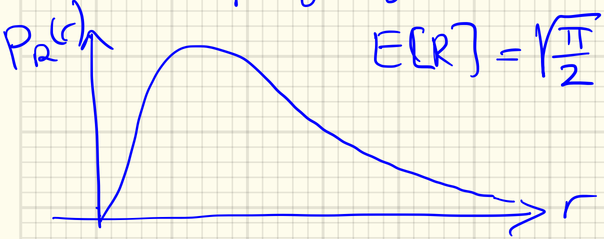
$$(X_1, X_2, \cdots X_n) \longrightarrow (Y_1, Y_2, \cdots Y_n) \rightarrow J = \begin{bmatrix} & ? & \end{bmatrix}_{n \times m}$$

$P_{X,Y}(x,y) = ?$ $P_X(x)$ $P_Y(y)$: joint pdf of X & Y

$$P_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)$$

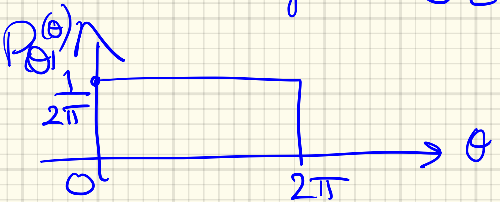
$$P_{R,\theta}(r,\theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\underbrace{(r^2\cos^2\theta + r^2\sin^2\theta)}_{r^2}} = \underbrace{\left(\frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}\right)}_{P_R(r)} \underbrace{\left(\frac{1}{2\pi}\right)}_{P_\Theta(\theta)}$$

R ~ Rayleigh pdf (σ^2)



$E[R] = \sqrt{\frac{\pi}{2}}\,\sigma$

Θ ~ uniform : $U[0, 2\pi]$



R, θ independent ✓

Recall: $Z = X + Y$, X & Y are independent.

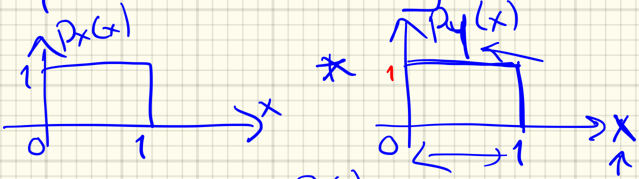
$$p_Z(z) = p_X(x) * p_Y(y)$$: convolution of the pdfs of X & Y.

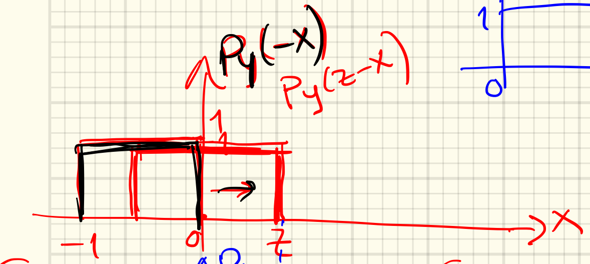
$$p_Z(z) \triangleq \int_{-\infty}^{\infty} p_X(x) \cdot p_Y(z-x)\, dx$$ Definition of convolution operation.

Mechanics of Convolution:

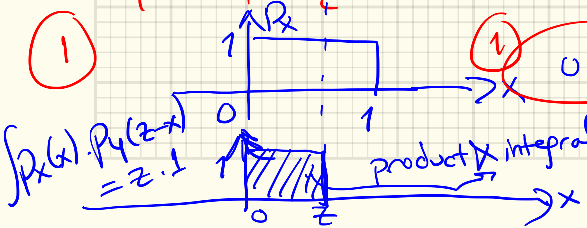
Ex: Let $X, Y \sim U[0,1]$ & indep: $Z = X+Y$

