YZV 231E

13.12.2021

Probability Theory & Stats

Gü.

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

X r.v. $\xrightarrow{g(\cdot)}$ $Y = g(X)$: g : one-to-one (g^{-1} exist), many-to-one

Given pdf (pmf) $p_X(x)$ $\longrightarrow$ $p_Y(y) = ?$

1.) change of variables formula

$$p_Y(y) = p_X(\underbrace{x}_{g^{-1}(y)}) \left| \frac{dx}{dy} \right| \qquad \frac{1}{\left|\frac{dy}{dx}\right|}$$

2) CDF way :

i) $P(\underbrace{Y}_{g(x)} \leq y) = F_Y(y)$ find $F_X(\cdot)$

ii) Differentiate $p_Y(y) = \frac{d}{dy} F_Y(y)$

3) $W = g(X, Y)$: Using conditioning

$W = X + Y$ $\longrightarrow$

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

$W = g(X, Y)$

$Z = h(X, Y)$

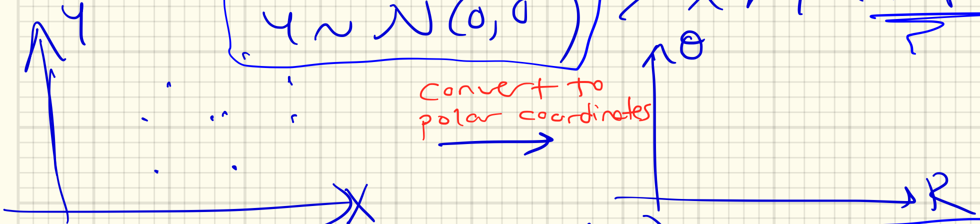
$$p_{W,Z}(w,z) = p_{X,Y}\left(g^{-1}(w,z), h^{-1}(w,z)\right) \left| \det\left(\frac{\partial(x,y)}{\partial(w,z)} \right) \right|$$

Jacobian of the transform

Ex: Received signal coordinates X, Y ; in radar/sonar.

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

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



R: magnitude of the signal

Θ: angle of the signal.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \longrightarrow \begin{pmatrix} R \\ \Theta \end{pmatrix} \qquad \begin{cases} R = \sqrt{X^2+Y^2}, & R \geq 0 \\ \Theta = \operatorname{atan}\frac{Y}{X}, & 0 \leq \Theta \leq 2\pi \end{cases}$$

$$\boxed{\begin{matrix} X = R\cos\Theta \\ Y = R\sin\Theta \end{matrix}}$$

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

Jacobian: $J = \frac{\partial(X,Y)}{\partial(R,\Theta)} = \begin{bmatrix} \frac{\partial X}{\partial R} & \frac{\partial X}{\partial \Theta} \\ \frac{\partial Y}{\partial R} & \frac{\partial Y}{\partial \Theta} \end{bmatrix} = \begin{bmatrix} \cos\Theta & -R\sin\Theta \\ \sin\Theta & R\cos\Theta \end{bmatrix} \rightarrow |\det J| = R \geq 0$

$(x_1, \ldots, x_m) \longrightarrow (y_1, \ldots, y_n)$

$P_{X,Y} = ?$

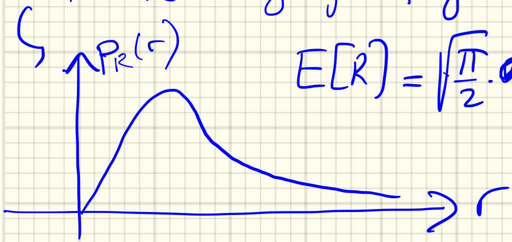
Jacobian $J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_m} \\ \vdots & & & \\ \frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix}_{n \times m}$

Joint PDF for X & Y

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

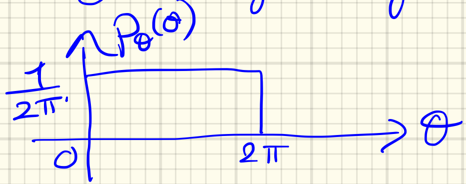
$$\Rightarrow p_{R,\Theta}(r,\theta) = \frac{(r)}{2\pi\sigma^2} e^{-r^2/2\sigma^2} = \underbrace{\left(\frac{r}{\sigma^2} e^{-r^2/2\sigma^2}\right)}_{p_R(r)} \cdot \underbrace{\left(\frac{1}{2\pi}\right)}_{p_\Theta(\theta)}$$

$R \sim$ Rayleigh pdf



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

$\Theta \sim$ uniform pdf



- Rayleigh distrib. is used in various engineering disciplines, including communications & signal proc.

Mechanics of Convolution:

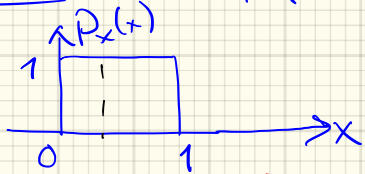
Recall: $Z = X + Y \longrightarrow X$ & Y are indep. r.v.s.

