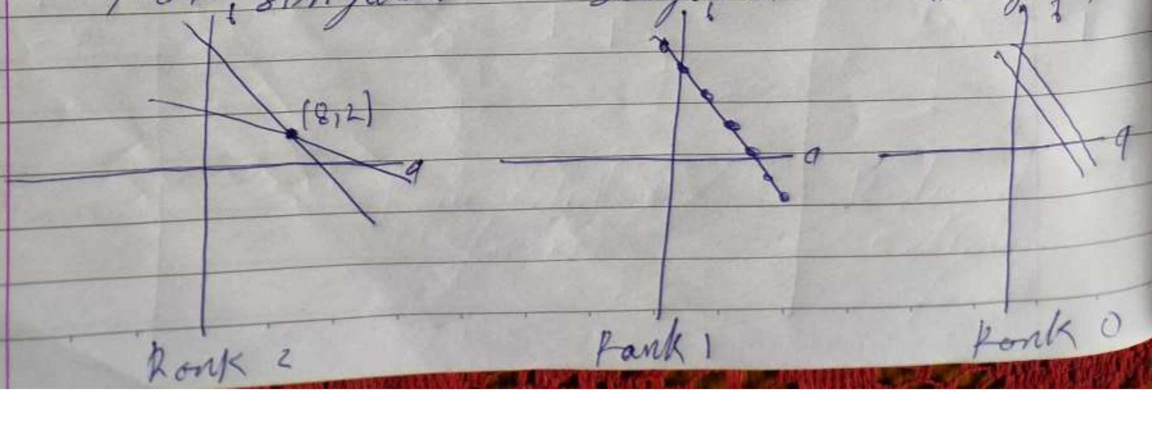
Course-1, Week-1

* System of sentences:

System 1	System 2	System 3
Dog is Black	Dog is Black	Dog is Black
Bird is Red	~~Bird is~~ Dog is Black	Dog is white
complete	Redundant	contradictory
Non-singular	singular	singular
Rank 2	Rank 1	Rank 0

* Systems of equations

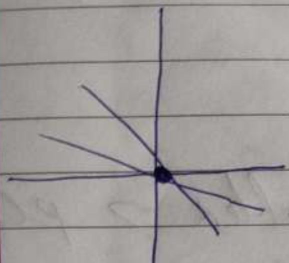
System 1	System 2	System 3
$a + b = 10$	$a + b = 10$	$a + b = 10$
$a + 2b = 12$	$2a + 2b = 20$	$2a + 2b = 24$
Unique solution	Infinite sol	No solution
$a = 8$, $b = 2$	$a = 8, 7, 6 \ldots$, $b = 2, 3, 4 \ldots$	
complete	Redundant	contradictory
Non-singular	singular	singular



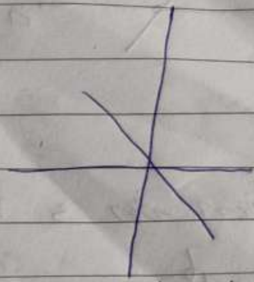
Rank 2   Rank 1   Rank 0

* System of equations as lines

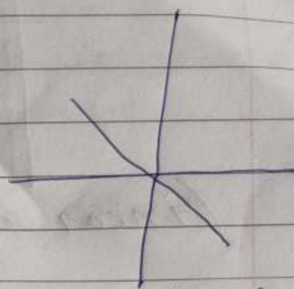
①	②	③
$a + b = 0$	$a + b = 0$	$a + b = 0$
$a + 2b = 0$	$2a + 2b = 0$	$2a + 2b = 0$



complete



Redundant



Redundant

* System of equations as matrices

①

$a + b = 0$
$a + 2b = 0$
non-singular system

$\begin{matrix} 1 & 1 \\ 1 & 2 \end{matrix}$
non-singular matrix

②

$a + b = 0$
$2a + 2b = 0$
singular system

$\begin{matrix} 1 & 1 \\ 2 & 2 \end{matrix}$
singular matrix

↓

second row is depend on first row

second row = first row * 2

↓

linearly dependent

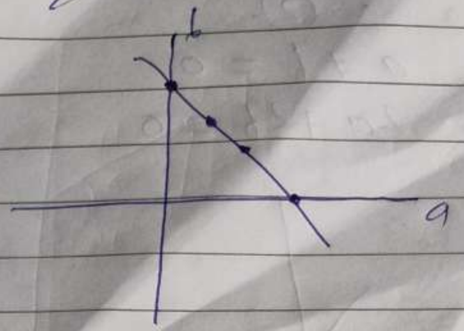
Dependent - Singular
Independent - Non-singular

* Determinant :

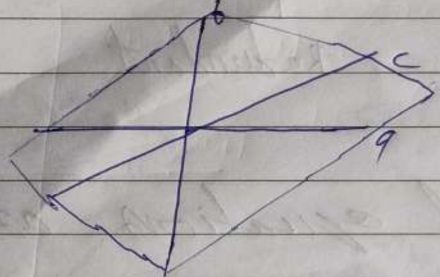
16
non-singular

0
Singular

Linear equation in 2 variables → Line



Linear equation in 3 variables → plane



course-1, Week 2

System of equations to matrices

original system	Intermediate system	solved system
$5a+b=17$	$a+0.2b=3.4$	$1a+0b=3$
$4a-3b=6$	$b=2$	$0a+1b=2$

original matrix	upper diagonal matrix	Diagonal matrix
$\begin{matrix}5 & 1\\ 4 & -3\end{matrix}$	$\begin{matrix}1 & 0.2\\ 0 & 1\end{matrix}$	$\begin{matrix}1 & 0\\ 0 & 1\end{matrix}$
	↓ Row echelon form	Reduced row echelon form