$p_Z(z) = ?$



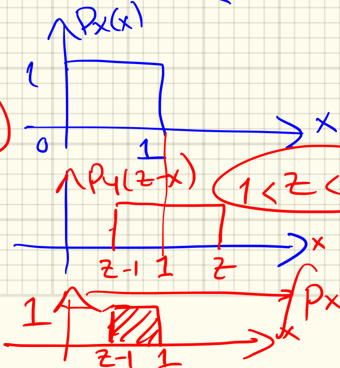


① ① $0 < z < 1$



$\int p_X(x) \cdot p_Y(z-x) = z \cdot 1$

② ② $1 < z < 2$



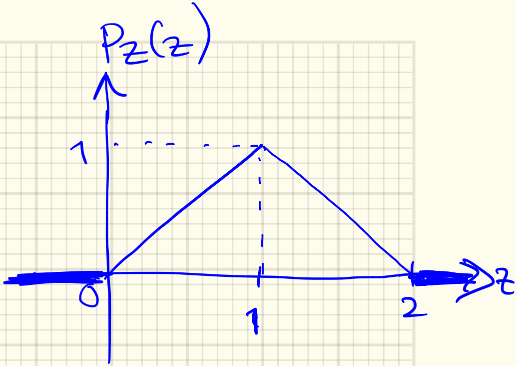
$\int p_X(x) \cdot p_Y(z-x) = 2-z$

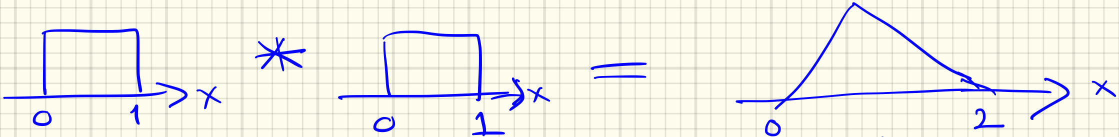
③ $z > 2$

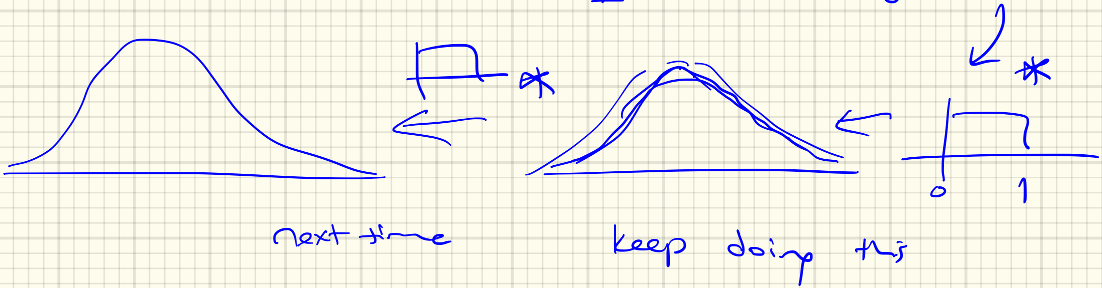
$\int p_X(x) p_Y(z-x)\, dx = 0$

no overlap

$$\rightarrow p_z(z) = \begin{cases} 0 \; , & z \le 0 \\ z \; , & 0 < z \le 1 \\ 2-z \; , & 1 < z \le 2 \\ 0 \; , & z > 2 \end{cases}$$



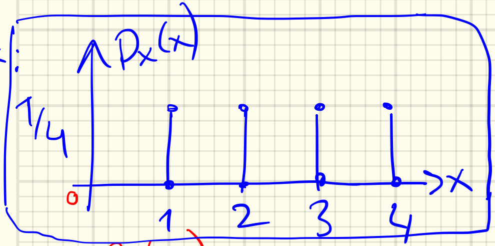




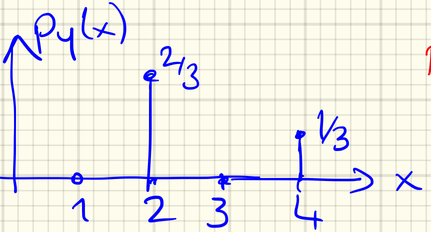
Example to Discrete Convolutions (Functions are discrete)

2 discrete r.v-s X & Y, independent: $Z = X+Y$, $P_Z(z) = ?$

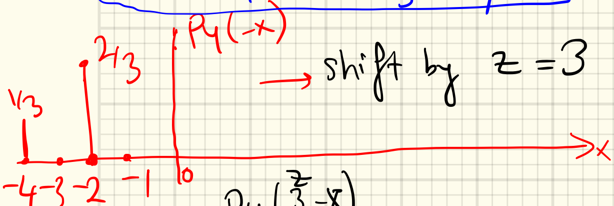
Ex:



$\ast$ convolve

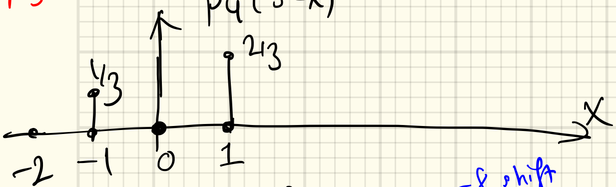


Flip $p_Y(y)$, shift it over $p_X(x)$ multiply them & add the result.



shift by $z=3$, then starts to overlap w/ $p_X(x)$.

$$P_Z(z) = \sum_x p_X(x) P_Y(z-x)$$

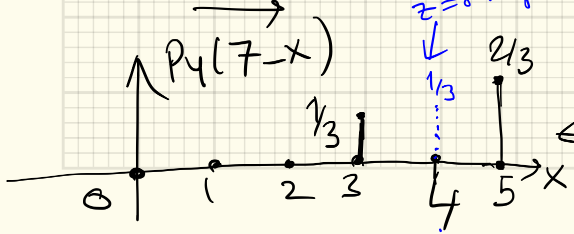


$$P_Z(z=3) = \sum_x p_X(x)\, P_Y(3-x) = \frac{1}{4}\cdot\frac{2}{3} = \frac{1}{6}$$

$$P_Z(z=4) = \frac{1}{6}$$

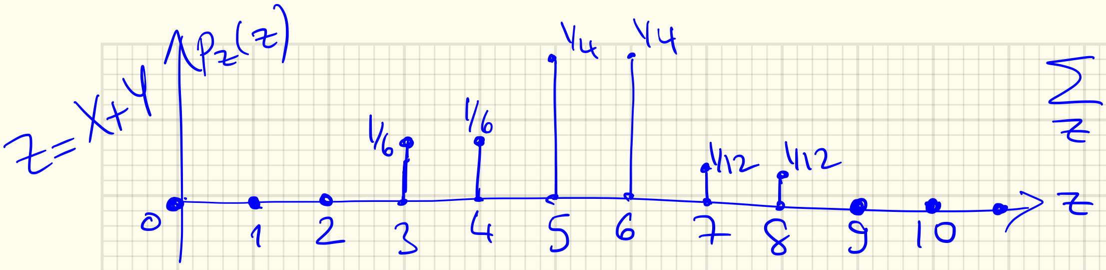
$$P_Z(z=5) = \frac{2}{3}\cdot\frac{1}{4} + \frac{1}{3}\cdot\frac{1}{4} = \frac{1}{4}$$

$$P_Z(z=6) = \frac{1}{4}$$



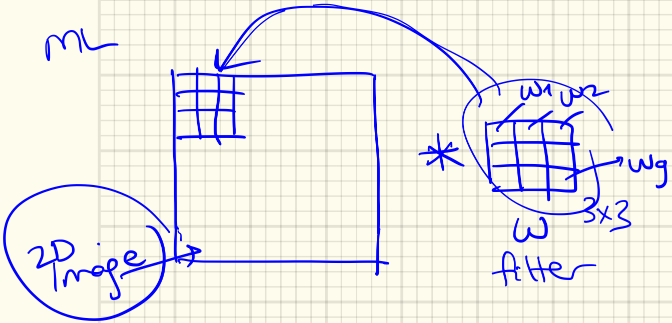
$$\Leftarrow P_Z(z=7) = \frac{1}{3}\cdot\frac{1}{4} = \frac{1}{12} = P_Z(\underline{z=8})$$

$P_Z(z)=0$ for $z \geq 9$ no overlap.