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

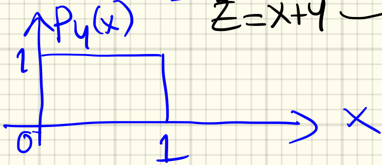
$$\boxed{p_Z(z) \triangleq \int_{-\infty}^{\infty} p_X(x)\, p_Y(z-x)\, dx}$$

Ex: Let $X, Y \sim U[0,1]$ & indep. $\longrightarrow p_Z(z) = ?$

$Z = X + Y$

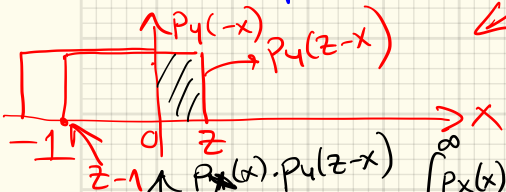


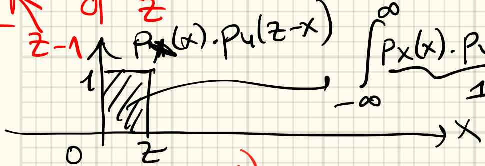
$*$



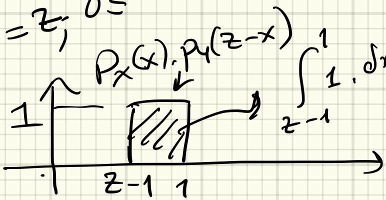
Result:

$$p_Z(z) = \begin{cases} 0, & z \leq 0 \\ z, & 0 < z \leq 1 \\ 2-z, & 1 < z \leq 2 \\ 0, & z > 2 \end{cases}$$

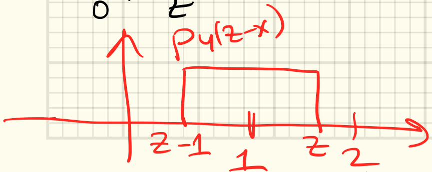




$$\int_{-\infty}^{\infty} \underbrace{p_X(x).p_Y(z-x)}_{1}\, dx = \int_0^z 1.dx = z; \quad 0 \leq z \leq 1.$$

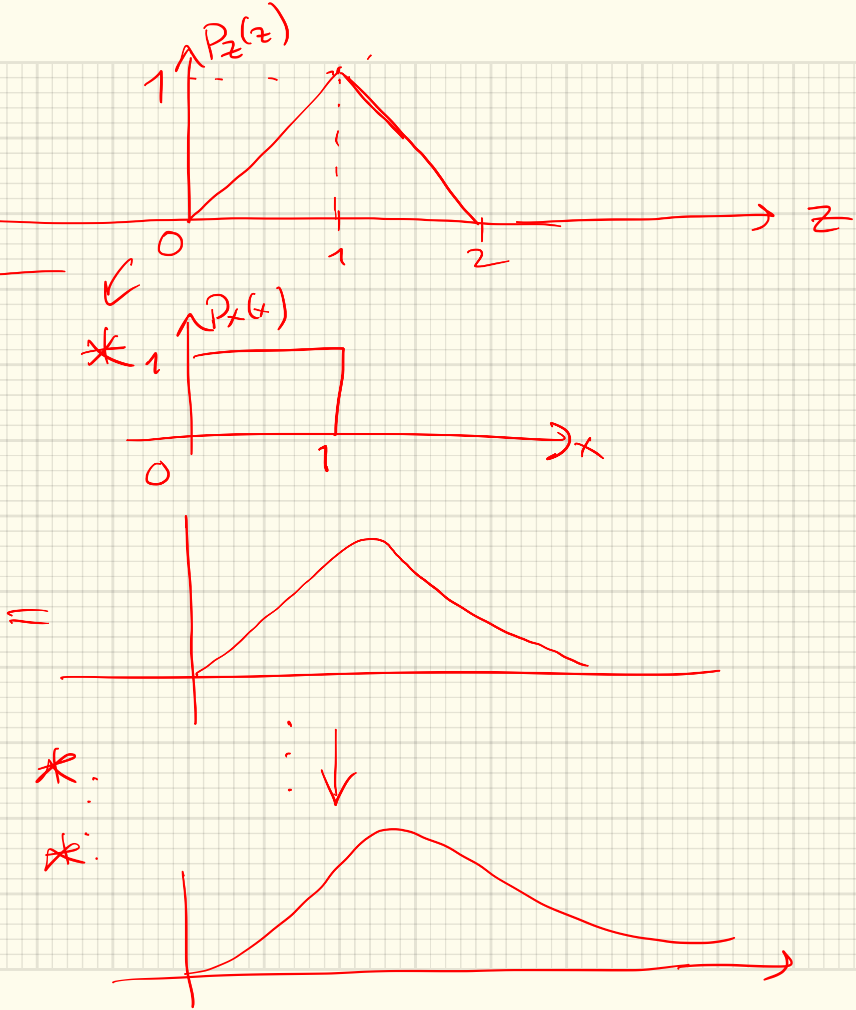


$$\int_{z-1}^{1} 1.dx = 2-z, \quad 1 \leq z \leq 2$$



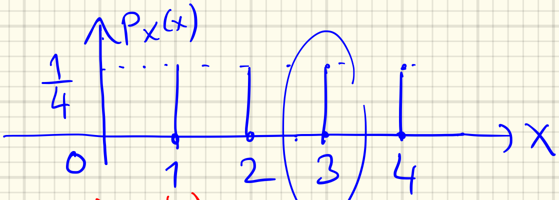
$p_X(x).p_Y(z-x) = 0$ (no overlap), $z > 2$