Rank : In row echelon form →
Number of ones in the diagonal

Reduced row echelon form

$$\begin{matrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{matrix}$$

It is in row echelon form

→ each pivot is 1

→ Any number above a pivot is 0

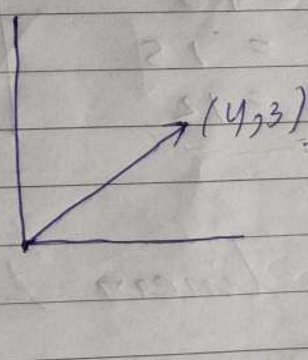
↓

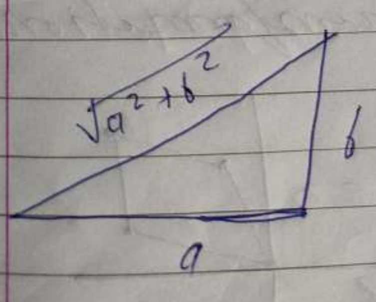
Rank
5

→ Rank of the matrix is the number of pivots

Course-1, Week 3

Vectors



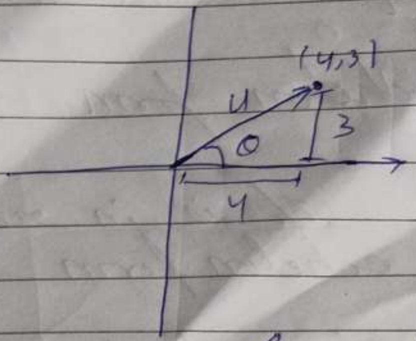


→ pythagorean theorem

L1-norm
$\|(a,b)\|_1 = |a| + |b|$

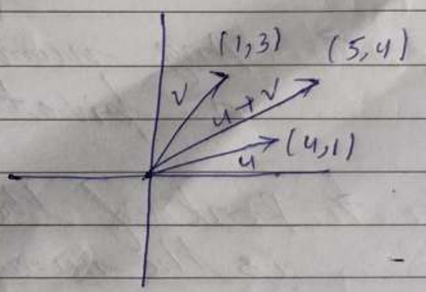
L2-norm
$\|(a,b)\|_2 = \sqrt{a^2+b^2}$

Direction of a vector



$\tan(\theta) = \frac{3}{4}$

Sum of vector



$u + v = (4+1,\ 1+3)$
$(5, 4)$

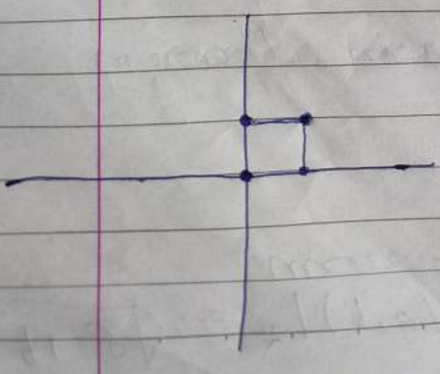
Equation as dot product

system of equations | matrix product

$$\begin{aligned} a + b + c &= 10 \\ a + 2b + c &= 15 \\ a + b + 2c &= 12 \end{aligned}$$

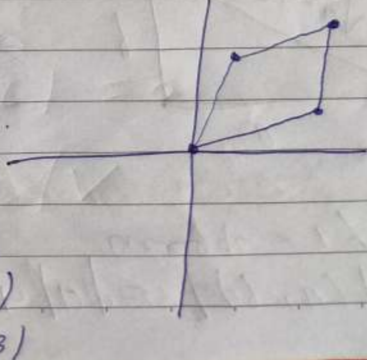
$$\begin{matrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{matrix} * \begin{matrix} a \\ b \\ c \end{matrix} = \begin{matrix} 10 \\ 15 \\ 12 \end{matrix}$$

Matrices as linear transformations



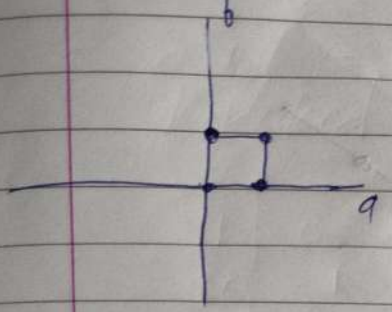
$$\begin{matrix} 3 & 1 \\ 1 & 2 \end{matrix}$$

$(0,0) \to (0,0)$
$(1,0) \to (3,1)$
$(0,1) \to (1,2)$
$(1,1) \to (4,3)$

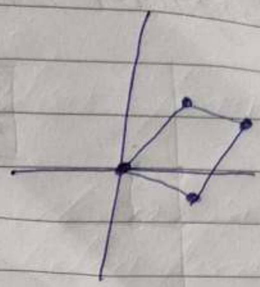


non singular matrices → inverses

Linear transformation as matrices



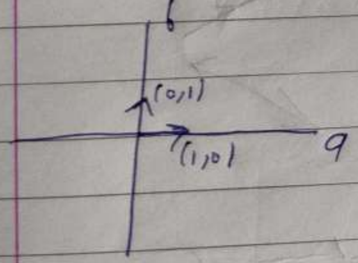
$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

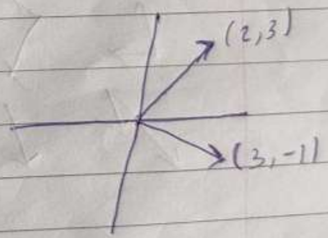


first $(0,0) \to (0,0)$
$(1,0) \to (3,-1)$
$(0,1) \to (2,3)$
$(1,1) \to (5,2)$

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{matrix} 2 \\ 3 \end{matrix}$$

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{matrix} 3 \\ -1 \end{matrix}$$





$$\begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{matrix} 2 \\ 3 \end{matrix}$$

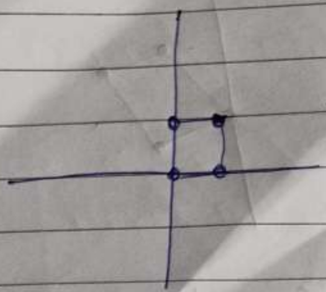
$$\begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{matrix} 3 \\ -1 \end{matrix}$$

$$\begin{bmatrix} 3 & 2 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{matrix} 2 \\ 3 \end{matrix}$$

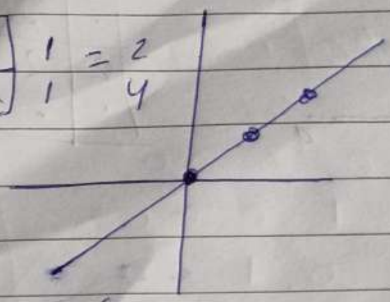
$$\begin{bmatrix} 3 & 2 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{matrix} 3 \\ -1 \end{matrix}$$

course-1 week 4

Singular transformation

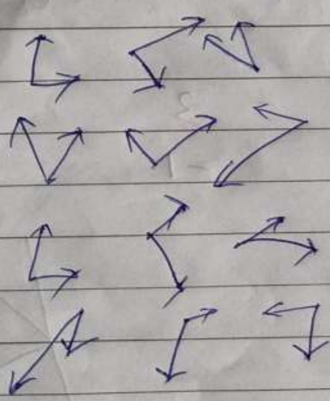


$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

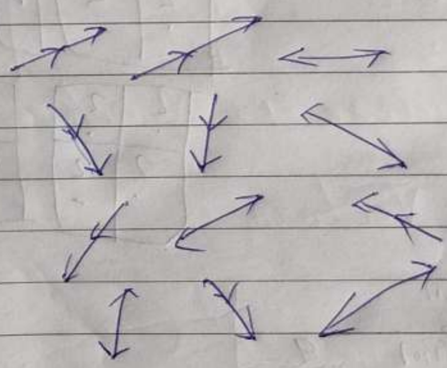


$(0,0) \rightarrow (0,0)$
$(1,0) \rightarrow (1,2)$
$(0,1) \rightarrow (1,2)$
$(1,1) \rightarrow (2,4)$

