Chapter-4

Lecture note:-

→ univariate normal distribution को लागि pdf

pdf: $$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

CDF: $$P(X \le x) = \Phi(x) = \int_{-\infty}^{x} \phi(t)\,dt$$

→ normal distribution symmetric भएको हुनाले

$$\Phi(-x) = 1 - \Phi(x)$$

→ Test for normality:

Shapiro wilk, Kolmogorov smirnov

→ pdf for a p-dimensional MVN random vector $\underset{\sim}{X} = (X_1, \ldots, X_p)$ is of the

$$f(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu})\right\}$$

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

→ when $p=2$, we have bivariate normal distribution

→ pdf for bivariate normal distribution is

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}} \exp\left\{\frac{-1}{2(1-\rho_{12}^2)}\left[\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2 + \left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right\}$$

Properties of MVN distribution:-

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

① Any Linear combination of variables $\underset{\sim}{a}^T \underset{\sim}{x} = a_1 X_1 + \dots + a_p X_p$ is said to follow $N(\underset{\sim}{a}^T \underset{\sim}{\mu}, \underset{\sim}{a}^T \Sigma \underset{\sim}{a})$

proof

② Conversly If $\underset{\sim}{a}' \underset{\sim}{X} \sim N(\underset{\sim}{a}^T \underset{\sim}{\mu}, \underset{\sim}{a}^T \Sigma \underset{\sim}{a})$ for every a, then $\underset{\sim}{X}$ must be $N_p(\underset{\sim}{\mu}$

proof

③ For q linear combination $\underline{A}\,\underset{\sim}{X}$, if $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$

$$\underline{A}\,\underset{\sim}{X} \sim N_p(\underline{A}\,\underset{\sim}{\mu}, \underline{A}\,\Sigma\,\underline{A}^T)$$

④ $\underset{\sim}{X}_{p\times1} + \underset{\sim}{d}_{p\times1} \sim N(\underset{\sim}{\mu} + \underline{d}, \Sigma)$, where d is constant vector.

⑤ All subset of $\underset{\sim}{X}$ are multivariate normally distributed.

Eg:- If we partition $\underset{\sim}{X}$, we know that its mean vector and covariance matrix will be

$$\underset{\sim}{X} = \begin{bmatrix} X^{(1)} \\ \hline X^{(2)} \end{bmatrix}_{p-q}^{q}, \quad \underset{\sim}{\mu} = \begin{bmatrix} \mu^{(1)} \\ \hline \mu^{(2)} \end{bmatrix}_{p-q}^{q}, \quad \Sigma = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]$$

Then $X^{(1)} \sim N_q(\underset{\sim}{\mu}^{(1)}, \Sigma_{11})$

↳ If $\underset{\sim}{X}^{(1)}, X_2^{(2)}$ are independent then $Cov(\underset{\sim}{X}^{(1)}, X^{(2)}) = 0_{q_1 \times q_2}$

↳ If $\underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{X_1} \\ \hline \underset{\sim}{X_2} \end{bmatrix}_{q_2}^{q_1} \sim N_{q_1+q_2}\left(\begin{bmatrix} \mu_1 \\ \hline \mu_2 \end{bmatrix}, \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]\right)$

then $\underset{\sim}{X_1}$ and $\underset{\sim}{X_2}$ are independent iff $\Sigma_{12} = \Sigma_{21}^T = 0$

↳ मतलब covariance matrix $\Sigma_{12} = \Sigma_{21}^T = 0$, zero है अगर then $\underset{\sim}{X_1}$ and $\underset{\sim}{X_2}$ are independent

Recall

If $Z_1, Z_2, \dots, Z_p$ are independent $N(0,1)$, then

$$\sum_{i=1}^{p} Z_i^2 \sim \chi^2_{(p)}$$

↳ Univariate का जहाँ जहाँ squre लिखेंगे, multivariate का that squre is replaced by उसी vector * उसका Transpose (Eg $Z^2 \Longleftrightarrow Z Z^T$)

↳ मतलब मलाई Z^2 लेख्नु छ भने I will write ZZ^T.

→ So In general, If $\underset{\sim}{Z} \sim N_p(0, I)$ then $ZZ^T \sim \chi^2_{(p)}$

<u>proof</u> since $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$

$$\Rightarrow \underset{\sim}{X} - \underset{\sim}{\mu} \sim N_p(\underset{\sim}{0}, \Sigma)$$

So, Let $$Z = \frac{\underset{\sim}{X} - \underset{\sim}{\mu}}{(\Sigma)^{1/2}} \sim N(\underset{\sim}{0}, I)$$

note

$$= (\underset{\sim}{X} - \underset{\sim}{\mu}) \Sigma^{-1/2} \sim N(\underset{\sim}{0}, I)$$

So Z^2

$$Z^T Z = [(\underset{\sim}{X} - \underset{\sim}{\mu}) \Sigma^{-1/2}]^T [(\underset{\sim}{X} - \underset{\sim}{\mu}) \Sigma^{-1/2}]$$

$$= (\underset{\sim}{X} - \underset{\sim}{\mu})^T \Sigma^{-1/2} \Sigma^{-1/2} (\underset{\sim}{X} - \underset{\sim}{\mu})$$

$$= (\underset{\sim}{X} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{X} - \underset{\sim}{\mu}) \sim \chi^2_{(p)}.$$

Sampling distribution of $\bar{X}$ and S.