Later: Keep adding more $\mathcal{U}(0,1)$

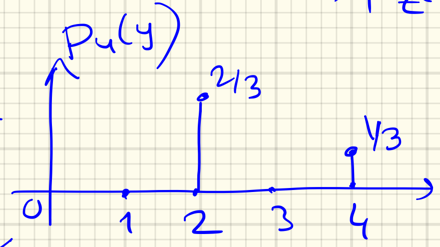


Example to Discrete Convolution operation.

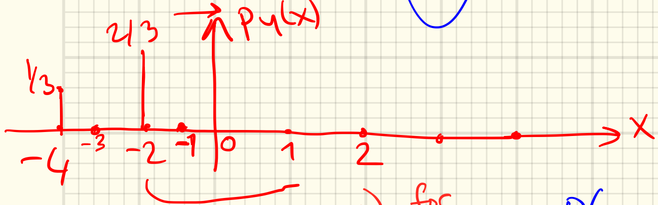
Ex: Discrete r.v.s X & Y : $Z = X + Y \rightarrow P_Z(z) = ?$



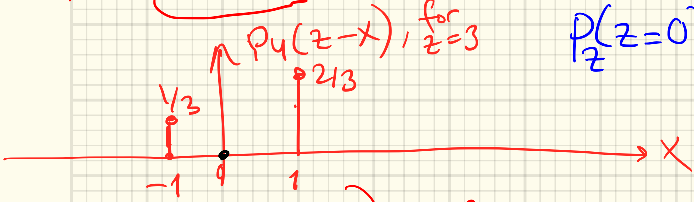
$*$



flip $p_Y(y)$, shift it over $p_X(x)$, multiply then add the result up.



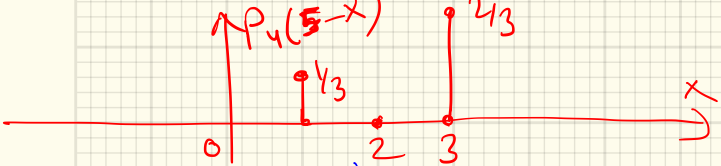
$$P_Z(z) = \sum_x P_X(x)\, P_Y(z-x)$$



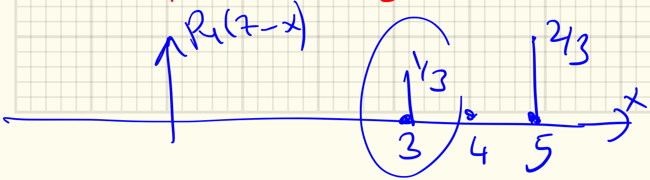
$$P_Z(z=0) = P_Z(z=1) = P_Z(z=2) = 0$$

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

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



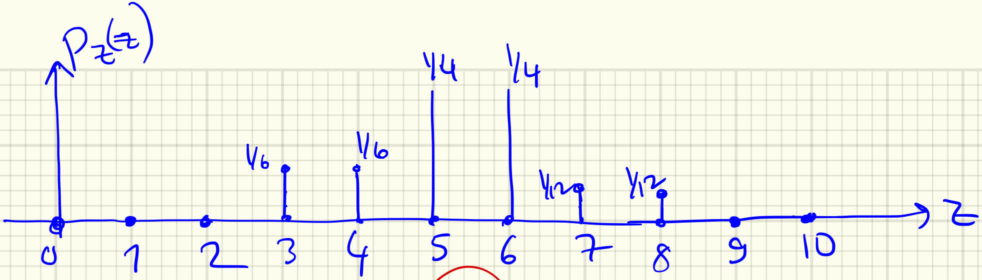
$$P_Z(z=5) = \frac{1}{3}\cdot\frac{1}{4} + \frac{2}{3}\cdot\frac{1}{4} = \frac{1}{4}$$