Bases



Not Bases



Inverse of a matrix $A^{-1} = \frac{1}{|A|} \cdot \text{Adj } A$

eigen vector :

$$|A - \lambda I| = 0$$

$$AV = \lambda V$$

MATHS 4 AI - Course 2 Notes

$$\text{slope} = \frac{\text{height}}{\text{width}}$$

$f(x) = x^2$

$f'(x) = 2x$

$$f'(x) = \frac{\Delta f}{\Delta x}$$

$$\frac{\Delta f}{\Delta x} = \frac{(x+\Delta x)^2 - x^2}{\Delta x}$$

$$= \frac{\cancel{x^2} + 2x\Delta x + \Delta x^2 - \cancel{x^2}}{\Delta x}$$

$$= \frac{\Delta x(2x + \Delta x)}{\Delta x}$$

$f'(x) = 2x + \Delta x$

$\Delta x \to 0$

$f'(x) \to 2x$

$f'(\sin x) = \cos x$

$f'(\cos x) = -\sin x$

$f'(e^x) = e^x$

$\left(1 + \frac{1}{x}\right)^x \to e$

corner | cusps
vertical tangent
jump discontinu—

$$f = g + h$$
$$f' = g' + h'$$

$$f = gh$$
$$f' = g'h + h'g$$

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

$$f(x) = x^2 + y^2$$

$$\frac{d f(x)}{dx} = 2x = 0, \quad x = 0$$

$$\partial y = 2y = 0, \quad y = 0$$

minima

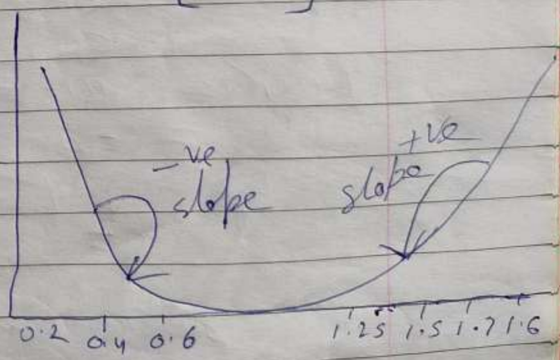
Gradient $\begin{bmatrix} 2x \\ 2y \end{bmatrix}$

Gradient at $(2,3)$ $\begin{bmatrix} 4 \\ 6 \end{bmatrix}$

new point = old point − slope

$$x_K = x_{K-1} - \alpha f'(x_{K-1})$$

one variable



2 variable

define a α

choose a starting point (x_0, y_0)

update:

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} x_{k-1} \\ y_{k-1} \end{bmatrix} - \alpha \nabla f(x_{k-1}, y_{k-1})$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

$$\sigma'(z) = \cancel{d}\, f'\left(\frac{1}{1+e^{-z}}\right)$$

$$= (1+e^{-z})^{-1}$$

$$= -1(1+e^{-z})^{-2}$$

$$= -(1+e^{-z})^{-2}\left(-(e^{-z})\right)$$

$$= (1+e^{-z})^{-2}(e^{-z})$$

$$= (1+e^{-z})^{-2}(e^{-z})$$

$$= \frac{e^{-z}}{(1+e^{-z})^2}$$

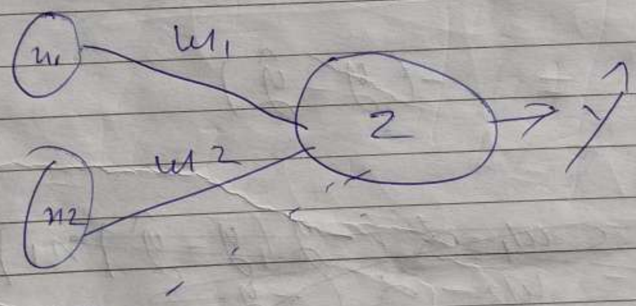
$$= \frac{e^{-z}+1-1}{(1+e^{-z})^2}$$

$$= \frac{1+e^{-z}}{(1+e^{-z})^2} - \frac{1}{(1+e^{-z})^2}$$

$$= \frac{1}{1+e^{-z}} - \left(\frac{1}{1+e^{-z}}\right)\left(\frac{1}{1+e^{-z}}\right)$$

$$= \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right)$$

$$= \sigma(z)\,(1 - \sigma(z))$$



b

$$z = x_1 w_1 + x_2 w_2 + b$$

$$MSE = L = \frac{1}{2}(y - \hat{y})^2$$

$$\frac{dL}{d\hat{y}} \quad f'(L) = -(y - \hat{y})$$

~~dy~~

$$\frac{dL}{dw_1} = \frac{dL}{d\hat{y}} \frac{d\hat{y}}{dw_1}$$

$$\frac{dL}{db} = \frac{dL}{d\hat{y}} \frac{d\hat{y}}{db}$$

$$\frac{dL}{dw_2} = \frac{dL}{d\hat{y}} \frac{d\hat{y}}{dw_2}$$

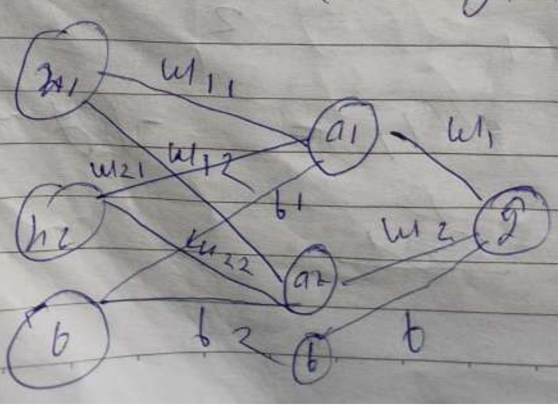
$$\frac{dL}{dw_1} = -(y-\hat{y})x_1$$

$$\frac{dL}{dw_2} = -(y-\hat{y})x_2$$

$$\frac{dL}{db} = -(y-\hat{y})$$

$$\cancel{w_1 = w_1 - \alpha(-y(y-\hat{y})x_1)}$$
$$\cancel{w_2 = w_2 - \alpha(-y}$$

$$\left\{ \begin{aligned} w_1 &= w_1 - \alpha(-(y-\hat{y})x_1) \\ w_2 &= w_2 - \alpha(-(y-\hat{y})x_2) \\ b &= b - \alpha(-(y-\hat{y})) \end{aligned} \right\}$$



Page No.:
Date:

$$a_1 = \sigma(z_1)$$
$$z_1 = x_1 w_{11} + x_2 w_{21} + b_1$$
$$a_2 = \sigma(z_2)$$
$$z_2 = x_1 w_{12} + x_2 w_{22} + b_2$$
$$\hat{y} = \sigma(z)$$
$$z = a_1 w_1 + a_2 w_2 + b$$

$$\frac{dL}{dw_{11}} = \frac{dz_1}{dw_{11}} \cdot \frac{da_1}{dz_1} \frac{dz}{da_1} \frac{d\hat{y}}{dz} \frac{dL}{d\hat{y}}$$

$$L(y, \hat{y}) = \cancel{-y \log}$$
$$-y \log(\hat{y}) - (1-y) \log(1-\hat{y})$$