In univariate case

1) $X \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

2) $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$

$$E\left[\frac{(n-1)s^2}{\sigma^2}\right] = E(\chi^2_{(n-1)})$$

$$\frac{\cancel{(n-1)}}{\sigma^2} E(s^2) = \cancel{(n-1)}$$

$$E(s^2) = \sigma^2$$

$\hookrightarrow$ $(n-1)s^2 \sim \sigma^2 \chi^2_{(n-1)}$ where $s^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1}$

In multivariate

3. If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then

a) $\underset{\sim}{\bar{X}} \sim N_p(\underset{\sim}{\mu}, \frac{1}{n}\Sigma)$ $\boxed{cov(\underset{\sim}{\bar{X}}) = \frac{1}{n}\Sigma \text{ prove}}$

b) $(n-1)\underline{\underline{S}} \sim W_{(n-1)}$ Wishart distribution

c) $\underset{\sim}{\bar{X}}$ and $\underset{\sim}{S}$ are independent (univariate का case मा easily देखाउन सकिन्छ just by ~~two~~ showing $s^2 = f(X_i - \bar{X})$)

Two Important theorem:-

1.) Law of Large number

$Y_1, \ldots, Y_n$ are independent observation from population with $E(Y_i) = \mu$. Then $\bar{Y} \xrightarrow{P} \mu$ i.e $P(|\bar{Y} - \mu| < \varepsilon) = 0 \ \forall \varepsilon > 0$

wrong

As a sequence of law of large number, we can say

$$S^2 \xrightarrow{P} \sigma^2 \quad \text{or} \quad \frac{1}{n}\Sigma(X_i-\bar{X})^2 \xrightarrow{P} \sigma$$

For the multivariate case

a) Each $\bar{X}_i \xrightarrow{P} \mu_i$, $i=1,\dots,p$ so that

$$\bar{\underset{\sim}{X}}_i \xrightarrow{P} \underset{\sim}{\mu}$$

b) Each sample covariance $S_{ik} \xrightarrow{P} \sigma_{ik}$, $k=1,2,\dots,p$ so that $\underline{\underline{S}} \xrightarrow{P} \Sigma$

② Central Limit Theorem :

Let $\underset{\sim}{X_1},\dots,\underset{\sim}{X_n}$ be independent observation from a population with mean $\underset{\sim}{\mu}$ and finite (non singular) covariance Σ, then

noted

$$\sqrt{n}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}) \sim N_p(0,\Sigma) \text{ when } n \text{ is Large}$$

It follows from here that

$$n(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu})\Sigma^{-1}(\bar{\underset{\sim}{X}}-\underset{\sim}{\mu}) \sim \chi^2_{(p)}$$

proof If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu},\Sigma)$. Then

$$\bar{\underset{\sim}{X}} \sim N_p(\underset{\sim}{\mu},\tfrac{1}{n}\Sigma)$$

Sol $(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}) \sim N_p(0, \frac{1}{n}\Sigma)$

$\sqrt{n}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}) \sim N_p(\underset{\sim}{0}, \Sigma)$

Let $\underset{\sim}{Z} = \sqrt{n}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\Sigma^{-1/2} \sim N_p(\underset{\sim}{0}, I)$

Then $Z^T . Z = [\sqrt{n}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\Sigma^{-1/2}]^T[\sqrt{n}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})\Sigma^{-1/2}]$

$= n(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu})^T \Sigma^{-1}(\underset{\sim}{\bar{X}} - \underset{\sim}{\mu}) \sim \chi^2_{(p)}$

Properties of Wishart Distribution:

→ Multivariate मा χ^2-distribution को analogues हुनेको Wishart Distribution हो।

1) If $Z_1, Z_2, \ldots, \underset{\sim}{Z}_m \overset{iid}{\sim} N_p(\underset{\sim}{0}, \Sigma)$. Then

$$\sum_{j=1}^{m} \underset{\sim}{Z}_j Z_j^T \sim W_m(\Sigma)$$

2) If $A_1 \sim W_{m_1}(\Sigma)$ independent of $A_2 \sim W_{m_2}(\Sigma)$. then

$$A_1 + A_2 \sim W_{m_1+m_2}(\Sigma)$$

3.) If $A \sim W_m(\Sigma)$, then $CAC^T \sim W_m(C\Sigma C^T)$

Book

If $\underset{\sim}{X} \rightarrow N_p(\underset{\sim}{\mu}, \Sigma)$. Then

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

mean $E(\underset{\sim}{a}', \underset{\sim}{X}) = \underset{\sim}{a}'\underset{\sim}{\mu}$

$var(\underset{\sim}{a}', \underset{\sim}{X}) = \underset{\sim}{a}'\Sigma\underset{\sim}{a}$.