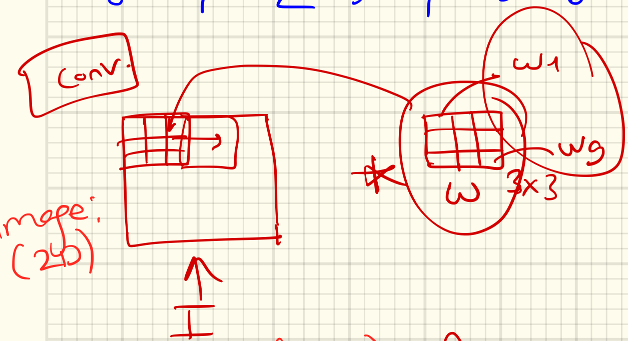


$$P_Z(z=6) = \frac{1}{4}$$

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

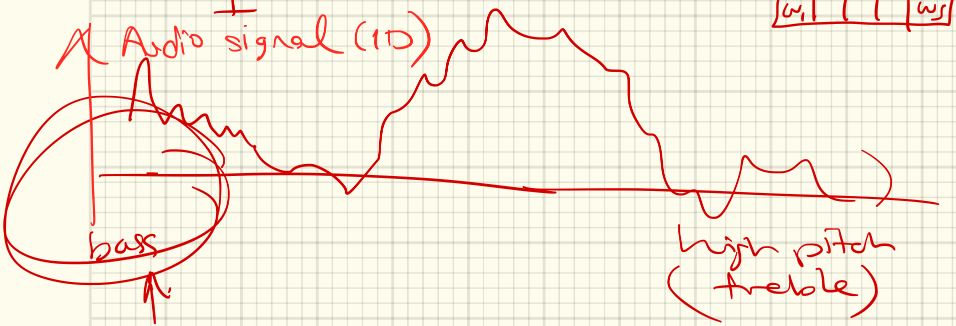
$\boxed{P_Z(z) = 0,\ z \geq 9}$ → no overlap

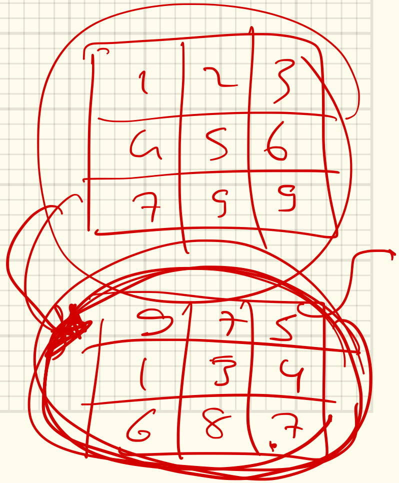




→ Fixed in standard filtering

Deep CNNs: $w_1, \dots, w_9$ are learnable filter weights





Moments of R.V.s :

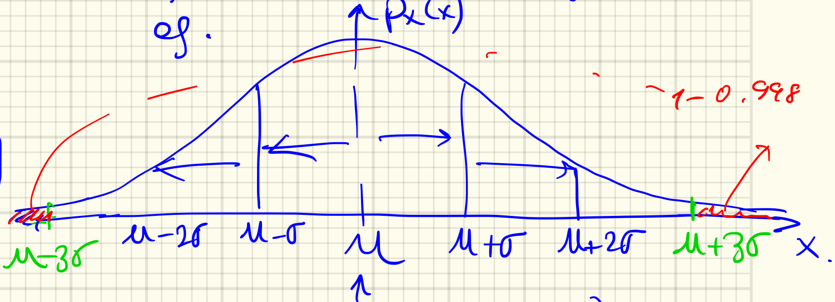
- 1^{st} moment : $E[X]$: Expected Value / mean of an r.v.
of.

- 2^{nd} moment: $E[X^2]$

Centralized moment (2^{nd} order)

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

$\sigma^2 \longrightarrow \sigma = \sqrt{Var(X)}$ same unit as the mean



For Gaussian: $P(|x-\mu| \leq \sigma)$

$$P(\mu-\sigma < X < \mu+\sigma) = 0.68$$

$$P(\mu-2\sigma < x < \mu+2\sigma) = 0.955$$

$$P(\mu-3\sigma < x < \mu+3\sigma) = 0.998$$

Generalized Moments:

n^{th} moment of an R.V

$E[(X-E[X])^n]$: n^{th} central moment

$E[X^n]$: not centralized n^{th} moment

Note: n^{th} moment exists if

$$E[|X|^n] < \infty$$

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

$$E[(\overbrace{X-\underbrace{E[X]}_{\mu_x}}^{g(x)})^n] = \int_{-\infty}^{\infty} (x-\mu_x)^n p_x(x)\, dx = E[g(X)] \quad \text{w/} \quad g(X) = (x-\mu_x)^n$$

If it is known that $E[X^s]$ exists, then $E[X^r]$ exists for $r<s$.

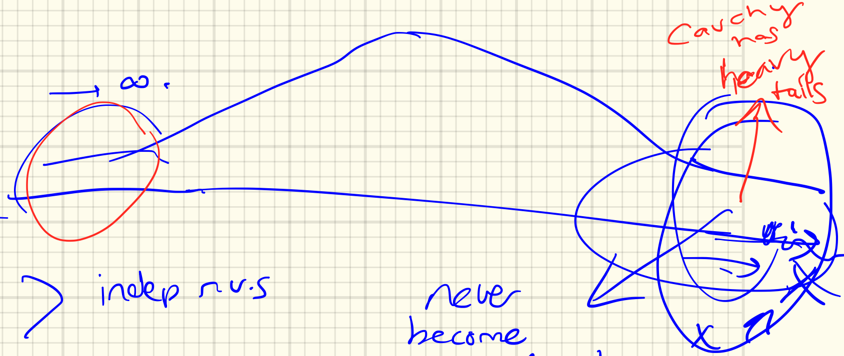
[Skay]. Prob. 6.23.