$$\frac{dL}{d\hat{y}} = \frac{-y}{\hat{y}} - \frac{(1-y)}{(1-\hat{y})}(-1)$$

$$= \frac{-y}{\hat{y}} + \frac{(1-y)}{(1-\hat{y})}$$

$$= \frac{-y(1-\hat{y}) + (1-y)\hat{y}}{\hat{y}(1-\hat{y})}$$

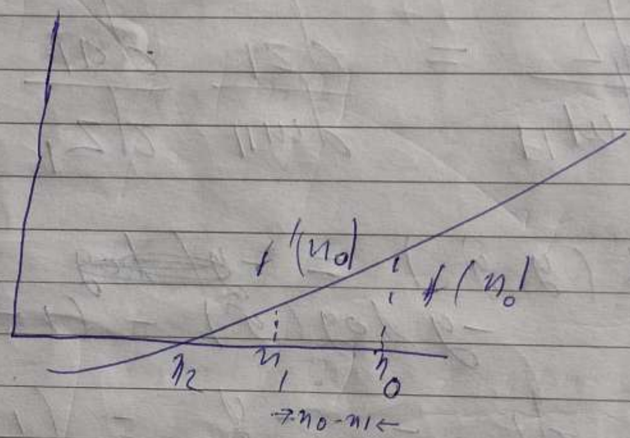
$$= \frac{-y + \cancel{y\hat{y}} + \hat{y} - \cancel{y\hat{y}}}{\hat{y}(1-\hat{y})}$$

$$= \frac{\hat{y} - y}{\hat{y}(1-\hat{y})}$$

$$\frac{dL}{dw_{11}} = x_1 \cdot a_1(1-a_1) \cdot w_1 \cdot \cancel{\hat{y}(1-\hat{y})} \cdot \frac{\hat{y}-y}{\cancel{\hat{y}(1-\hat{y})}}$$

$$\boxed{f'(w_{11}) = x_1 w_1 a_1 (1-a_1)(\hat{y}-y)}$$

$$w_{11} = w_{11} - \alpha \frac{dL}{dw_{11}}$$



$$f'(x) = \frac{f(x)}{x_0 - x_1}$$

$$x_0 - x_1 = \frac{f(x)}{f'(x)}$$

$$x_1 = x_0 - \frac{f(x)}{f'(x)}$$

Newton's method $x_{k+1} = x_k - \dfrac{f(x_k)}{f'(x_k)}$

NM for optimization

→ start with x_0
→ update

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

~~second~~

$$f(x, y) = 2x^2 + 3y^2 - xy$$

$$\frac{d'(x,y)}{d(x)} = 4x - y$$

$$\frac{d'(x,y)}{d(y)} = 6y - x$$

$$4x - y \xrightarrow{x} 4 \qquad 4x - y \xrightarrow{y} -1$$

$$6y - x \xrightarrow{x} -1 \qquad 6y - x \xrightarrow{y} 6$$

Hessian matrix $H = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$

~~f(x,y)~~

$$\begin{tikzcd}
 & & f_{xx}(x,y) \\
 & f_x(x,y) \arrow[ur, "x"] \arrow[dr, "y"] & \\
f(x,y) \arrow[ur, "x"] \arrow[dr, "y"] & & f_{xy}(x,y) \\
 & & f_{yx}(x,y) \\
 & f_y(x,y) \arrow[ur, "x"] \arrow[dr, "y"] & \\
 & & f_{yy}(x,y)
\end{tikzcd}$$

$$H(x,y) = \begin{bmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{bmatrix}$$

→ If eigen values are +ve or >0
(0,0) is a minimum

→ If eigen values are -ve or <0
(0,0) is a maximum

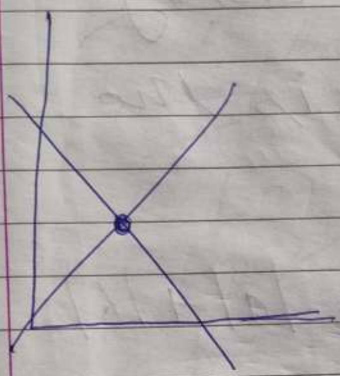
→ = Saddle point any eigen value is -ve

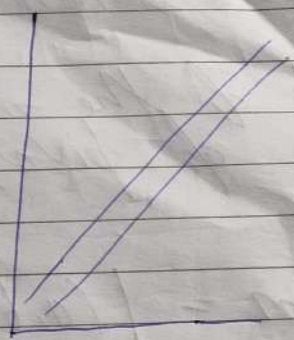
Netwon
$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k)\, \nabla f(x_k, y_k)$$

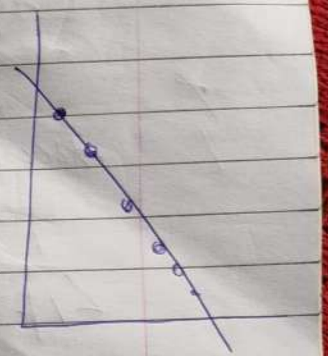
complete → non-singular → unique solution

contradictory → singular → no solution

Redundant → singular → infinite solution







Row echelon form

1	0.6
0	1

Reduced row echelon

1	1
0	1

Rank 2

Rank of matrix → The number of 'ones' in diagonal in row echelon.

0	0
0	1

Rank 1

0	0
0	0

Rank 0

Page No.:
Date:

Row echelon form

2	*	*
0	1	*
0	0	3

Rank → 3

Each row has a pivot
↓
left most non-zero entry

→ Rank of the matrix is the number of pivots.

L_1-norm $= \|(a,b)\|_1 = |a| + |b|$

L_2-norm $= \|(a,b)\|_2 = \sqrt{a^2+b^2}$

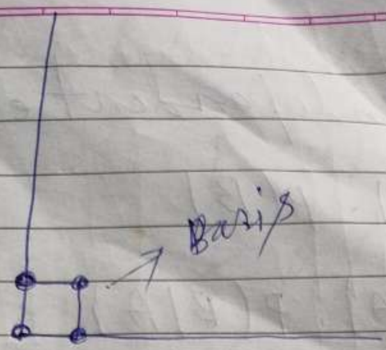
→ matrices as linear transformation

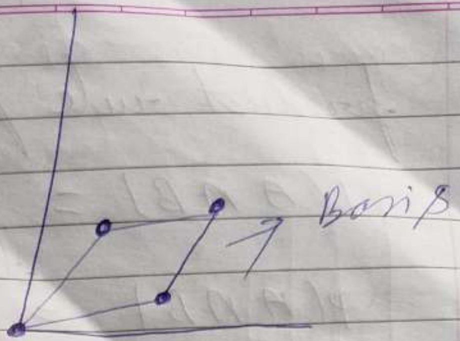
$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\text{''} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

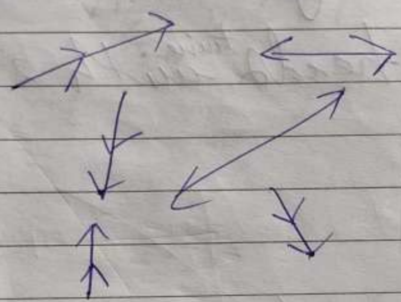
$$\text{''} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\text{''} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$





Not Basis



Inverse of a matrix :

$$A^{-1} = \frac{1}{|A|} \cdot Adj\ A$$

eigen values $|A - \lambda I| = 0$

eigen vector $AV = \lambda V$

---

course-3 stats & prob

$$P'(x) = 1 - P(x)$$

Disjoint $P(A \cup B) = P(A) + P(B)$

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

aba 3
dbacaba 7

product rule for independent events

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

else

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

conditional probability

→ prob of getting heads twice with 2 coin

$\frac{1}{4}$

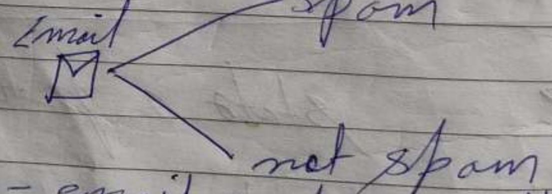
→ with condition the first one is head

$\frac{1}{2}$

Bayes theorem -

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

→ prior →



Event - email contains lottery

→ posterior →

P(spam | lottery)

