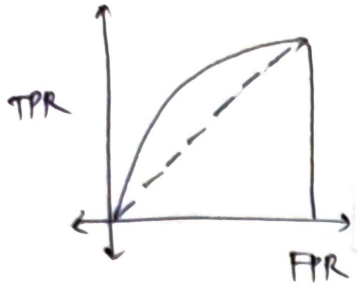
AUC-ROC

→ AUC = 0.9 ; what does it mean?



ans if we have 2 pts of different classes (here ⊕ ⊖) and we dont know which is +ve and -ve,

AUC = 0.9 → 90% it will classified correctly

AUC : 0.5 → random model ⇒ If AOC < 0.5 swap 0 and 1

AUC = 0.1 → [R] 0→1, 1→0 → AOC = 0.9

* In KNN why manhatan > Euclidean in high dimensionality

Euclidean = $(x_1 - x_2)^2$ → as sq → it amplifies difference / distance

manhatan = $|x_1 - x_2|$

Type I ≡ sig level

Hypothesis Testing

1) Null hypothesis — ideal
2) Alternate hypothesis — opp of null

take sample from popul ∴ and calculate statistic

pval = P (obs | Ho is true)

p-value

H_0: μ time spent = 20 min → take a sample of N = 100, find $\bar{\mu}$, std (popul ∴ $\mu = \bar{\mu} = 25$)

H_a: μ time spent ≥ 20 min

$(\bar{\mu} \geq 25)$

P.val = P $(\bar{\mu} = 25 \mid H_0)$

prob that our observation occurs given Ho is true

α → **significance level** : if $p < \alpha$; Ho rejected

$p \geq \alpha$; Ho accepted

generally $\alpha = 0.05$

Type I and Type II error

P(Type I error) = α

	Ho is true	Ho is false
Reject Ho	TYPE I error	correct
Fail to Reject Ho	Correct	TYPE II Error

} reality

Type I = FP

Type II = FN

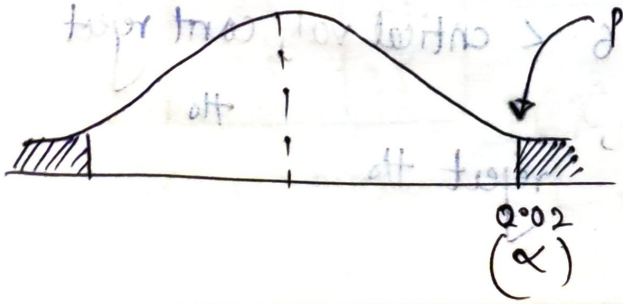
Type I : actual is True, you say False

Type II : actual is False, you predict True

Type I

reject Ho when it was True

→ assume Ho is true and $\alpha = 2\%$



Type I :- reject Ho when its true

Type II : accpt Ho, when its false

→ ∴ there is 2% chance that you made on Type II error

* conditions for picking sample from ppln

① Random

② Normal — atleast 10 sucess and failure

③ Independent — size(sample) ≤ 10% of population

z-table

x → p(≤x)

↳ value in z table

problem → speak ≥1 lang

$H_0 : p = 0.26$

$H_a : p > 0.26$

; sample → 120 ppl

(40/120) speak more than one language

signif^n test abt proportion

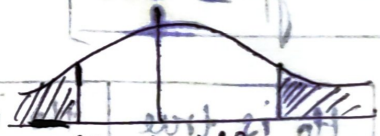
p_0 – assumed popⁿ proportion

$$z = \frac{\text{obtained prop} - \text{assumed proportion}}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$ ← std

→ deviation from p_0

$$z = \frac{0.33 - 0.26}{\sqrt{\frac{.26(1-.26)}{120}}} = 1.83$$

∴ P-value ≈ $p(z \geq 1.83) = 0.0336$



Significant test abt mean

CHI-S^2

check if our theory fits observation

→ theory : equal probs = 25, 25, 25, 25 (out of 100 students)

obsⁿ : 20, 20, 25, 35

Question

$\chi^2 = 6$

$\alpha = 0.05$

freedom = 3

→ get **critical val from χ^2 table**

if $\chi^2 = 6$ < critical val, can't reject

else, reject H_0

H_0 : our theory is correct

H_1 : theory is not correct

Chi-Square Test

χ^2 is a distribution of distributions (picked from normal distribution)

$\chi^2_1 = X^2_{0,1}$ [$X \sim N(0,1)$]