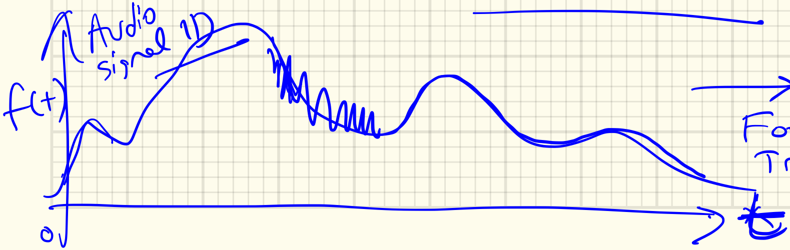
$$\sum_{z} p_Z(z) = 1 \quad \checkmark$$

ML

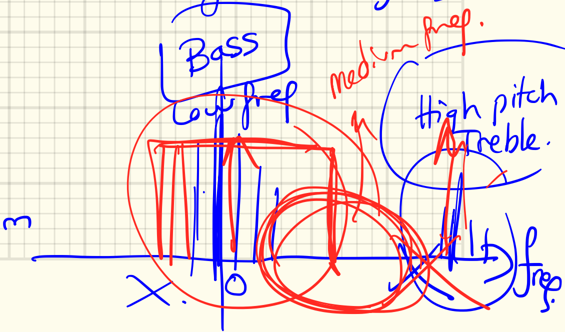


fixed weights in standard filtering

CNNs → $w_1, \dots, w_9$: learnable filter weights



Fourier Transform →



Moments of R.V.s :

1st Moment: $E[X]$: Expected Value / Mean of an r.v.

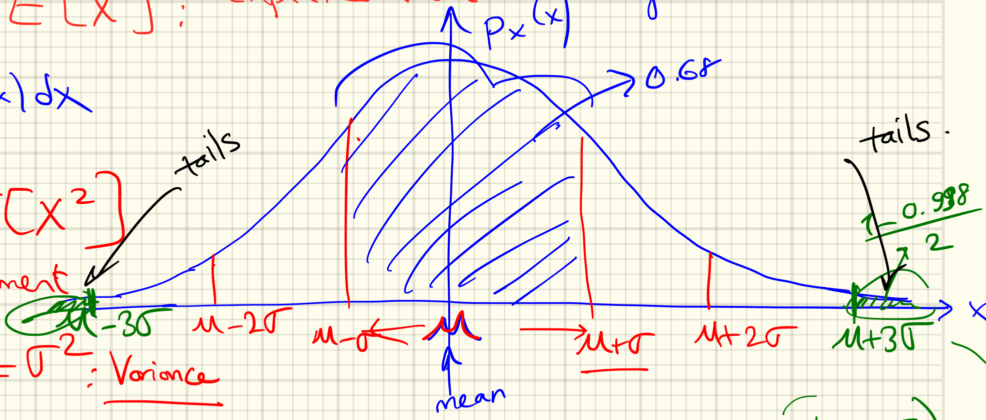
$$E[X] = \int_{-\infty}^{\infty} x \, p_x(x)\, dx$$

2nd Moment: $E[X^2]$

Centralized 2nd Moment

$$E[(X-\mu)^2] = \sigma^2$$: Variance

$\sigma = \sqrt{Var}$: Standard Deviation



For a Gaussian: $P(|X-\mu| \le \sigma)$

$$P(\mu-3\sigma < X < \mu+3\sigma) = \boxed{0.998}$$
$$P(\mu-2\sigma < X < \mu+2\sigma) = 0.955$$
$$P(\mu-\sigma < X < \mu+\sigma) = 0.68$$

Generalized Moments of an Rv.

nth moment of an r.v.

$$E[X^n] = \int x^n \cdot p_x(x)\, dx$$

✱ Note: nth moment of an r.v. exists 1 $E[|X|^n] < \infty$

centralized nth moment

$$E[(X-\mu_x)^n] = \int \underbrace{(x-\mu_x)^n}_{g(x)} p_x(x)\, dx = E[g(X)]$$

Note 2 : If we know that $E[X^s]$ exists, then $E[X^r]$ exists for $r<s$.

[SKay Prob. 6.23)

Ex: $X \sim \mathcal{N}(0,1)$ & indep.