Exercise: Cauchy Distribution $p_X(x) = \dfrac{1}{\pi(1+x^2)}$;

Check Skay pdf

$E[X]$ does not exist! We need $E[|X|^n] < \infty$.

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



never become negligible.

Exercise;

$X \sim N(0,1)$
$Y \sim N(0,1)$
indep r.v.s

Let $Z = \dfrac{Y}{X} \longrightarrow$ Cauchy distrib.

check its uses.

Check. [Skay]

Ex: Calculate moments of an exponential r.v.:

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

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

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

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

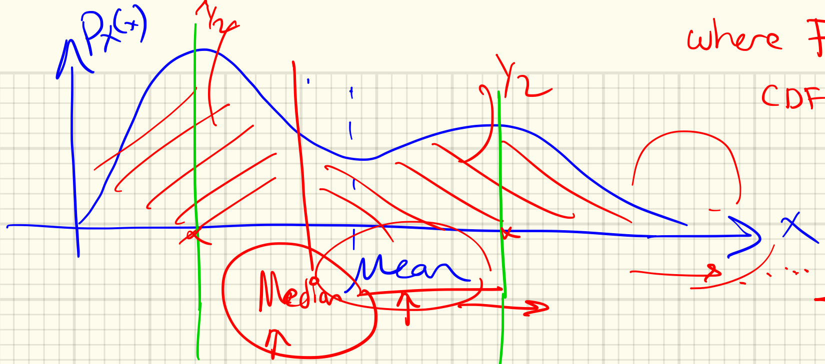
$$E[X^3] = \frac{3}{\lambda} E[X^2] = \frac{6}{\lambda^3}$$

$$Var(X) = \frac{1}{\lambda^2}$$

$$E[X^n] = \dots = \frac{n}{\lambda} E[X^{n-1}] \quad \left(= \frac{n!}{\lambda^n} \right)$$

[Note: Characteristic Functions : relates moments of a distribution X the Fourier transform of the pdf.]

We skip this part. [Skay] has this topic (6.7 & 11.7).



where $F_X(x^*) = \frac{1}{2}$

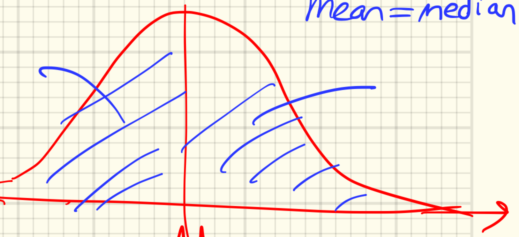
CDF

x^* : median value of the pdf.

robust statistic.

modes of the distribution : (local) maxima of the pdf.

★ For symmetrical distributions (like the Normal)

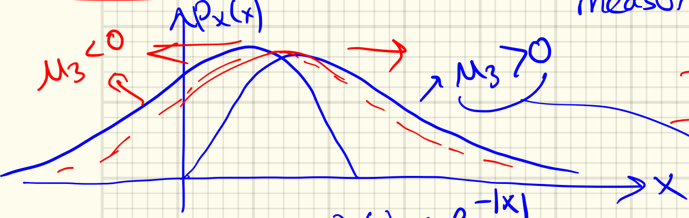


mean = median

μ = median = mode.

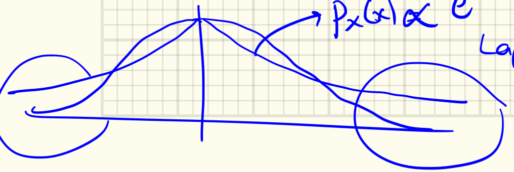
3rd Central Moment: $\mu_3 = E\left[(X-\mu)^3\right]$

measure of symmetry



- $\mu_3 = 0$ for normal distribution
- Any symmetric distribution

- Data is skewed to the right : $\mu_3 > 0$
- Data " " left $\mu_3 < 0$



$p_X(x) \propto e^{-|x|}$

Laplace

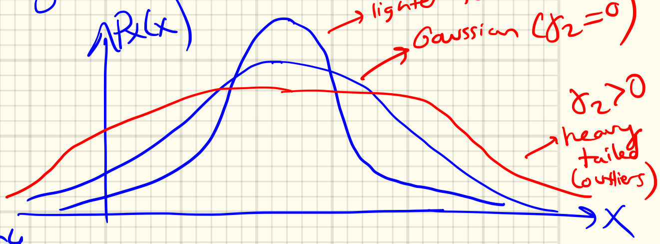
$\mu_3 = 0$

Coefficient of skewness $\gamma_1 \triangleq \dfrac{\mu_3}{\sigma^3}$

Right Tail > Left Tail

$\gamma_1 > 0$.

4th Central Moment : used to compare a given distrib. to a Gaussian.



$$\mu_4 = \int_{-\infty}^{\infty} (x-\mu)^4 \, p_x(x) \, dx = E\left[(x-\mu)^4\right]$$

for Gaussian: $\mathcal{N}(\mu,\sigma^2)$:

$$\mu_4 = \int_{-\infty}^{\infty} (x-\mu)^4 \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \, dx = 3\sigma^4$$

Define : Kurtosis (Coefficient of Excess) $\gamma_2 \triangleq \frac{\mu_4}{\sigma^4} - 3$

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

A: No! We have valid pdfs w/o some/all moments existing.