$\chi^2_2 = \chi^2_1 + \chi^2_1$

χ^2_i → degree of freedom (no of terms)

chi-table

degree of freedom	0.99	0.8	- - -	0.05	- - 0.1
1					
2					
3				7.81	
:					
n					

pdf → p critical

cdf → p-val

for χ^2 of degree = 3,

$P(\chi^2 = 7.81) = 5\% = 0.05$

$P(\chi^2 \geq 7.81) \geq 0.05$

Example

take mcq exam with 4 options

H_0 : equal prob of choices

H_a : unequal

$$\chi^2 = \sum_{i=1}^{n} \frac{(obs_i - expected)^2}{exp\text{-}val}$$

sample size = 100 → find probability of getting this | given H_0 is true

choice	expected	obs
A	25	20
B	25	20
C	25	25
D	25	35

chi²-stat χ

$$\chi^2 = \frac{(20-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(25-25)^2}{25} + \frac{(35-25)^2}{25}$$

$\chi^2 = 6$

degree of freedom = $(n-1) = 3$

↳ because n^{th} can be calculated using ③

$P(\chi^2 \geq 6) \geq 10\%$

↳ can be taken as p-value

let $\alpha = 0.05$

mean Est — Estimation ← point estimation / Interval estimation

(Q1) company tells pcs over 25 months with $\mu = 370.16$ and $\sigma = 75$. Estimate mean demand for next month with 95% confidence

→ $P\left(\bar{x} \pm Z_{(\alpha/2)} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$

α = significance level

$1 - \alpha = 95\% \Rightarrow \alpha = 0.05$

$Z_{0.025} = 1.96$; $370.16 \pm 1.96 * \frac{75}{\sqrt{25}} \Rightarrow 340 < \mu < 399$ reg

Chi-square independence test → check if 2 variables r independent

goodness of fit ⇒ check theory true

* 2 categorical variables.

Ex:- a group of 120 ppl are given to choose their favorite social media app. Is social media app preference independent of gender

Ho: gender & app r independent

	male	female	
FB	15	20	35
INSTA	30	35	65
TIK	5	15	20
	50	70	120

OBSERVED

$F_{male} \cdot F_{female} / Total$

	male	female	
FB	14.6	20.4	35
INSTA	27.1	37.9	65
TIK	8.3	11.7	20
	50	70	120

EXPECTED

→ calculate $\chi^2 = \sum \frac{(obs - exp)^2}{exp}$

if independent, $P(male \cap FB) = P(male) \cdot P(FB)$

$\frac{35 \times 50}{120}$

degree freedom = $(3-1)(2-1) = 2$

frequencies angle

Chebyshevs Theorem

* if data is gaussian distribution

$\mu \pm \sigma \rightarrow 68\%$
$\mu \pm 2\sigma \rightarrow 95\%$
$\mu \pm 3\sigma \rightarrow 99.7\%$

* if data is not normally distributed

Atleast $\left(1 - \frac{1}{K^2}\right)$ values will fall within $(\mu + K\sigma)$ and $(\mu - K\sigma)$

$0.xx \rightarrow xx\%$

$(\mu \pm K\sigma)$ contains $\left(1 - \frac{1}{K^2}\right)$ proportion of values

sample μ and σ^2 → why $(n-1)$ in D^r?

because when we take sample, it was found that it was underestimated

$\therefore$ reduce d^r by 1 to compensate

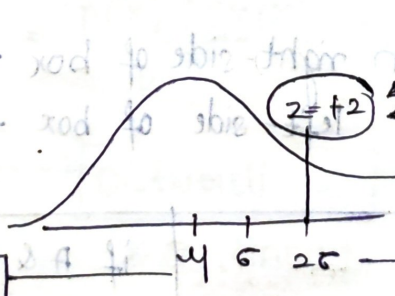
$$\mu = \frac{\sum x_i}{n-1}$$

$$\sigma^2 = \frac{\sum (x - \bar{x})}{n-1}$$

Z scores it represents no of SDs a value (x) is above or below the mean, when dist^r is normal.

$$z = \frac{x - \mu}{\sigma}$$

Ex: $z = +2$: shows that value (x) is 2 SDs above mean



*used to compare SD's with different means

Coef of variation

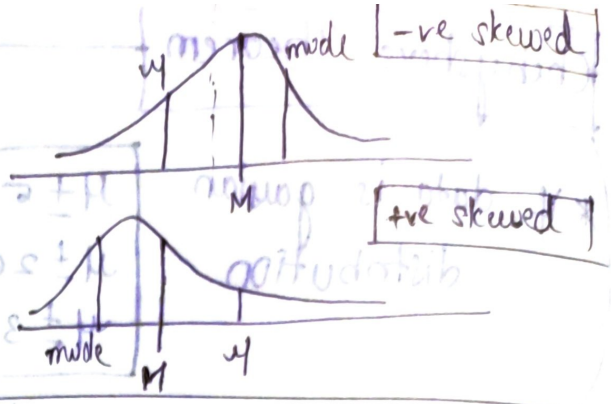
$$CV = \frac{\sigma}{\mu}(100) \Rightarrow$$ The SD is $(CV)\%$ of the mean