$Y \sim \mathcal{N}(0,1)$

$$Z = \frac{Y}{X} \quad = h(x,y)$$

→ Cauchy Distrib. $p_Z(z) = ?$

Let $W = X$ $\quad g(x,y)$ (auxiliary r.v.)

$$\underbrace{(X,Y) \xrightarrow{(g,h)} (W,Z)}$$

$-\infty < w < \infty$

$-\infty < z < \infty$

$p_{W,Z}(w,z) = ?$

$$X = g^{-1}(w,z) = w$$

$$Y = h^{-1}(w,z) = z.w$$

Jacobian (of the inverse transformation): $\dfrac{\partial(x,y)}{\partial(w,z)} = \begin{bmatrix} \overset{\partial x/\partial w}{1} & \overset{\partial x/\partial z}{0} \\ \overset{\partial y/\partial w}{z} & \overset{\partial y/\partial z}{w} \end{bmatrix}$

$$|\det J| = |w|$$

$$p_{X,Y}(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

$$p_{W,Z}(w,z) = p_{X,Y}\left(g^{-1}(w,z), h^{-1}(w,z)\right).|w|$$

Joint $p_{W,Z}(w,z) = \dfrac{1}{2\pi} \exp\left(-\frac{1}{2}\left(w^2 + (wz)^2\right)\right)|w| = \dfrac{1}{2\pi} \exp\left[-\frac{1}{2}(1+z^2).w^2\right].|w|.$ →

$$p_Z(z) = \int_{-\infty}^{\infty} p_{W,Z}(w,z)\,dw = \int_{-\infty}^{\infty} \underbrace{\frac{1}{2\pi}\exp\left(-\frac{1}{2}(1+z^2)w^2\right)|w|}_{\text{integrand: even fn.}}\,dw$$

Marginalize joint pdf.

$$= \cancel{2}\cdot\int_{0}^{\infty} \frac{1}{\cancel{2}\pi}\exp\left(-\frac{1}{2}(1+z^2)w^2\right)\cdot w\,dw$$

exercise derive this.

(Lorentz) Cauchy density

$$\boxed{p_Z(z) = \frac{1}{\pi(1+z^2)}}, \quad -\infty < z < \infty$$

more general form

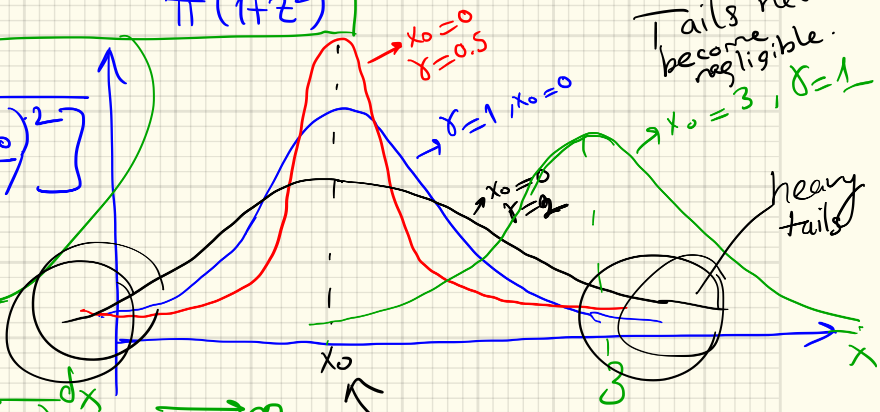
$$p_X(x) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

x_0: location

γ: scale

$x_0 = 0,\ \gamma = 1$

$$E[X] = ?\ \int_{-\infty}^{\infty} x\cdot\frac{1}{\pi(1+x^2)}\,dx \longrightarrow \infty.$$



E[X] does not exist!

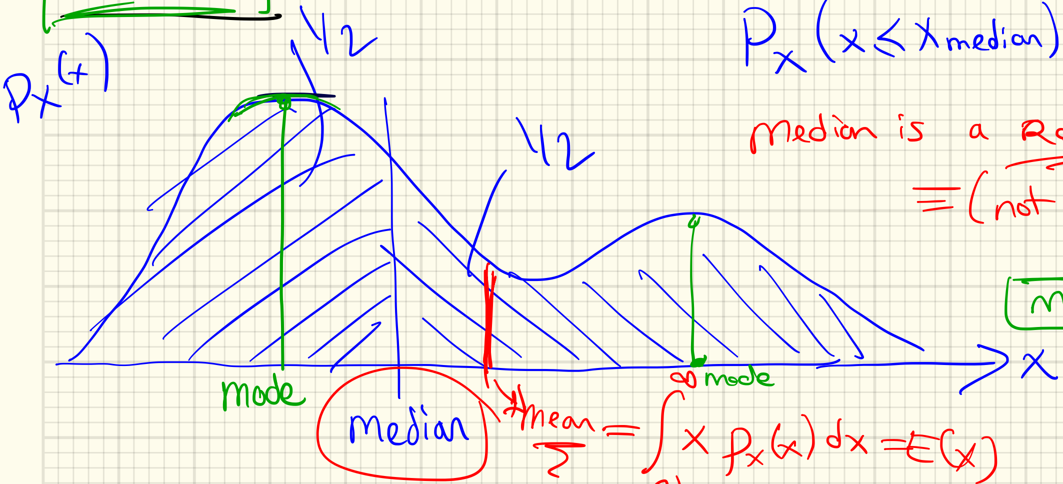
so no other higher moments also do not exist.

→ Median could be used to estimate the location parameter x_0.
b/c I cannot use the $E[X]$!

$\boxed{\text{Median:}}$

X_{median} is where $F_X(X_{median}) = \frac{1}{2}$

$$P_X(x \leq X_{median}) = \frac{1}{2} = \int_{-\infty}^{X_{median}} p_X(x)\,dx$$

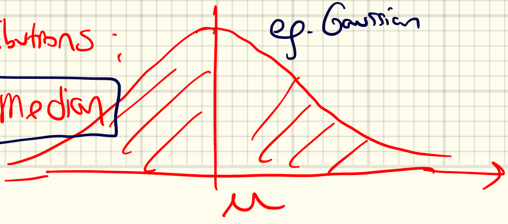


$$\text{mean} = \int_{-\infty}^{\infty} x\, p_X(x)\,dx = E(X)$$

Median is a Robust Statistic
= (not sensitive to extreme values or outliers)

$\boxed{\text{Modes}}$ of the distribution
= (local) maxima of the pdf.

