Lecture 4

To check whether it is really multivariate normal distribution or not

Normality

$$X = \begin{array}{c|cccc} & X_1 & X_2 & & X_p \\ x_1. & x_{11} & x_{12} & \cdots & x_{1p} \\ x_2. & x_{21} & x_{22} & & x_{2p} \\ \vdots & \vdots & & & \\ x_n. & x_{n1} & x_{n2} & - & x_{np} \end{array}$$

here every column is also normally distributed

$$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \sim N_p(\underset{\sim}{\mu}, \Sigma)$$

$$a' = [a_1, a_2, \dots a_p].$$

$$a'\underset{\sim}{X} \sim N(a'\mu, a'\Sigma a)$$

here we get n $a'\underset{\sim}{X}$

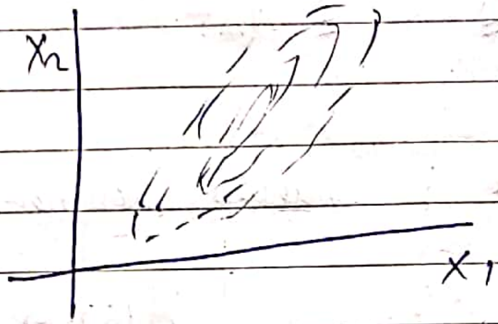
Marginal distribution will be univariate normal.
means single variables.

if X_1 follows ~~multi~~ univariate normal and X_2 also follows same, then

X_1 & X_2 may not follow the same

i.e X_1 & X_2 may not be multivariate normal.

if $x_1, x_2 \dots x_p$ are univariate normal then distribution may follow multivariate normal distribution, (necessary but not sufficient)



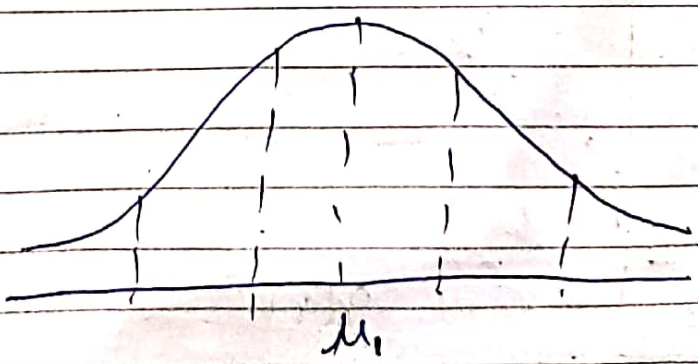
In scatter plot if it's elliptical means it is nomally distributed

google → bivariate normal scatter plot

$$X_1 \to n \quad (\bar{x}_1 - \sqrt{s_{11}},\ \bar{x}_1 + \sqrt{s_{11}})$$

$$(\mu_1 - \sqrt{\sigma_{11}},\ \mu_1 + \sqrt{\sigma_{11}}) \to 0.683.$$

$$(\mu_1 - 2\sqrt{\sigma_{11}},\ \mu_1 + 2\sqrt{\sigma_{11}}) \to 0.954$$



Date:
Page No.

Q-Q plots

$x_1, x_2, \dots, x_n$ — n data points for 1 variable

Ordered:

$x_{(1)}, x_{(2)}, \dots, x_{(n)}$

$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

$x_{(j)}$ there are j values less than or equal to $x_{(j)}$

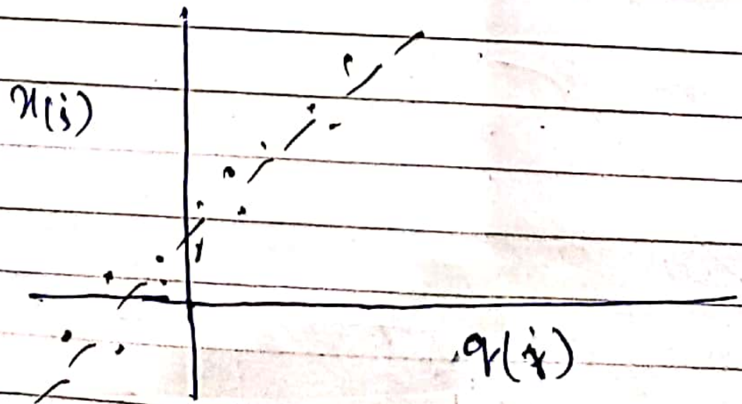
proportion $\frac{j-1/2}{n}$

$x_{(j)}$ is quantile

$\boxed{\sigma q_{(j)} + \mu}$

$q_{(j)}$ $P(Z \leq q_{(j)}) = \frac{j-1/2}{n}$ → find from table

if we plot $q_{(j)}$ and $x_{(j)}$ they should fall in a straight line.