↳ relative component of SD to its mean

* Coef of skewness :-

$$S_k = \frac{3(\mu - M)}{\sigma}$$

where,
M = median



Kurtosis

* Amount of peakedness of a distribution

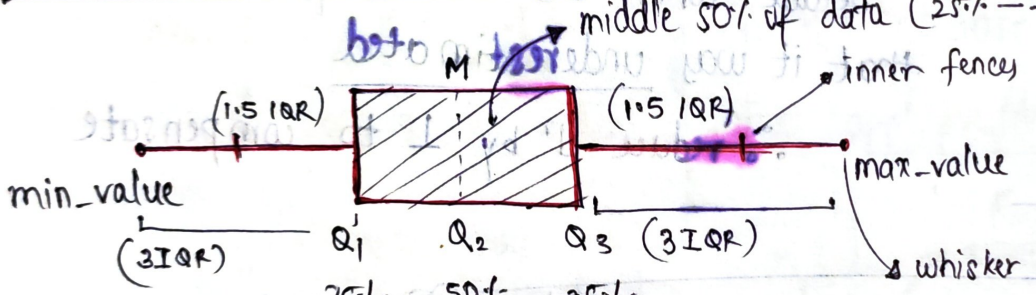
Quartiles

IQR = $Q_3 - Q_1$ (50% of data)

$Q_1 \rightarrow 25^{th}$ percentile
$Q_2 \rightarrow 50^{th}$ percentile / median
$Q_3 \rightarrow 75^{th}$ percentile

median

Box and Whisker plot



INTERPRETN

* values after / before / outside inner fences may be treated as outliers
* values b/w (Q_1, Q_3) and inner fences ⟶ mild outliers
* is dist^n skewed :- if median is in right-side of box – left skewed
  " " ⟶ left-side of box – right skewed

Prob law of addition

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) + P(A \cap B) = P(A) + P(B) = 1$$

if A & B are mutually exclusion ⟶
$P(A \cap B) = 0$

Probability

* apriori → can be determined before the experiment

* mutually exclusive events — $P(A \cap B) = 0$

* Independent events
  → ocurence of one event does not affect the other event in any other way
  $P(X|Y) = P(X)$ and
  $P(Y|X) = P(Y)$

Sampling from population

with replacement

$(N)^n$

N — population size
n — sample size

with replacement

$^{N}C_{n} = \frac{N!}{(N-n)!\,n!}$

Probability laws

addition

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

↓ prob of either A or B happening

* multiplication

$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$

→ $P(A) \cdot P(B)$ if A & B are independent

↓ prob of A and B both happening

Conditional prob

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} = \frac{P(X)\,P(Y|X)}{P(Y)} = \frac{P(X) \cdot P(Y|X)}{P(X) \cdot P(Y|X) + P(\bar{X})\,P(Y|\bar{X})}$$

Distributions

* discrete distrns
  1. Binomial
  2. Poissons
  3. Hypergeometric

* continous distrubns
  1. uniform
  2. normal
  3. Exponential
  4. t-distributns
  5. chi-square
  6. F-distribution

Binomial Distribution

$$P(x) = {}^{n}C_x \cdot p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} \cdot p^x \cdot q^{n-x}$$

$\mu = np$

$\sigma = \sqrt{npq}$

asumptions
- n identical trials
- two possible outcomes
- Independent trials → with repl
- p(sucess) and p(fail) remain same throughout all trials

$n \rightarrow$ no of trials
$x \rightarrow$ no of desired sucesses
$p \rightarrow$ probability of sucess in 1 trial
$q \rightarrow (1-p)$

Poissons Distribution → focus only no of ocurenurences over a interval

law of rare events

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$x = 0, 1, \ldots n$ (no of ourences per interval)
λ = long-run avg
$e = 2.7182$

Hypergeometric distribution

* Binomial dist^n is for sampling with replacement

Sample without replacement (trials r not independent)

Conditions
* size of popula^n is known
* The no of sucesses in the popula^n must be known

$$P(x) = \frac{{}^{A}C_x \cdot {}^{N-A}C_{n-x}}{{}^{N}C_n}$$

N – size of population
n – sample size
A – no of sucesses in pop^n
x – no of sucesses in sample

} sampling is done without replacement

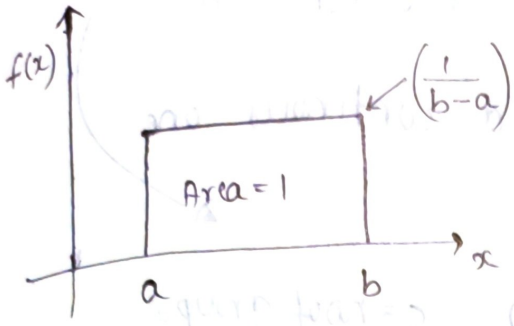
when to use hyper instead of Binomial?
1. Sampling without replacement
2. sample size $\geq$ 5% of population