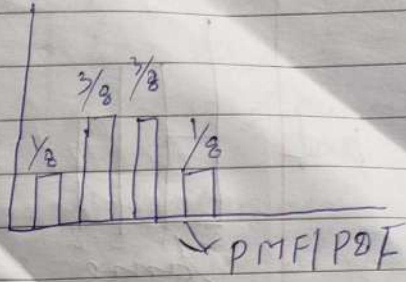
Page No.:
Date:
YOUVA

Discrete probability distribution :

number of heads 3 ~~time~~ coin

$p_x(x) \geq 0$

$\sum p_x(x) = 1$



Binomial distribution :

I flip 5 coins, 2 of them land in head

$^nC_k \, p^k (1-p)^{n-k}$

$\frac{\lfloor 5}{\lfloor 2 \, \lfloor 5-2}$ $\binom{5}{2}$ - Binomial $\binom{n}{k}$

Head Trail

$k = 0, 1, 2 \ldots n$ and $\frac{\lfloor n}{\lfloor k \, \lfloor n-k}$

Bernoulli distribution :

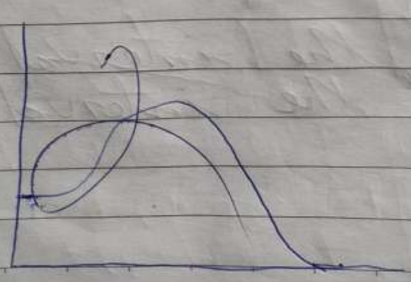
Number of heads flip a coin

$p(X=1) = 0.5$ success $\quad p(X=0) = 0.5$ failure

Discrete to continuous? interval

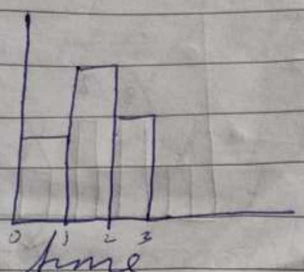
limit

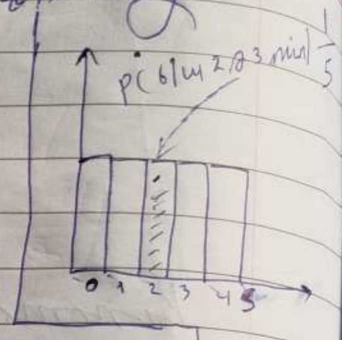
Discrete sum of heights equal 1
continuous : Area under the curve equals 1



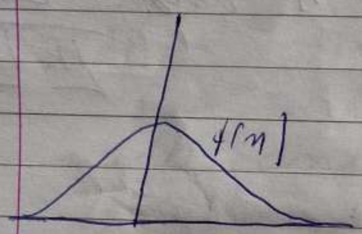

cumulative distributed probability

pdf

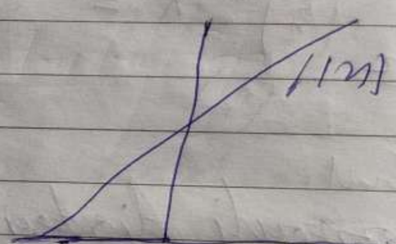


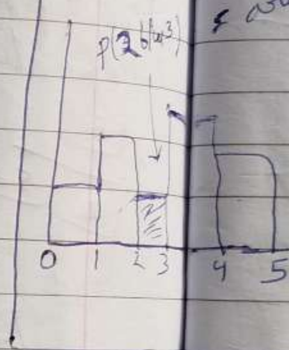


PDF



CDF





area = 1
always +ve

left end point is 0
right end point is 1
always +ve and
increasing

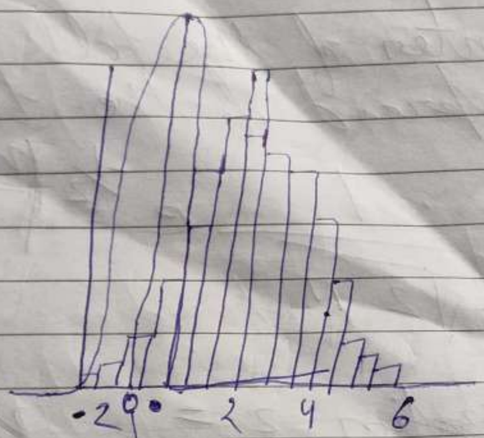
Uniform Distribution :

$$f_x(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x \in (a,b) \end{cases}$$

a : begining of the interval
b : end of the interval

Normal Distribution :-

$e^{-\frac{x^2}{2}}$



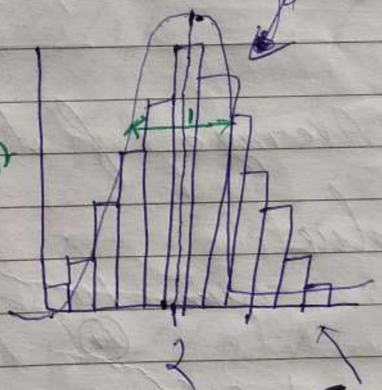
$\mu = 2$

$\mu = 0$

$e^{-\frac{1}{2}(x-2)^2}$

s area

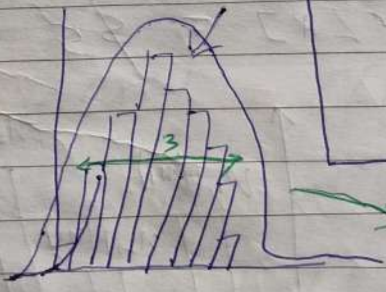
P(exactly at 2 mins) = 0 = area



6 how thick the curve is

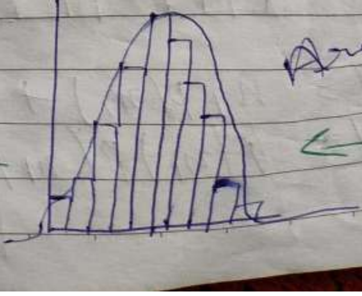
$\sigma = 1$

$\sigma = 3$



Area $\frac{1}{3\sqrt{2\pi}}$

$\frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-2}{3}\right)^2}$

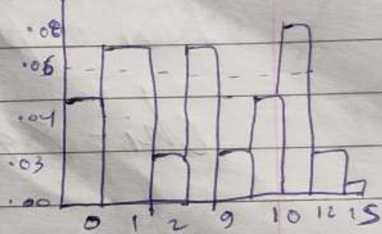


Area 1

uniform distri...