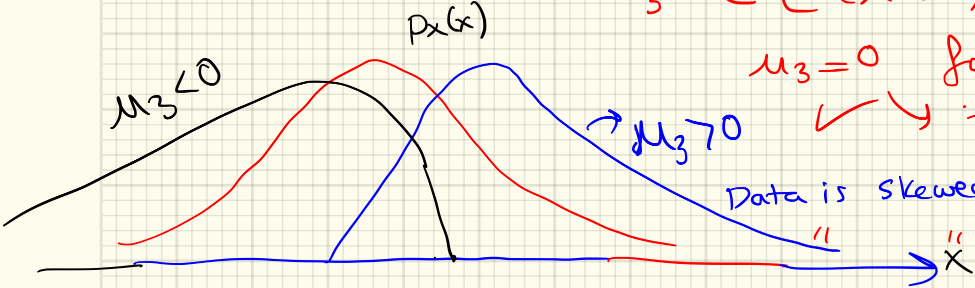
For symmetrical distributions:
If mean exists, $\boxed{\text{Mean} = \text{median}}$



3rd Central Moment : $\mu_3 = E[(x-\mu)^3]$: measure of symmetry

$\mu_3 = 0$ for normal distribution

↳ for any symmetric distrib.



Data is skewed to the right : $\gamma_1 > 0$

" " left : $\gamma_1 < 0$.

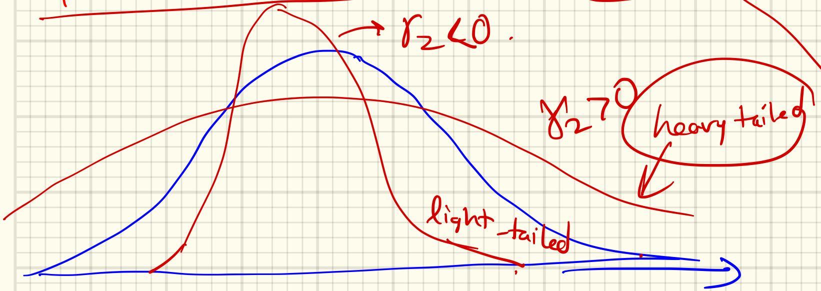
Coefficient of Skewness : $\gamma_1 \triangleq \frac{\mu_3}{\sigma^3} \Leftarrow (\sigma)^3$

→ std. $\sqrt{\sigma^2}$

4th Central Moment ; Kurtosis : compares any distrib. to a Gaussian

$$\mu_4 = E[(x-\mu)^4]$$

$$\boxed{\gamma_2 \triangleq \frac{\mu_4}{\sigma^4} - 3}$$



$\gamma_2 = 0$ for a Gaussian.

Q. Is a pdf of an r.v uniquely described by its moments?

$$\mathcal{N}(0,\sigma^2) \quad \text{vs.} \quad \text{Laplacian}\ (0,\sigma^2)$$

A. No

But exception: Gaussian r.v. (μ,σ^2) → pdf. defined by its 1st & 2nd moments.

Exercise: Calculate moments of an exponential r.v.

$$p_x(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & o/w \end{cases}$$

$$E[X] = \frac{1}{\lambda}$$

$$Var(X) = E[(X - E(X))^2] = \frac{1}{\lambda^2}$$

$$= E[X^2] - (E(X))^2$$

$$E[X^3] = \frac{3}{\lambda} E[X^2] = \frac{6}{\lambda^3} = \frac{3!}{\lambda^3}$$

$$E[X^n] = \dots = \frac{n}{\lambda} E[X^{n-1}] = \frac{n!}{\lambda^n}.$$

Note:

Characteristic Function of an r.v.

Relates moments of a distribution and the Fourier transform of the pdf.

We won't study this part

[Skay] 6.7
11.7.

you're not responsible!

COVARIANCE : btw 2 r.v.s . (multiple r.v.s)

$$\mathrm{Var}(\underbrace{X+Y}_{Z}) = E\left[\left(X+Y-(\mu_x+\mu_y)\right)^2\right]$$

$$= E\left\{\left((X-\mu_x)+(Y-\mu_y)\right)^2\right\}$$

$$= E\left[(X-\mu_x)^2\right] + E\left[(Y-\mu_y)^2\right] - 2\underbrace{E\left[(X-\mu_x)(Y-\mu_y)\right]}_{\triangleq \text{Covariance}(X,Y)}$$

$$\mathrm{Cov}(X,Y) \triangleq E_{X,Y}\left[(X-\mu_x)(Y-\mu_y)\right] = E_{X,Y}\left[XY - X\mu_y - Y.\mu_x + \mu_x\mu_y\right]$$

$$= E(XY) - \underbrace{E[X]}_{\mu_x}.\mu_y - \underbrace{E[Y]}_{\mu_y}.\mu_x + \mu_x\mu_y$$

$$= \underbrace{E_{X,Y}[X.Y]}_{} - \mu_x.\mu_y$$

$$E_{X,Y}[X.Y] = \int_{-\infty}^{\infty}\int x.y\, p_{X,Y}(x,y)\, dx\, dy$$

★ If X & Y are independent: $E_{X,Y}[X.Y] = \iint x.y.p_X(x)p_Y(y)dxdy$

$$= E[X].E[Y]$$

Independence implies (always) uncorrelatedness!