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

But ☆ Gaussian r.v $\rightarrow (\mu, \sigma^2) \rightarrow$ its pdf is defined by its 1st & 2nd moments.

COVARIANCE: between 2 r.v.s (or multiple r.v.s): whether they co-vary w/ each other or not!

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

$$= E\left[\left[(X-\mu)+(Y-\mu_y)\right]^2\right]$$

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

$$= \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$\mathrm{Cov}(X,Y) \triangleq E_{x,y}\left[(X-\mu_x)(Y-\mu_y)\right] \left(= E_{x,y}(X\cdot Y) - \mu_x \underbrace{E_{x,y}[Y]}_{\mu_y} - \mu_y \underbrace{E_{x,y}[X]}_{\mu_x} + \mu_x\mu_y\right)$$

$$\mathrm{Cov}(X,Y) = \underbrace{E_{x,y}[X\cdot Y]}_{} - \mu_x\cdot\mu_y$$

$$E_{x,y}[X\cdot Y] = \iint_{-\infty}^{\infty} x\,y\,P_{x,y}(x,y)\,dx\,dy$$

eg. $\boxed{\text{When } X \,\&\, Y \text{ are independent}}$

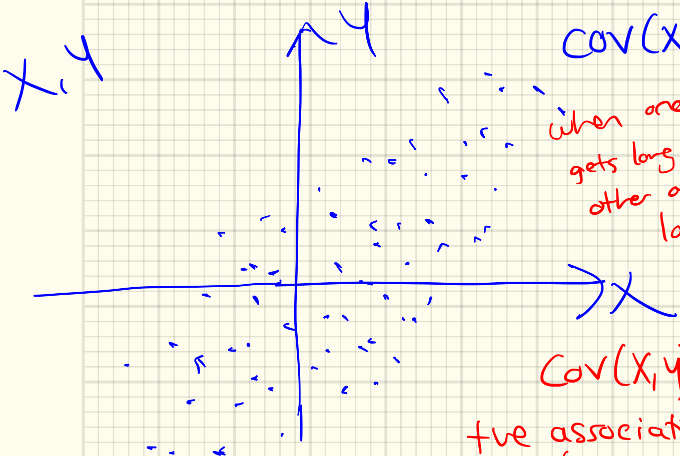
$$E_{x,y}[X\cdot Y] = \iint X\cdot y\, P_x(x)\cdot P_y(y)\,dx\,dy = E[X]\cdot E[Y]$$

$\Rightarrow$ When X, Y are independent: $Cov(X,Y) = \underbrace{E_{X,Y}[X \cdot Y]}_{= E[X]\,E[Y]} - E[X] \cdot E[Y]$

$Cov(X,Y) = 0$ always holds when $X \perp\!\!\!\perp Y$ are indep.

$X \perp\!\!\!\perp Y$ independent $\xrightarrow{\text{implies}}$ $X \perp\!\!\!\perp Y$ are uncorrelated $(Cov(X,Y)=0)$

But $X \perp\!\!\!\perp Y$ uncorrelated $\not\Rightarrow$ independent.

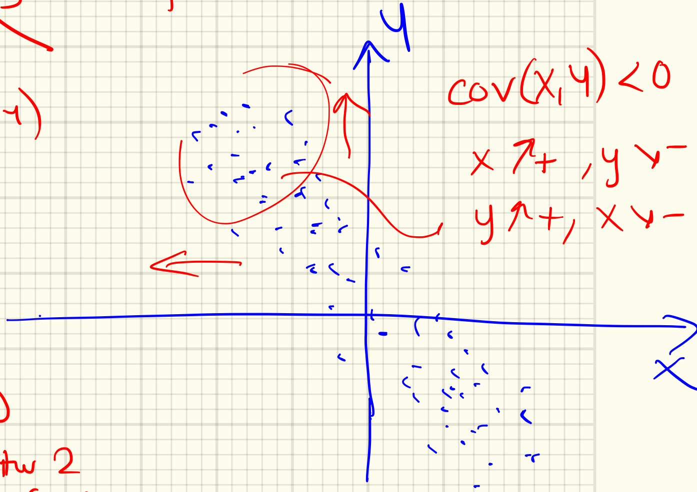


$cov(X,Y)$

when one (x or y) gets large, the other also gets large

$Cov(X,Y) > 0$

+ve association btw 2 r.v.s (correlation)



$cov(X,Y) < 0$

$x \nearrow +,\ y \searrow -$

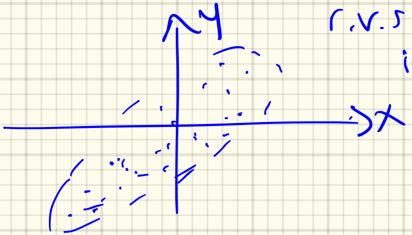
$y \nearrow +,\ x \searrow -$

Correlation Coefficient:

$$\rho_{X,Y} \triangleq \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}, \qquad |\rho_{X,Y}| \leq 1$$