Uniform Distribution

$$f(x) = \begin{cases} \left(\frac{1}{b-a}\right) & \text{for } a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

$$\mu = \frac{a+b}{2}$$

$$\sigma = \frac{b-a}{\sqrt{12}}$$



* Area enclosed = 1

$l \times b = (b-a)(\text{Height}) = 1$

$\therefore \boxed{\text{height} = \frac{1}{(b-a)}}$

Probabilities in a uniform distr

$$P(x) = \frac{x_2 - x_1}{b-a}$$ where, $a \le x_1 \le x_2 \le b$

Probability of val lies b/w x_1 and x_2

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

standardize → $\mu = 0$, $\sigma = 1$

$$Z_i = \frac{x_i - \mu}{\sigma} \quad \forall\, x_i \in X$$

standard distribution = ND with $\mu = 0, \sigma = 1$

Exponential Distribution

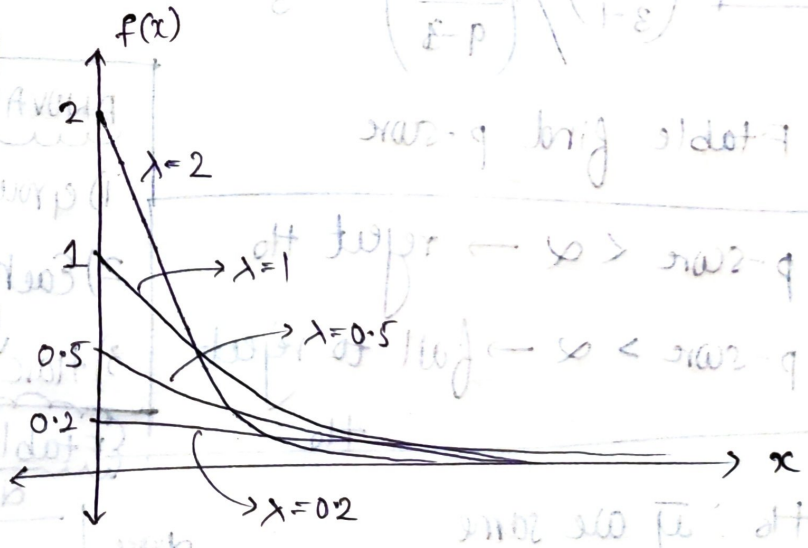
* continous and skewed to the right
* $0 \le x \le \infty$ and apex is at $x = 0$
* as $x \uparrow$, curve steadily decrease

$$f(x) = \lambda e^{-\lambda x}$$

$$\mu = \frac{1}{\lambda}$$

$$P(x \ge x_0) = e^{-\lambda x_0}$$

$x_0 \ge 0$



Ex:- prob of $x \ge 2 \mid \lambda = 1.2$

$P(x \ge 2) = e^{-(1.2)(2)} = 0.0907$

ANOVA

samples

ONE-WAY ANOVA → can we say each grp is dif wrt continous

→ chuk if for a categorical variable given a continous one means are significantly different

Ex: fi: categorical with 3 groups ($n_1+n_2+n_3$) c = no of groups

$N = n_1+n_2+n_3$

$$SS_{Between} = \sum (\bar{x}_i - \bar{\bar{x}}_g)^2 = \sum n_i (\bar{x}_i - \bar{\bar{x}}_g)^2$$

$$SS_{within} = \sum (x_i - \bar{x}_i)^2$$

$$F = \frac{MS_B}{MS_W} = \frac{\left[\frac{SS_B}{c-1}\right]}{\left[\frac{SS_W}{n-c}\right]}$$

F-score

Ex:- g1 (1,3,5) g2 (5,7,9) g3 (4,5,6)

$\bar{x}_1 = 3$ $\bar{x}_2 = 7$ $\bar{x}_3 = 5$ → $\bar{\bar{x}}_g = \frac{3+7+5}{3} = \frac{15}{3} = 5$

$SS_W = [(-2)^2+0^2+2^2] + [-2^2+0^2+2^2] + [-1^2+0^2+1^2] = 18$

$SS_B = 3(3-5)^2 + 3(7-5)^2 + 3(5-5)^2 = 24$

$\therefore$ Fscore $= \left(\frac{24}{3-1}\right) \Big/ \left(\frac{18}{9-3}\right) = \frac{12}{3} = 4$

use F-table find p-score

n = no of datapts
c = categories #
MS = mean square
SS = sum of squares

if p-score $< \alpha$ → reject H_0

p-score $> \alpha$ → fail to reject H_0

ANOVA assumptions

1) groups are condi= independent
2) Each group is normall distr
3) Have same variance