When X & Y are indep: $\text{Cov}(X,Y) = \underbrace{E_{X,Y}[XY]}_{E[X].E[Y]} - E[X]E[Y]$

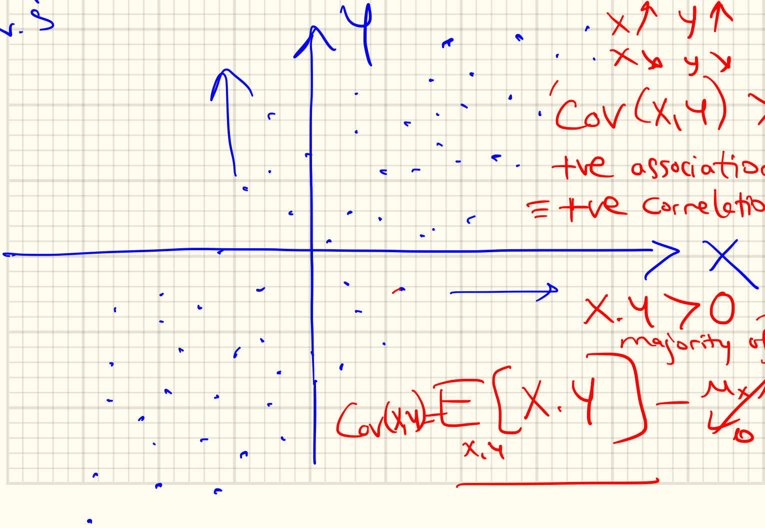
$E[X].E[Y] - E[X]E[Y]$

$\text{Cov}(X,Y) = 0$ ✓

X&Y Independent → X & Y are uncorrelated. $(\text{Cov}(X,Y) = 0)$)

But X&Y uncorrelated $\not\rightarrow$ X&Y are ~~independent~~

X,Y: r.v.s scatter plot of data sampled from X,Y



x↑ y↑

x↘ y↘

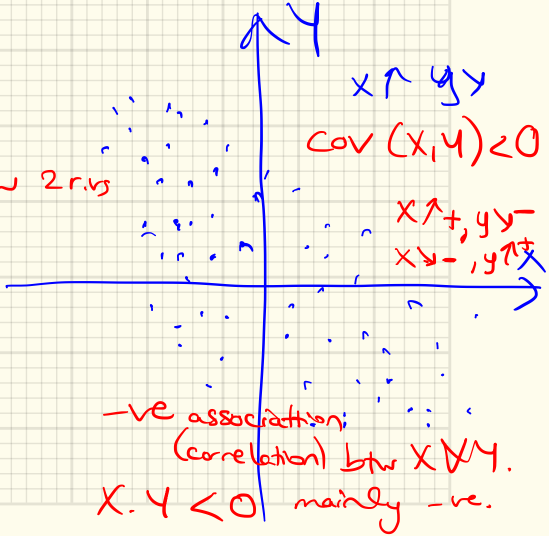
$\text{Cov}(X,Y) > 0$

+ve association btw 2 r.vs

≡ +ve correlation

$X.Y > 0$ for majority of X.Y

$\text{Cov}(X,Y) = E_{X,Y}[X.Y] - \cancelto{0}{\mu_x \mu_y}$



x↑ y↘

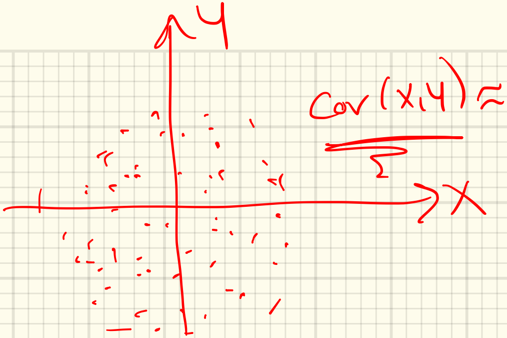
$\text{cov}(X,Y) < 0$

X↑+, y↘−

X↘−, y↑+

−ve association (correlation) btw X & Y.

$X.Y < 0$ mainly −ve.



$Cov(X,Y) \approx 0$

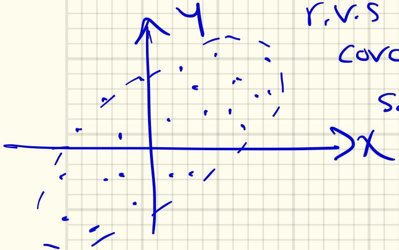
Correlation Coefficient : $\rho_{X,Y}$

(like in std → change the unit of Covariance to the "correct" unit.

$$\boxed{\underbrace{\rho_{X,Y}}_{\text{normalized version of Covariance.}} \triangleq \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}}$$

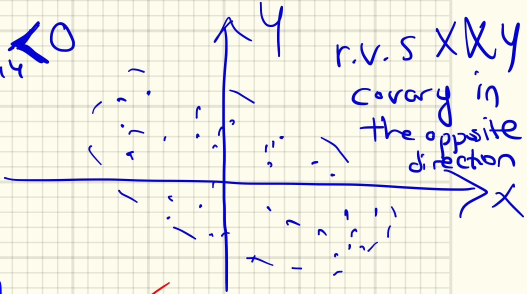
Case 1 $\rho_{X,Y} > 0$

r.v.s X & Y covary in the same dir



Case 2 $\rho_{X,Y} < 0$

r.v.s X & Y covary in the opposite direction



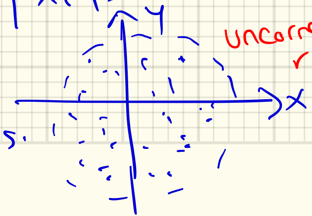
$$|\rho_{X,Y}| \leq 1$$

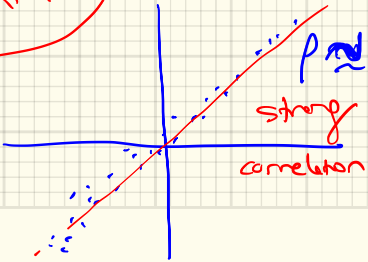
$$-1 \leq \rho_{X,Y} \leq 1$$

Case 3 $\rho_{X,Y} = 0$ correlation

uncorrelated r.v.s.