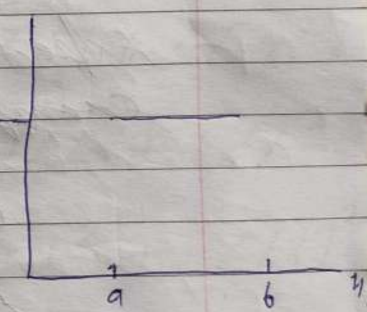
Any value b/w 0 & 15 minutes must have the same frequency of occurrence

The pdf must be constant for all values in the interval (0,15)



which constant

$15 \times h = 1$, $h = 0.06$



parameters:

μ: center of the bell

σ: spread of the bell

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

standardization $z = \dfrac{x-\mu}{\sigma}$

Chi-square distribution

$W = z^2$

Median: Average of the 2 middle ones

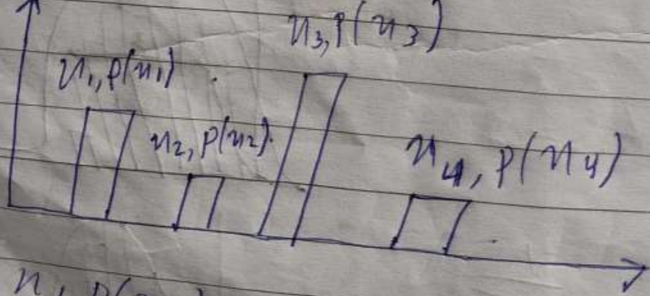
Expected value

example1: win 10 dollars or nothing

(H) (T) 0

$0.5 \times 10 + 0.5 \times 0 = \$5$

$E[X] = \$5$



$E[X] = x_1 p(x_1) + x_2 p(x_2) + x_3 p(x_3) + x_4 p(x_4)$

$E[f(x)] = f(x_1)p(x_1) + f(x_2)p(x_2) + f(x_3)p(x_3) + \cancel{f(x_4)}\, f(x_4)p(x_4)$

Sum of Expectations

(H) win \$1   + , win \$1 \$2 \$3 \$4 \$5 \$6

(T) win nothing

$E[X_{coin}] = \$0.5$, $E[X_{die}] = \$3.5$

$$E[X] = E[X_{coin}] + E[X_{die}]$$
$$= 0.5 + 3.5 = \$4$$

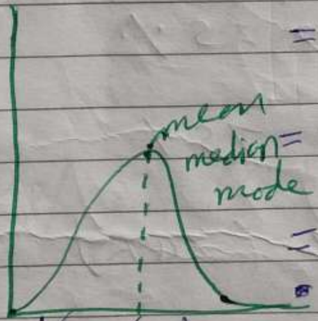
$$\boxed{E[X_1 + X_2] = E[X_1] + E[X_2]}$$

Variance

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

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

$\mu = E[X]$



$$= E[X^2] - 2\mu E[X] + E[\mu^2]$$
$$= E[X^2] - 2E[X]E[X] + \cancel{\mu}^2$$
$$= E[X^2] - 2E[X]^2 + E[X]^2$$

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

$$std(X) = \sqrt{Var(X)}$$

standardize a distribution

$$\frac{X - \mu}{\sigma}$$

Page No.:
Date:

Skewness $E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right]$

$\longrightarrow$ >0 +vely skewed	$\longleftrightarrow$ $=0$ not skewed	$\longleftarrow$ -vely skewed <0

kurtosis

$$E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right]$$

Q_2 Median → second quartile → 50% quantile

Q_1 first quartile → 25%

Q_3 third quartile → 75%

Interquartile range (IQR)

$IQR = Q_3 - Q_1$

* Joint distribution:

$P_{XY}(x, y) = P(X = x, Y = y)$

Total 10 child $P(9, 30) = \frac{1}{10}$

Age X \ Height Y	30	40	50	60
7	0	0	2/10	0
8	3/10	0	0	0
9	1/10	0	0	0

prob

X and Y are independent

$P_{XY}(x, y) = P(x) \cdot P(y)$

* Marginal distribution

Distribution of one variable while ignoring the other.

* Conditional distribution

If age = 9, what is the distribution ~~acr~~ across the height variable

$P_{Y|X=9}(y) = P(Y = y \mid X = 9)$

* divide by row sum

$$P_{Y|X=x}(y) = \frac{P_{XY}(x, y)}{P_X(x)}$$

Joint PDF of X & Y

marginal distribution of x

$$\mathrm{cov}(X, Y) = \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n}$$

equal probabilities

unequal probabilities :

$$\sum P_{XY}(x_i, y_i)(x_i - \mu_x)(y_i - \mu_y)$$

covariance Matrix

$$\begin{bmatrix} Var(X) & cov(X,Y) \\ cov(X,Y) & Var(Y) \end{bmatrix}$$

correlation coefficient →

$$\frac{cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}$$

Multivariate Gaussian Distribution

If W, H were independent

$$f_{HW}(h, w) = f_H(h)\, f_W(w)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_H} e^{-\frac{1}{2}\frac{(h-\mu_H)^2}{\sigma^2_H}} \cdot \frac{1}{\sqrt{2\pi}\,\sigma_W} e^{-\frac{1}{2}\frac{(w-\mu_W)^2}{\sigma^2_W}}$$

$$= \frac{1}{2\pi\sigma_H\sigma_W} e^{-\frac{1}{2}\left(\frac{(h-\mu_H)^2}{\sigma^2_H} + \frac{(w-\mu_W)^2}{\sigma^2_W}\right)}$$

Page No.:
Date:

Course-3

population → the entire group

Sample → subset of the population

population proportion

$$P = \frac{\text{number of items with a given characteristic } (x)}{\text{population } (n)}$$

Sample Variance

$$Var(x) = \frac{\Sigma (x - \bar{x})^2}{n-1}$$

Law of large numbers :-

As the sample size increases the average of the sample will tend to get closer to the average of the entire population

n : number of samples
X_i : some estimate X for a sample size i

as $n \to \infty$

$$\frac{1}{n} \sum_{i=1}^{n} X_i \to E[X] = \mu_x$$

under certain condition :

→ sample is randomly drawn
→ sample size must be sufficiently large
→ independent observation

Central Limit Theorem :

$\mu = np = nP(H)$
$\sigma^2 = np(1-p) = nP(H)P(T)$

* As n increases, the probability distribution becomes closer to a gaussian distribution

$\mu_{\bar{x}} = \mu$ ← population mean

mean of the sample means

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ ← population variance

Variance of the sample means

Maximum Likelihood :

H H H H H T H H T T

Coin 1 | Coin 2 | Coin 3

$P(H) = 0.7$ | $P(H) = 0.5$ | $P(H) = 0.3$
$P(T) = 0.3$ | $P(T) = 0.5$ | $P(T) = 0.7$

coin 1 has the maximum likelihood to get more heads. ~~___~~

$p = P(H)$

Likelihood $L(p;\, 8H) = p^8(1-p)^2$

log likelihood $= \log\left(p^8(1-p)^2\right)$

$= 8\log p + 2\log(1-p)$

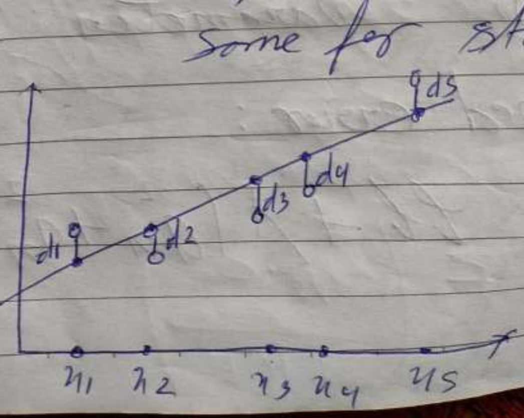
$$\frac{d}{dp}\left(8\log(p) + 2\log(1-p)\right) = \frac{8}{p} + \frac{2}{1-p}(-1) = 0$$

$$\hat{p} = \frac{8}{10}$$

* The best distribution is the one where the mean of the distribution is the mean of the sample.