F table

degree freedom N^r $(c-1)$

degree ND^r $(N-c)$ → ▨ → critical val

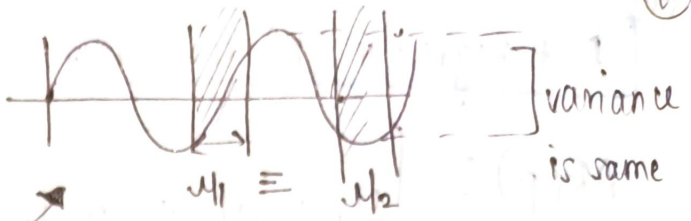
if $F <$ crit → accept

H_0 : $\bar{y}$ are same

H_1 : statically different

TIME SERIES

Stationary Time Series



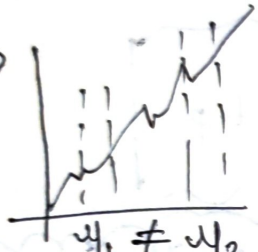
1. Mean constant over time
2. variance constant
3. Co-variance is only a function of gap/width ⟶ take a gap/width of K

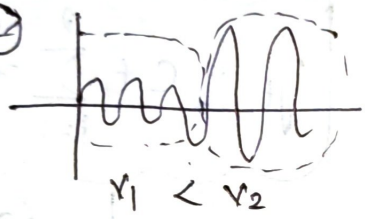
C_1	C_2
t	t+K
t-1	t+K-1
t-2	t+K-2
⋮	⋮

gap = K

compute covariance b/w C_1 and C_2

Ex where non-stationary

1. 

2. 

periodicity
* event occurs periodically
Ex:- christmas season

trending
* overall trend
Ex:- inflation

Notations

observations : $y_1, y_2, y_3 \cdots y_{t-1}, y_t, y_{t+1} \cdots y_n$

predictions : $\hat{y}_1, \hat{y}_2 \cdots \hat{y}_{t-1}, \hat{y}_t, \hat{y}_{t+1} \cdots \hat{y}_n$

$\hat{y}_t$ was predicted using data $(y_1 y_2 \cdots y_{t-1})$

errors : $e_K = y_K - \hat{y}_K$

$C = \mu$ = mean of all values seen till now

⟶ used in stationary-time series

ARIMA → moving average

- auto regresive
- Integrating

(48)

AutoRegresion (p) model

→ assumes you can model $\hat{y}_t$ using historic data

AR(p):

$$y_t = \mu + (\alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p}) + \epsilon_t$$

- μ → constant
- $\alpha = [\alpha_1, \ldots, \alpha_p]$ → hyper-param
- ϵ_t → Error term

* to generate o/p at y_t we use data from y_t to y_{t-p}.

$$y_t - \epsilon_t = \mu + \sum_{i=1}^{p} \alpha_i y_{t-p} \Rightarrow \hat{y}_t = c + \sum_{i=1}^{p} \alpha_i y_{t-p}$$

moving average models

→ Same as Linear regresion on prev data

MA(q)

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

$\epsilon_i = y_i - \hat{y}_i$

↳ errors

use linear regressions

- (AR) – model on prev values
- (MA) – model on prev errors

ARMA (p,q)

$$y_t = \mu + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

Non-stationary → stationary

① Differencing

$y_t = y_t' + y_{t-1}$ ...

* $y_t' = y_t - y_{t-1} \equiv \frac{\Delta y}{\Delta x = 0} = \Delta y$
* $y_t'' = y_t' - y_{t-1}'$

$y_t - 2y_{t-1} + y_{t-2}$

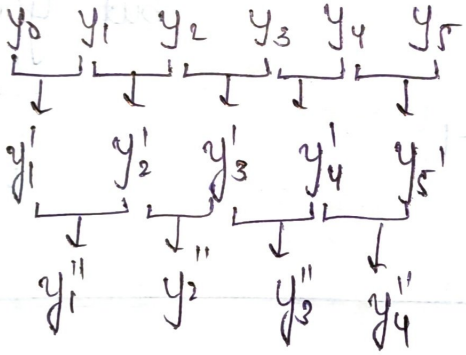
$y_t' = y_t - y_{t-1}$	→ 1st order
$y_t'' = y_t - 2y_{t-1} + y_{t-2}$	→ 2nd order

original :- $y_0\ y_1\ y_2\ y_3\ y_4$

$y_1'\ y_2'\ y_3'$

we can say $y_i' = y_i - y_{i-1}$

$y_i'' = y_i' - y_{i-1}'$



(1st order, d=1)

(2nd order, d=2)

Here, new series formed by d = 1, 2 .. 3 may be stationary

Opposite of differencing → Integrating

ARIMA → hyper-params = p, q, d

AR(p) + I + MA(q)

= ARIMA(p, q, d) → ARIMA is linear model applied on time series post integrate/differentiation.