Scanned with CamScanner

Date:
Page No.

$x_{(j)}$	$\frac{j-1/2}{n}$	$q_{(j)}$
-1.00	0.05	-1.645
-0.10	0.15	-1.036
$.16$	0.25	-0.674
$.41$		
$.62$		
⋮	⋮	
	0.95	

Check for multivariate Normality

$x_{1\cdot}, x_{2\cdot}, x_{3\cdot}, \cdots, x_{n\cdot}$

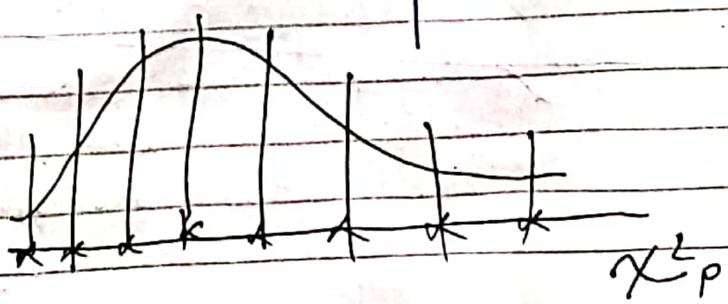
$d_j^2 = (x_{j\cdot} - \bar{\underset{\sim}{x}})' S^{-1} (x_{j\cdot} - \bar{\underset{\sim}{x}}) \Rightarrow$ squared statistical distance

$d_1^2, d_2^2, \cdots, d_n^2$

ordered $d_{(1)}^2, d_{(2)}^2, \cdots, d_{(n)}^2$

$\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$

$(\underset{\sim}{X} - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{X} - \underset{\sim}{\mu}) \sim \chi_p^2$



C → Chi Squared
p → p degrees of freedom

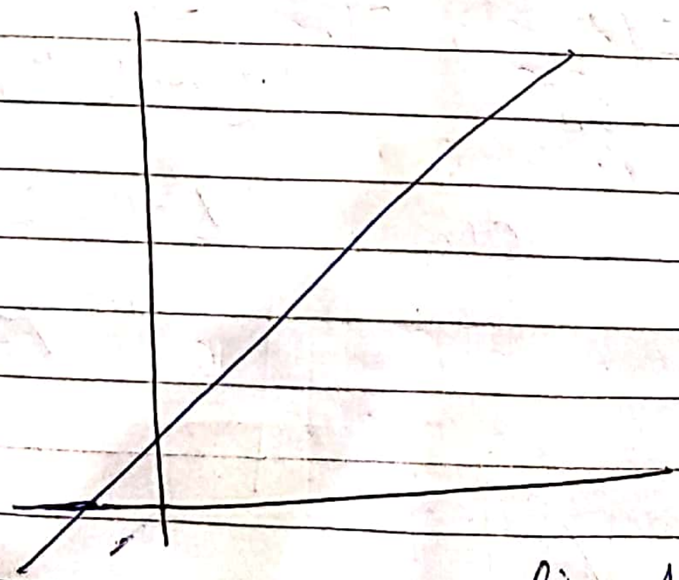
$$\left(q_{c,p}\left(\left(\frac{j-1/2}{n}\right), d_{(j)}^2\right)\right)$$

$$q_{c,p}\left(\frac{j-1/2}{n}\right) = \chi_p^2\left(\left(n-j+1/2\right)/n\right)$$

α $1-\alpha$

$p=2$.

j	$d_{(j)}^2$	$q_{c,2}(j-1/2/10)$
1	0.3	0.10
2	0.62	0.33.
3	1.16	⋮
4	⋮	⋮
5	⋮	
⋮		
10	4.38.	5.99



line should be similar to straight line.