↓

same for standard deviation



Likelihood :

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_2^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_3^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_4^2} \cdot \frac{1}{\sqrt{2\pi}} e$$

Maximize :

$$e^{-\frac{1}{2}d_1^2} \cdot e^{-\frac{1}{2}d_2^2} \cdot e^{-\frac{1}{2}d_3^2} \cdot e^{-\frac{1}{2}d_4^2} \cdot e^{-\frac{1}{2}d_5^2}$$

$$e^{-\frac{1}{2}(d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2)}$$

minimize :

$$d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2$$

$\downarrow$ Least squares error

Regularization term :

Model : $y = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$

Log-loss : ~~ld~~ ll

L2 regularization error : $a_n^2 + a_{n-1}^2 + \dots + a_1^2$

" parameter : λ

Regularized error :

$$ll + \lambda (a_n^2 + a_{n-1}^2 + \dots + a_1^2)$$

course - 4

Confidence Interval :

lower limit $< \bar{x} <$ upper limit

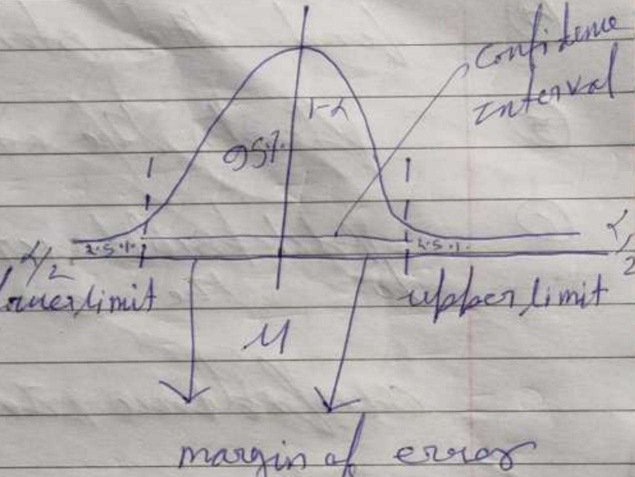
confidence level

$1-\alpha$, $1-0.05$

$0.95\ \%$

$\alpha = 0.05$

Significance level



mean of the sample means $\mu_{\bar{x}}$

standard deviation of the sample means

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

→ As n increases, the confidence interval shrinks.

$$\text{margin of error} = \underbrace{Z_{\alpha/2}}_{\text{critical value}} \cdot \underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{standard error}}$$

Page No.:
Date:

Confidence Interval - calculation s...

STEPS :-

1. find the sample mean
2. Define a desired confidence level (1-...
3. Get the critical value ($Z_{\alpha/2}$)
4. find the standard error ($\sigma/\sqrt{n}$)
5. find the margin of error
6. Add / subtract the margin of error to the sample mean

Confidence Interval =

$$\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Assumptions:

→ Simple random sample
→ Sample size > 30 or population is appro... normal

* Calculating Sample size.

$$n = \left(Z_{\alpha/2} \cdot \sigma \right)^2$$

Confidence Interval - t Distribution

σ ?

$$\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow \text{normal distribution}$$

$$\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \quad \text{student's t} \quad \text{''}$$

unknown σ

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

degrees of freedom : $n-1$

* Confidence Interval for proportions

Confidence interval = $\hat{p} \pm$ margin of error

margin of error = $Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$$p = \frac{x}{n}$$

Hypothesis testing

Motivation :

Null hypothesis H_0 email is ham

Alternative hypothesis H_1 email is spam

→ Not labeling the email spam, doesn't mean the email is ham (~~lack~~ lack of evidence)

→ plenty of evidence against H_0
↓
reject H_0 (and accept H_1)

Type I error : False positive

Type II error : False Negative

* Significance Level

What is the greatest probability of type I error you are willing to tolerate ?

(a ham email is spam 5% time if $\alpha = 0.05$)

0 — 1 → α

Both option are awful decision makers

0 ↓ Emails are always considered ham ↓ no type I error

You want a small α, but not zero

1 ↓ every email is considered spam ↓ every time an email is ham you make a type I error

maximum prob of type 1 error
significance level

catch:

↓ Type I error ↑ Type II error

→ for a given sample, α will determine if you reject H_0 or not.

$\bar{x} = 68.442$
↳ observed statistic

→ The mean height for 18 Y/O in the US in the 70 was 66.7 in.

Test statistic $\bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i$

3 questions — 3 sets of hypothesis

Right - Tailed Test $H_0: \mu = 66.7$ VS $H_1: \mu > 66.7$

Left - Tailed Test $H_0: \mu = 66.7$ VS $H_1: \mu < 66.7$

Two - Tailed Test $H_0: \mu = 66.7$ VS $H_1: \mu \neq 66.7$

null hypothesis
Alternative hypothesis
66.7 μ

Type I error
$\mu > 66.7$
when mean didn't change

If $\bar{x} >> 66.7$
Reject H_0

Type II error
don't reject $\mu = 66.7$ when $\mu > 66.7$

known σ
↓

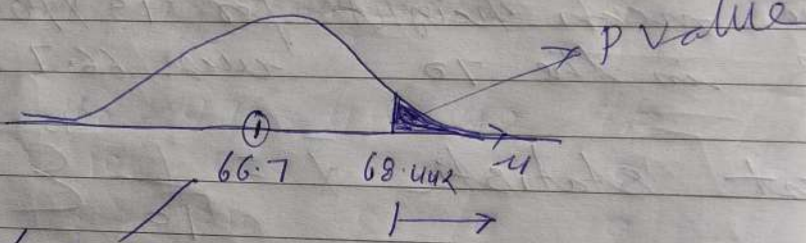
Right-Tailed Test for Gaussian Data

$\sigma = 3, \; n = 10, \; \bar{x} = 68.442$

Goal: Type I error prob $< \alpha = 0.05$

$P(\bar{X} > 68.442 \mid \mu = 66.7) = 0.0407$
$< \alpha$

conclusion: reject H_0
(with a 5% significance level)



null hypothesis

Decision Rule:
If p-value $< \alpha$ reject H_0
If p-value $> \alpha$ don't reject H_0

Z-Statistics

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

If H_0 is true $\quad Z \sim N(0,1)$

Page No.:
Date:

Decision	Reality: H_0 True (Not spam)	H_0 False (spam)
Reject H_0 (Decide spam)	Type I error	correct
Don't reject H_0 (Decide not spam)	correct	Type II error

Steps for performing Hypothesis testing

(1) State your hypotheses

Null hypothesis: the baseline → $H_0: \mu = 66.7$
Alternative ": the statement u want to prove
$H_1: \mu > 66.7$