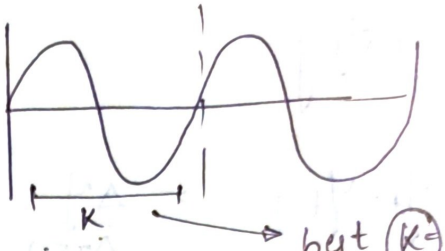
p → no of terms

q → no of errs

past p terms and past q errs

(44) **Auto Corelation** ACF → it can be used for p,q initialization

Why? to find best (K)



(K) is value by which if you shift you will get perfect corelation ie=1.

→ How to calculate

① pick (K)

② construct 2 vectors V_1 and V_2

V_1	y_t	y_{t+1}	y_{t+2}
V_2	y_{t-K}	y_{t-K+1}	y_{t-K+2}

→ cor (V_1, V_2)

③ best K will have corelation → 1 (it will be 1 for K=0, but you cant pick 0)

↳ to predict y_t you have only y_{t-1}, y_{t-2}..

* Best corelation value expected maybe 1 or -1

ARIMA improvements

use features obtained like $y_{t-1}, y_{t-2}, y_{t-3}$

$\epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3}$

+ your own categorical features

Ex: day, holiday

→ better regresion algorithms (DF, GBDT)

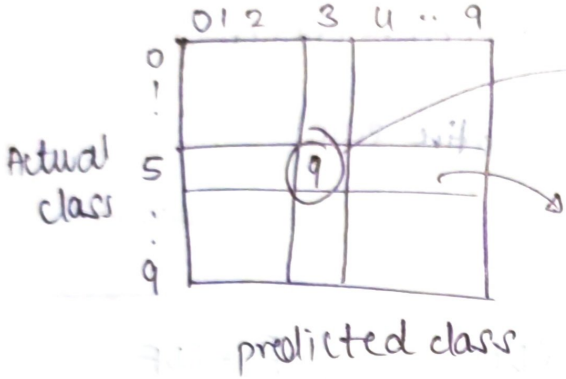
if you ~~ARIMA~~

Linear Regresion = ARIMA

↳ ar

Precision Recall matrix

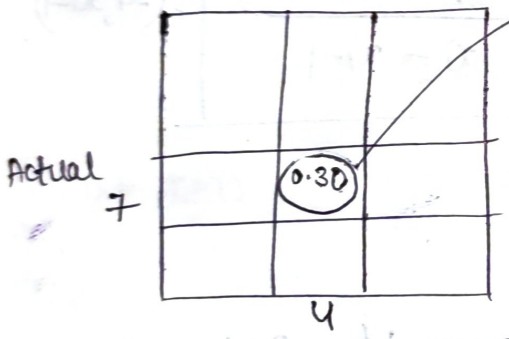
step 1 :- compute a confusion matrix



- 9 values which ∈ to class 3 (5) are predicted to ∈2 class (3)
- sum (row) = total vals which ∈2 class (5)

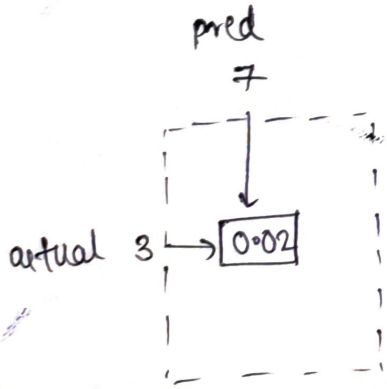
Precision matrix

CT + Colluemn normalized → sum of each col = 1



- Out of all pts predicted to ∈2 class (4) 30% of them actually ∈2 class (7)
- (*) precision → % of ~~out~~ pred^n are correct
- * colluem-up + out of all pred as (P) we have ~~or~~ val % ∈2 actual class (K)

Recall matrix → CT + row-sum = 0



- (*) recall → how many % of actual class were predicted correctly
- row-clean
- 3% of actual class (K) are predicted to ∈2 class (P)

* of all pts that actually ∈2 class (3), 2% are classified as ∈2 (7)

Ideal diagonals = 1 — of all pts ∈2 class (2), 100% are predicted ∈2 (2)

16

Compute PDF given CDF

discrete → prob mass function
continous → PDF

Random variable is discrete

i/p : list of $x_i \rightarrow CDF(x_i)$

o/p find $PDF(x_i)$

① sort acc to x_i

② $PDF(x_i) = CDF(x_i) - CDF(x_{i-1})$

(CDF ≡ cumulative sum)

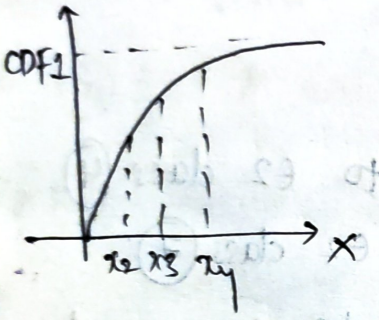
RV is continous

→ sort x_i

⟹ PDF = derivative of CDF

$$\lim_{(x_2 - x_1) \to 0} \frac{F(x_2) - F(x_1)}{x_2 - x_1}$$

F = CDF
f = PDF



So → ∀ x_i, CDF(x_i) in arr:

$$PDF(i) = \frac{CDF(x_i) - CDF(x_{i-1})}{x_i - x_{i-1}}$$

approximation → nearest (x_i, x_{i-1})