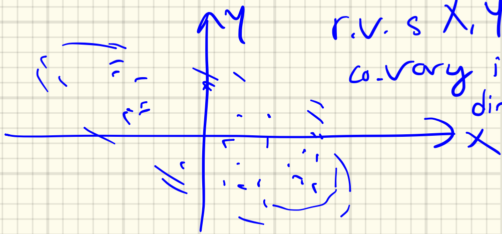
see proof prop 7.7 SKay

Case 1 $\rho_{X,Y} > 0$ $\left(cov(X,Y) > 0 \right)$



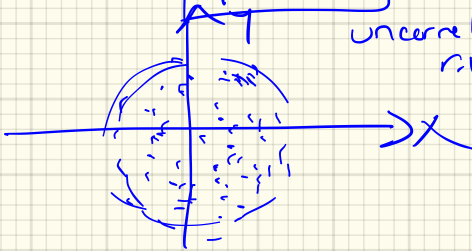
r.v.s X,Y co-vary in the same dir.

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



r.v.s X,Y co-vary in opposite directions

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



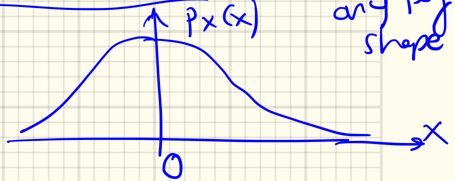
uncorrelated r.v.s.

Uncorrelatedness $\underset{\Leftarrow}{\not\Rightarrow}$ Independence

★ Correlation Coeff. (also known as Pearson's correlation) detects only linear dependencies between two variables.

★ Independence is more general.

$\boxed{\text{Ex:}}$ R.V. X is symmetrically distributed around 0.



any pdf shape

Let $Y = X^2$. $(\mu_x = E[X] = 0)$

$$Cov(X,Y) = E_{x,y}[X \cdot (\underbrace{Y}_{X^2} - \mu_y)]$$

$$= \underbrace{E_{x,y}[X^3]}_{=0} - E[X \cdot \mu_y] \qquad = \underbrace{E[X]}_{0} \cdot \mu_y = 0$$

$=0$ b/c $p_x(x)$ is symmetric (skewness is 0 for any symmetric distrib.)

1) $\Rightarrow X$ & Y are uncorrelated as $Cov(X,Y) = 0$

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

(★ Uncorrelation does not imply independence) No! They are dependent!

Recall 2 independent std Normal r.v.s:

$X \sim N(0,1)$; $Y \sim N(0,1)$

$$p_{X,Y}(x,y) = p_X(x)\, p_Y(y) = \frac{1}{2\pi} \exp\left\{-\frac{(x^2+y^2)}{2}\right\}$$

when $\rho = 0$

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

joint pdf for 2 correlated Gaussians: $N(0,1)$

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

ρ: correlation coefficient between X & Y ; $\boxed{-1 < \rho < 1}$.

for s.b.n. $\rho = \mathrm{cov}(X,Y)$

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}}$$

let's look at

$p_{X,Y}(x,y) =$ constant. $\rightarrow$ solve for (x,y)

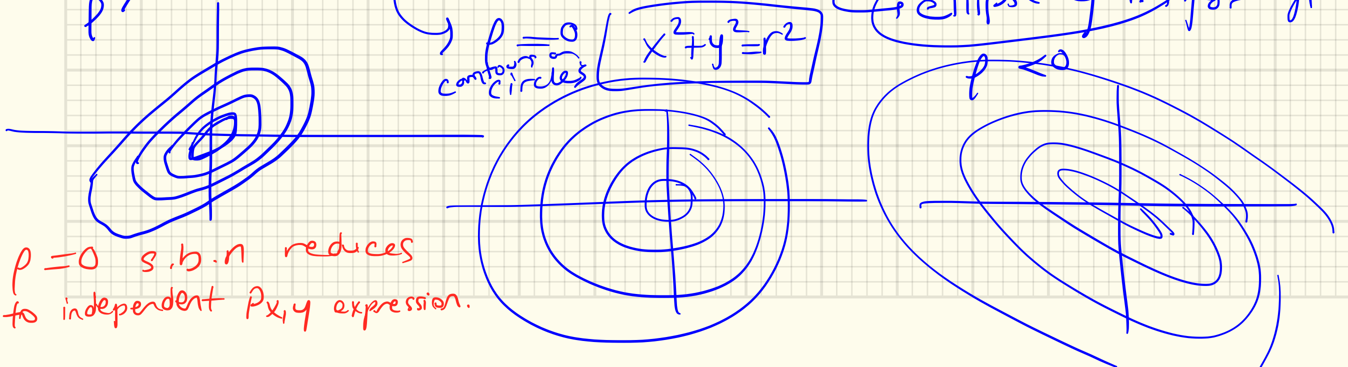
$\boxed{x^2 - 2\rho xy + y^2 = r^2}$ $\rightarrow$ contours of constant density

ellipse eqn. for different ρ.