(2) Design your test

Decide the test statistic to work with $\bar{X}$
Decide the significance level → $\alpha = 0.05$

(3) Compute the observed statistic (based on your sample) → $\bar{x} = 68.442$

(4) Reach a conclusion:
If the p-value is less than the significance level reject H_0
→ $P(\bar{X} > 68.442 \mid \mu = 66.7) > 0.05$

unknown σ^2

t- distribution and T-Statistic

Degrees of freedom (V) = Sample size - 1

$V = 10-1$

$s = \sqrt{\frac{1}{10-1}\sum_{i=1}^{10}(X_i-\bar{X})^2}$ $T = \frac{\bar{X}-\mu}{s/\sqrt{10}} \sim t_9$

Remaining notes:-

Marginal distribution:

		Height						
		45	46	47	48	49	50	
Age	7	1/10	2/10	0	0	0	0	3/10
	8	0	0	2/10	0	0	0	2/10
	9	0	0	0	0	3/10	1/10	4/10
	10	0	0	0	0	0	1/10	1/10
		$\frac{1}{10}$	$\frac{2}{10}$	$\frac{2}{10}$	0	$\frac{3}{10}$	$\frac{2}{10}$	

Marginal distribution of Age

Age (X)	7	3/10
	8	2/10
	9	4/10
	10	1/10

Marginal distribution of height

Height (Y)

45	46	47	48	49	50
1/10	$\frac{2}{10}$	$\frac{2}{10}$	0	$\frac{3}{10}$	[illegible]

$P(Y = 49 \mid X = 9) = \frac{3}{4}$

Conditional PDF

$P(A, B) = P(A) \cdot P(B/A)$

$$P(Y=49 \mid X=9) = \frac{P(X=9, Y=49)}{P(X=9)}$$

$$= \frac{\frac{3}{10}}{\frac{4}{10}} = \frac{3}{4}$$

Covariance ~~of a joint~~ matrix with 3 Variable

	x	y	z
x	Var(x)	cov(x,y)	cov(x,z)
y	cov(x,y)	Var(y)	cov(y,z)
z	cov(x,z)	cov(y,z)	Var(z)

$$\text{Likelihood} = \prod_{i=1}^{n} P_{x_i}(x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\left(\sum_{i=1}^{n} x_i\right)}(1-p)^{\left(n-\sum_{i=1}^{n} x_i\right)}$$

Log likelihood = log (") =

$$\left(\sum_{i=1}^{n} x_i\right)\log(p) + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-p)$$

Page No.:
Date:

Independent Two sample t-Test

Height of 18 y/o in the US:

$n_x = 10$ $\bar{x} = 68.442$

$S_x = 3.113$

Height of 18 y/o in Argentina

$n_y = 9$ $\bar{y} = 65.949$

$S_y = 3.106$

$H_0: \mu_{us} = \mu_{Ar}$ v/s $H_1: \mu_{us} > \mu_{Ar}$

" v/s $H_1: \mu_{us} < \mu_{Ar}$

" v/s $H_1: \mu_{us} \neq \mu_{Ar}$

$$* \quad \frac{(\bar{x}-\bar{y}) - (\mu_{us} - \mu_{Ar})}{\sqrt{\frac{\sigma^2_{us}}{10} + \frac{\sigma^2_{Ar}}{9}}} \sim N(0,1)$$

we don't know σ_{us}, σ_{Ar}, replace σ_{us}, σ_{Ar} with $S^2_{\bar{x}}, S^2_{\bar{y}}$.

Right tailed test

$H_0: \mu_{us} - \mu_{Ar} = 0$ v/s $H_1: \mu_{us} - \mu_{Ar} > 0$

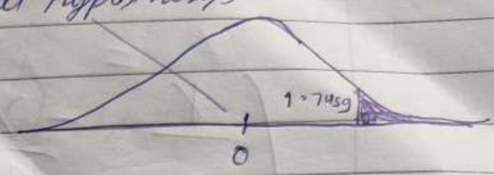
$\alpha = 0.05$

$$T = \frac{(\bar{x}-\bar{y}) - 0}{\sqrt{\frac{S^2_x}{10} + \frac{S^2_y}{9}}} = 1.7459$$

p-value $0.0495 < 0.05$

reject H_0

null hypothesis



paired t-Test

$$\frac{(x_1 - y_1) + (x_2 - y_2) + \cdots + (x_{10} - y_{10})}{10}$$

$$\bar{D} = \frac{D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_9 + D_{10}}{10}$$

$$\frac{\bar{D} - \mu_D}{\sigma_D / \sqrt{10}} \sim N(0,1)$$

σ_D is unknown

$$\sigma_D \to S_D \to \sqrt{\frac{\sum_{i=1}^{10} (D_i - \bar{D})^2}{10-1}} \to$$

$$T = \frac{\bar{D} - \mu_D}{S_D / \sqrt{10}} \sim t_{10-1}$$

Page No.:
Date:

A/B Testing : Conversion Rates

If H_0 is true $\rightarrow P_A = P_B = P$

$$\frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - (P - P)}{\sqrt{\frac{P(1-P)}{n_A} + \frac{P(1-P)}{n_B}}} \sim N(0,1)$$

$\downarrow$

$$P(1-P)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)$$

$$= P(1-P)(n_A + n_B)\frac{1}{n_A n_B}$$

$$\frac{\left(\frac{X}{n_A} - \frac{Y}{n_B}\right) - 0}{\sqrt{(n_A + n_B)P(1-P)}}\sqrt{n_A n_B} \overset{a}{\sim} N(0,1)$$

You don't know P

replace it by estimation $\hat{P} = X + Y$

Test statistic

$$\frac{\left(\frac{x}{n_A} - \frac{y}{n_B}\right) - 0}{\sqrt{(x+y)\left(1-\frac{x+y}{n_A+n_B}\right)}} \sqrt{n_A n_B} \overset{a}{\sim} N(0,1)$$

$H_0 : p_A - p_B = 0$ v/s $H_1 : p_A - p_B < 0$

(P_B has high conversion rate)

$n_A = 80 \quad n_B = 20 \quad \alpha = 0.05$

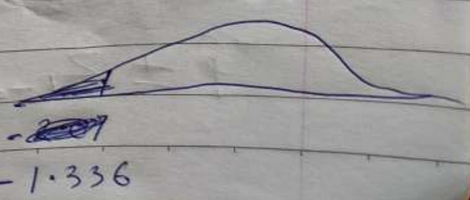
$x = 20 \quad y = 8$

$$z = \frac{\left(\frac{20}{80} - \frac{8}{20}\right) - 0}{\sqrt{(20+8)\left(1 - \frac{20+8}{80+20}\right)}} \sqrt{80 \cdot 20}$$

$z = -1.336$

p-value $= 0.091$

Don't reject H_0



assignment -3

t test for 2 independent sample

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{\sqrt{n_1}} + \frac{s_2^2}{\sqrt{n_2}}}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\sqrt{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}}$$

z test for 2 Independent sample

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\hat{p} = \frac{n_1 + n_2}{n_2 + n_2}$$

~~$\hat{p} \pm z_{\alpha/2}$~~

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{\ }}$$