Lack of association btw 2 r.v.s.





$\rho \approx 1$: 2 r.v.s are linearly related.

strong correlation.

★ If X & Y are independent r.v.s then so are $g(X)$ & $h(Y)$ for any $g(\cdot)$ & $h(\cdot)$

★ If 2 r.v.s are independent → $\rho = 0$ (converse is not true).

★ Correlation coefficient (aka Pearson's correlation) detects linear dependencies btw 2 variables.

★ Independence is more general.

Ex: X : r.v. symmetrically-distributed around 0 → pdf of X can have any shape

Let $Y = X^2$. Q1: Are X & Y correlated? Q2: Are X & Y independent?

$$Cov(X,Y) = E_{X,Y}\left[(X - \underbrace{\cancel{\mu_X}}_{0})(\underbrace{Y}_{X^2} - \mu_Y)\right]$$

$$= E_{X,Y}[X^3 - X.\mu_Y] = \underbrace{E[X^3]}_{=0} - E[\overset{0}{\cancel{X}}].\mu_Y$$

(Skewness $=0$ for any symmetric distrib). $=0$ b/c $p_X(x)$ is symmetric)

① $Cov(X,Y) = 0$ ⇒ X & Y are uncorrelated.

② Are X & Y independent? $Y = g(X) = X^2$; Y is completely determined by X.

No! They are dependent!

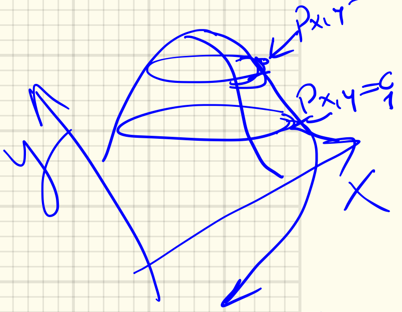
Uncorrelatedness does NOT imply independence.

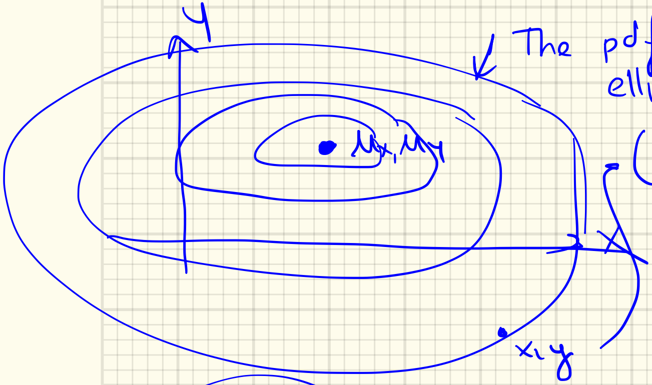
Two independent Normal r.v.s.

$X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$: independent

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$$

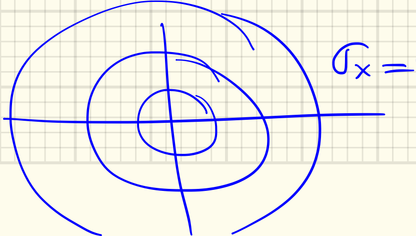
$$= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[\underbrace{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}} \right]$$



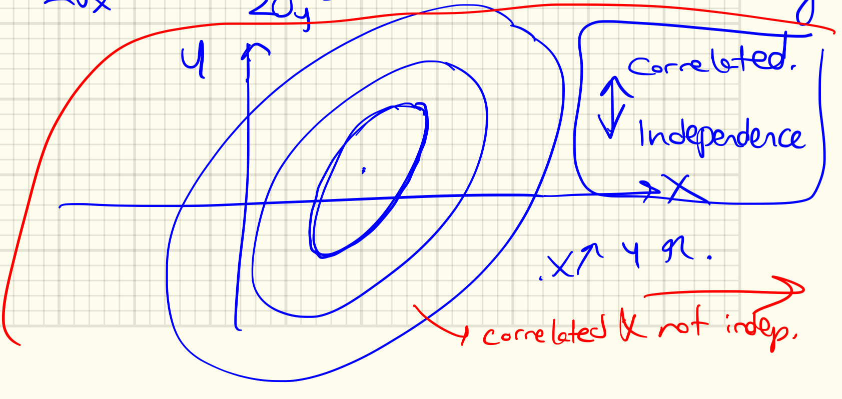


The pdf is constant on the ellipse where

$$\frac{(x-\mu_x)^2}{2\sigma_x^2} + \frac{(y-\mu_y)^2}{2\sigma_y^2} \text{ is constant}$$



$\sigma_x = \sigma_y$



Standard Bivariate Normal : (s.b.n).