$\rho = 0$ $\boxed{x^2+y^2=r^2}$

contours are circles

$\rho > 0$ $\rho < 0$



at $\rho = 0$ s.b.n reduces to independent $p_{X,Y}$ expression.

→ ∴ For Gaussian r.v.s only! Uncorrelatedness implies Independence!

$$\rightarrow \frac{x^2 - 2\rho xy + y^2}{(1-\rho^2)} = \begin{bmatrix} x \\ y \end{bmatrix}^T \underbrace{\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}}_{} \begin{bmatrix} x \\ y \end{bmatrix} \cdot \frac{1}{1-\rho^2}$$

$\rightarrow (1-\rho^2) = \det(\underline{\underline{Q}})$

For multivariate Gaussians (r.v.s)

$\underline{\underline{Q}} = \underline{\underline{C}}^{-1}$: $\underline{\underline{C}}$: covariance matrix

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

In general Covariance Matrix

$$\underline{\underline{C}} = \begin{bmatrix} \sigma_x^2 & \rho \cdot \sigma_x \cdot \sigma_y \\ \rho \cdot \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix}$$

$$\rho = \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}$$

$$\underline{\underline{C}}_{2\times 2} = \begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{bmatrix}$$

when $Cov(X,Y) = 0$:

$$\underline{\underline{C}} = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$$: Diagonal matrix

Covariance Matrix for multi-variate r.v.'s

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

→ Let's assume $X_1, \ldots, X_n$ with X_i zero mean.

$$E\left[(X_1 + \cdots + X_n)^2\right] = Var(X_1 + X_2 + \cdots + X_n)$$

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

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

cross-terms : correlations.

☆ Correlation between r.v.s : Does it imply causation ?

Higher Chocolate Consumption leads to #Nobel Prizes received by a country ???

a past paper !!!

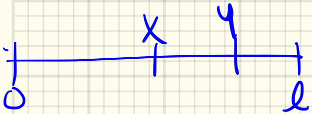
Correlation vs Causation.

Conditional Expectations:

→ Given the value of y of an r.v.

$$E[X|Y=y] = \int_{-\infty}^{\infty} x \cdot p_{X|Y}(x|y)\,dx \qquad \left(\sum_x x\, p_{X|Y}\right)$$

→ Recall stick ex. break uniformly at y; break again uniformly at X



$$E[X|Y=y] = \frac{y}{2} \leftarrow \text{unknown}$$

$$E[X|Y] = \underbrace{\frac{Y}{2}}_{} \longrightarrow g(Y)$$

is an r.v.

$g(Y)$: a general fn of Y

If $E(X|Y) = g(Y)$ is considered as an r.v.: it has an expectation

$$E[E[X|Y]] \longrightarrow$$

Law of Iterated Expectations:

$$E[E[X|Y]] = \sum_y E[X|Y=y]\, p_Y(y)$$

$E[X|Y]$ is a r.v. & is a fn. of Y.

$$= E[g(Y)]$$

$$= \sum_y E[X|Y=y]\, p_Y(y)$$

$$= E[X]$$

Recall:
⇝ Total Expectation Theorem.

$$\sum_y \sum_x x\, p_{X|Y}(x|y)\, p_Y(y)$$

$$\sum_x x \sum_y p_{X,Y}(x,y)$$

$$\sum_x x \cdot p_X(x)$$

$$= E[X]$$

$\cancel{E(E[g(x)|Y]) = E[g(x)]}$

$$\boxed{\underbrace{E[E[X|Y]]}_{\text{expectation of a conditional expectation}} = \underbrace{E[X]}_{\text{Unconditional expectation}}}$$

In the stick ex $E[E[X|Y]] = E[X] = \frac{\ell}{4}$.

Conditional Variances:

$Var(X|Y)$ is an r.v.

↳ we'll talk about its expectation as well.

$$\underbrace{Var(X|Y=y)}_{\text{for a specific } y \text{ value}} = E\left[\left(X - E[X|Y=y]\right)^2 \,\middle|\, Y=y\right]$$

→ $Var(X|Y)$: as an r.v.

Law of Total Variance :

$$\boxed{\underbrace{Var(X)}_{\text{Total Variance}} = \underbrace{E[Var(X|Y)]}_{1} + \underbrace{Var(E[X|Y])}_{2}}$$

(2nd term)

proof: $Var(X) = E[X^2] - (E[X])^2$

$Var(X|Y) = E[X^2|Y] - (E[X|Y])^2$ valid for cond. universe

take expectation

1st term RHS

$$E[Var(X|Y)] = \underbrace{E[E[X^2|Y]]}_{E[X^2]} - \underbrace{E[(E[X|Y])^2]}_{\text{cancels w/ the 1st term on the next page.}}$$

2

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

$$\mathrm{Var}\left(E[X|Y]\right) = E\left[\left(E[X|Y]\right)^2\right] - \left(E\left[E[X|Y]\right]\right)^2$$

$$\to \quad 1 + 2 \Rightarrow E[X^2] - \left(E[X]\right)^2 = \mathrm{Var}(X) = \mathrm{LHS}.$$

Next time:

Continue w/ examples.