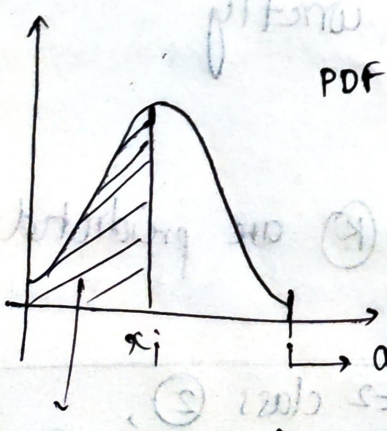
How is CDF calculated?

↳ Integrate PDF

∴ CDF last valis ①

∴ PDF values may be > 1
but PMF values are always b/w 0 to 1

area under whole PDF



area = $P(x \leq x_i) \equiv CDF(x_i)$

0 to 1

Permutation sampling — get p-value

Q: 2 samples of spending of men and women

$S_m = [\ldots\ 100]$

$S_w = [\ldots\ ]$ $n=100$

Question: compare S_m and S_w

metric: $\Delta = |\mu_{men} - \mu_{women}|$

→ observed metric = $|\mu_m - \mu_w| = 0.21$ ← given

Permute + resample

① Null Hypothesis = men and women spending habits are same

② mix $S_m + S_w = [\ldots]$ $n=100$ (as H_0: same)

why? if μ_1 and μ_1 are same (H_0) then samples r also similar

randomly sample S_m and S_w with 100 samples each and calculate metric] repeat (K) times

③ let K=1000 (u get 1000 test-metrics — (ts))

Now,

p-value = $P(X \geq obs) = p(ts > 0.21) = 0.xx$

↳ prob of atleast seeing value equal to or greater than observed

p-val: prob of getting val atleast as extreme as observat=

Result

→ if p-val = $p(ts > 0.21) < 5\%$ → reject H_0 bcoz in our testing less than 5% of ~~met~~ samples had metric atleast ($\geq$) observed

ie prob of observing metric is very low

↳ else — accept H_0

(9) **KS Test** check if ~~dis~~ both rv r from same distribution

↳ parametric

test-statistic

$$D_{n,m} = \sup_x \left| F_{1,n}(x) - F_{2,m}(x) \right|$$

$F_{1,n}$ → CDF value of x_1 with n samples

sup = supremum fn (maximal gap b/w CDF of x_1 and x_2)

*Null hypothesis is rejected at $\alpha = 0.05$

if $D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{nm}}$

$$c(\alpha) = \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}$$

Note:
if m,n are large,
CDF's are almost similar (very less gap)

Exact p-value

$$D\sqrt{\frac{nm}{n+m}} > \sqrt{-\frac{1}{2}\ln\left(\frac{\alpha}{2}\right)}$$

↓ squaring and mul by -1

$$-D^2\left(\frac{nm}{n+m}\right) < \frac{1}{2}\ln\left(\frac{\alpha}{2}\right)$$

$$\exp\left[-2D^2\left(\frac{nm}{n+m}\right)\right] < \alpha$$

$$\therefore \text{p-val} = \exp\left[-2D^2\left(\frac{nm}{n+m}\right)\right]$$

if p-val $< \alpha$, reject

VIF variable inflation factor

Step 1 :- Try to predict (f_i) using other independent features

Step 2 : If they can predict f_i well
→ multi-colinearity

$$VIF(f_i) = \frac{1}{1-R^2}$$

$R^2 = \left(1 - \frac{SS_{res}}{SS_{tot}}\right)$ of model which fits f_i
$(f_i \ldots f_j) \longrightarrow f_i$