$-1 < \rho < 1$ correlation coeff. btw X & Y.

joint pdf for 2 correlated Gaussians : $\mathcal{N}(0,1)$

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(x^2 - 2\boxed{\rho}\, x\cdot y + y^2\right)\right]$$

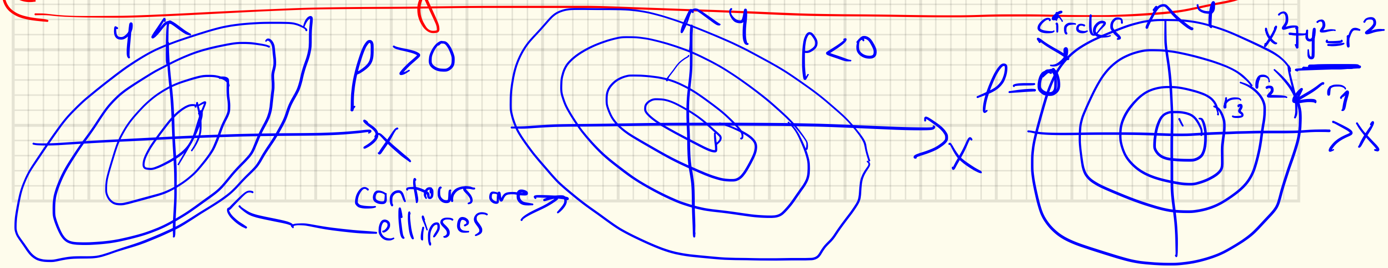
ep. insert $\rho = 0$; (X & Y are uncorrelated)

$$p_{X,Y}(x,y) = \frac{1}{2\pi}\exp\left[-\frac{1}{2}\underline{(x^2+y^2)}\right] = \frac{1}{2\pi} e^{-\frac{1}{2}x^2} \cdot e^{-\frac{1}{2}y^2}$$

$$= p_X(x) \cdot p_Y(y)$$

For Gaussian joint density Uncorrelated $(\rho=0)$ $\xrightarrow{\text{implies}}$ Independence.

exception! Normally → " $\not\longrightarrow$ " for other distributions.



Re arrange $\rightarrow$

$$\frac{x^2 - 2\rho x y + y^2}{\underbrace{1-\rho^2}_{\det Q}} = \begin{bmatrix} x \\ y \end{bmatrix}^T \underbrace{\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}}_{\underline{\underline{Q}} = \underline{\underline{C}}^{-1} \Rightarrow \det Q = 1-\rho^2} \begin{bmatrix} x \\ y \end{bmatrix}.$$

$\boxed{\text{Covariance Matrix}}$: symmetric matrix

for bivariate r.v.'s

Covariance matrix $\underline{\underline{C}} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$

$$\underline{\underline{C}} = \begin{bmatrix} \sigma_x^2 & \rho \cdot \sigma_x \sigma_y \\ \rho \cdot \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix} \leftarrow$$

General defn. for any 2 r.v.s.

Q. When X & Y are uncorrelated: $Cov(X,Y) = 0 \Rightarrow \underline{\underline{C}} = \begin{bmatrix} Var(X) & 0 \\ 0 & Var(Y) \end{bmatrix}.$

Covariance matrix for multi-variate r.v.s.

$X_1, X_2, \dots, X_N$:

$$\underline{\underline{C}} = \begin{bmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_N) \\ Cov(X_1,X_2) & Var(X_2) & & \\ & & \ddots & \\ Cov(X_1,X_N) & \cdots & Cov(X_{N-1},X_N) & Var(X_N) \end{bmatrix}$$

pairwise covariances.

→ diagonal: variances

* $X_1, \dots, X_n$ w/ X_i's zero-mean.

$$\boxed{Z = \sum_{i=1}^{n} X_i \quad E[Z]=0}$$

$$E[(\underbrace{X_1 + \cdots + X_n})^2] = Var(\underbrace{X_1 + X_2 + \cdots + X_n}_{Z})$$

Z : a sum r.v.

$$= E\left[\sum_{i=1}^{n} X_i^2 + \sum_{(i,j)} X_i X_j\right]$$

$$Var(\sum_i X_i) = \sum_{i=1}^{n} \underbrace{E[X_i^2]}_{Var(X_i)} + \underbrace{\sum_{i,j\ i \neq j} E[X_i X_j]}_{Cov(X_i, X_j)}$$

individual variances

cross-terms : correlations.

Q. : when X_i's are independent.

$$Var(\sum_i X_i) = ?$$

$$= \sum_i^n Var(X_i)$$

$$Cov(X_i, X_j) = 0 \quad i \neq j$$

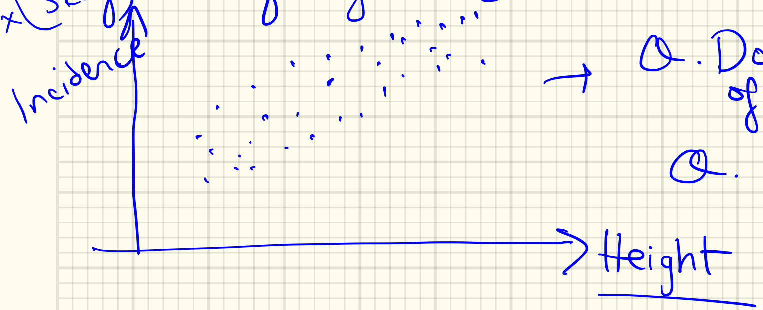
---

exercise : Derive $Var(\sum_i X_i)$ when mean's are not zero :

$$E[X_i] = \mu.$$

Correlation : Does it imply causation ?

ex (Skay): survey : age > 55 in the US : height & ~~prostabe~~ prostate cancer.



→ Q. Does this indicate a strong correlation of cancer w/ height ? Yes.

Q. Does getting taller cause an increased incidence of cancer ? No!

Correlation btw 2 variables only indicates ASSOCIATION (linear dependence) NOT CAUSATION ! No causal (physical) relation.

Do not deduce causation!

Paper:

Higher Chocolate Consumption leads to increased # Nobel prizes received by a country ! ! ! ! !