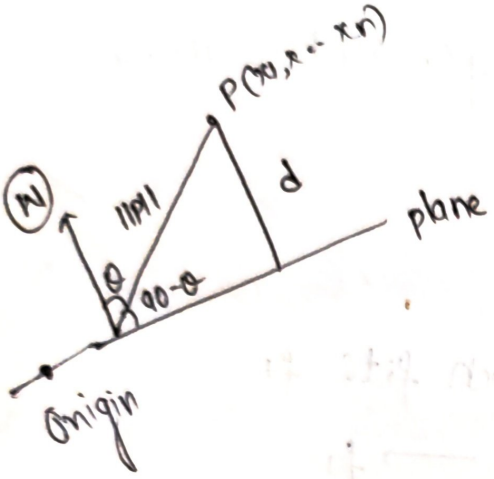
as $R^2 \longrightarrow 1$, better fit

as $R^2 \longrightarrow 1$, VIF ↑

→ VIF = 0 (no colinearity, $R^2 = 0$) doesn't fit

VIF > 10, 15 → multi colinearity

50 Distance of point from plane



$$\sin(90-\theta) = \frac{d}{\|P\|} \Rightarrow \boxed{d = \|P\| \cos\theta}$$

$$w \cdot P = \|w\| \|P\| \cos\theta$$

$$\frac{w \cdot P}{\|w\|} = \|P\| \cos\theta = (d)$$

$$\therefore \boxed{d = \frac{w \cdot P}{\|w\|}}$$

Ex: plane in 2d = line

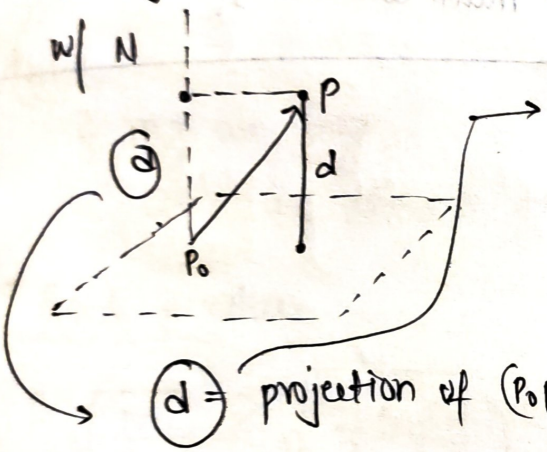
$P(7, 9)$

Line $= 7x + 4y + 120 = 0$

$w = (7, 4) - 120$

$$d = \frac{(7 \cdot 7) + (4 \cdot 9) - 120}{49 + 16}$$

(ii) Projection view



$$\boxed{d = \frac{\vec{w} \cdot \vec{P}}{\|\vec{w}\|}}$$

(d) = projection of $(P_0 P)$ on N/w $= \text{Proj}_{\vec{w}} \vec{P} = d$

Shortest distance b/w 2 parallel plane π_1 π_2

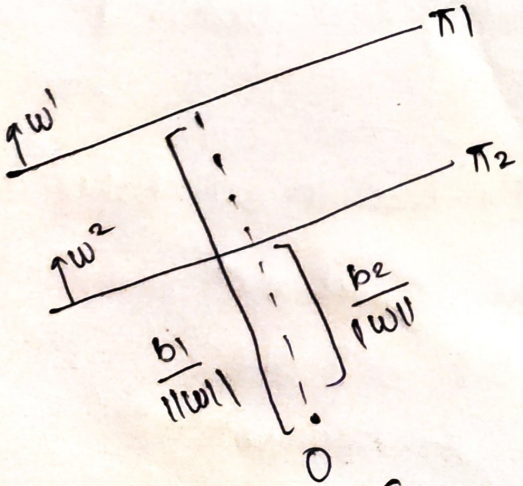
if $\pi_1 \| \pi_2$,

then $w_1 = w_2$,

only intercept/b change

$\longrightarrow$ $\pi_1 : w^T x + b_1$

$\pi_2 : w^T x + b_2$



$$\therefore d(\pi_1 - \pi_2) = d(\text{origin}, \pi_1) - d(\text{org}, \pi_2)$$

$$= \frac{b_1}{\|w\|} - \frac{b_2}{\|w\|} \longleftarrow = \boxed{\frac{\|b_1 - b_2\|}{\|w\|}}$$

$$\left(\frac{|b_1 - b_2|}{\|w\| \|w\|}\right)$$

dist of origin from plane

$\frac{b}{\|w\|}$ → intercept term of π

Eqn of π passing throug origin = $w^T x = 0$

↳ not passing $0(0,0)$ = $w^T x + b = 0$

$d = \frac{w \cdot P + b}{\|w\|}$ dist of point from plane which doesnt pass origin

$P = 0(00\ldots0)$

$d = \frac{w \cdot [0\ldots0] + b}{\|w\|} = \frac{b}{\|w\|}$

∴ shortest dist b/w parallel planes π_1 and π_2 = $\frac{|b_1 - b_2|}{\|w\|}$

R^2 and adjusted R^2

$R^2 = 1 \leftarrow$ Best $(SS_{res} = 0)$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum(x_i - \hat{x})^2}{\sum(x_i - \bar{x})^2}$$

↓

when more features are added → as n_feat ↑ SS_{res} ↓ → R^2 ↑

(whenever u are more feature (R^2) will always increase but maybe the added feature is not correlated with y)

↓ penalize more features

$$R^2_{adj} = 1 - \frac{(1-R^2)(N-1)}{N-p-1}$$

← penalize features when newly added features are not correlated with y.

(N) = no of sample

(P) = no of predictors (features)

↳ no of independent features