$\therefore \underset{\sim}{a}'\underset{\sim}{X} \rightarrow N_p(\underset{\sim}{a}'\underset{\sim}{\mu}, \underset{\sim}{a}'\Sigma\underset{\sim}{a})$

Book Example

(The equivalence of zero covariance and independence for normal variable).

Let $\underset{3\times 1}{\underset{\sim}{X}} \sim N_3(\underset{\sim}{\mu}, \Sigma)$ with $\Sigma = \begin{array}{c} \quad\;\; \sigma_{12} \\ \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} \end{array}$

Are X_1 and X_2 independent? What about (X_1, X_2) and X_3?

Since X_1 and X_2 have covariance $\sigma_{12} = 1$, they are not independent. However partitioning X and Σ as

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \hline X_3 \end{bmatrix} \quad \text{and} \quad \Sigma = \left[\begin{array}{cc|c} 4 & 1 & 0 \\ 1 & 3 & 0 \\ \hline 0 & 0 & 2 \end{array}\right]$$

we see that $X_1 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and X_3 have covariance matrix $\Sigma_{12} \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Therefore (X_1, X_2) and X_3 are independent. This implies X_3 is independent of X_1 and also of X_2.

* The End *

Q.) Given $\underset{\substack{n \times p \\ 3 \times 3}}{\overline{X}} = \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$. Find the following $\boxed{3}$

a.) $\underset{\sim}{\overline{X}}_{3 \times 1}$

b.) S_n

c.) Consider the linear Combinations $2X_1 + 2X_2 - X_3$ and $X_1 - X_2 + 3X_3$

i.) Find their means and variances respectively

ii.) Determine $Cov(2X_1 + 2X_2 - X_3 , X_1 - X_2 + 3X_3)$

Solu

a.) $\underset{\sim}{\overline{X}} = \frac{1}{n} \overline{X}^T \mathbb{1}$

$$= \frac{1}{3} \begin{bmatrix} 1 & 4 & 4 \\ 2 & 1 & 0 \\ 5 & 6 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \frac{1}{3} \begin{bmatrix} 9 \\ 3 \\ 15 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$$

b.) $S_n = \frac{1}{n} \underline{X}^T \left(I - \frac{1}{n} \mathbb{1}\mathbb{1}^T\right) \underline{X}$

$$= \frac{1}{3}\begin{bmatrix} 1 & 4 & 4 \\ 2 & 1 & 0 \\ 5 & 6 & 4 \end{bmatrix} \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \right\} \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} 1 & 4 & 4 \\ 2 & 1 & 0 \\ 5 & 6 & 4 \end{bmatrix} \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \right\} \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} 1 & 4 & 4 \\ 2 & 1 & 0 \\ 5 & 6 & 4 \end{bmatrix} \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \right\} \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} 1 & 4 & 4 \\ 2 & 1 & 0 \\ 5 & 6 & 4 \end{bmatrix} \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} -2 & 1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 5 \\ 4 & 1 & 6 \\ 4 & 0 & 4 \end{bmatrix}$$

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

c) Let $\underset{\sim}{b}^T \underset{\sim}{X} = 2X_1 + 2X_2 - X_3$

$$= \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

and $\underset{\sim}{c}^T \underset{\sim}{X} = X_1 - X_2 + 3X_3$

$$= \begin{bmatrix} 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$$

Sample mean of $\underset{\sim}{b}^T \underset{\sim}{X} = \underset{\sim}{b}^T \underset{\sim}{\bar{X}}$

$$= \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}$$

$$= 3$$

Sample Variance of $\underset{\sim}{b}^T \underset{\sim}{X} = \underset{\sim}{b}^T S_n \underset{\sim}{b}$

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

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

Sample variance of $\underset{\sim}{c}^T \underset{\sim}{X} = \underset{\sim}{c}^T S_n \underset{\sim}{c}$

$$= (1 \ -1 \ 3) \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 3 \\ -2/3 \\ 5/3 \end{bmatrix}$$

$$= 26/3$$

Sample Covariance of $\underset{\sim}{b}^T \underset{\sim}{X}$ and $\underset{\sim}{c}^T \underset{\sim}{X}$

$$\text{cov}(\underset{\sim}{b}' \underset{\sim}{X}, \underset{\sim}{c}' \underset{\sim}{X}) = \underset{\sim}{b}' S_n \underset{\sim}{c}$$

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

$$= \begin{bmatrix} 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ -2/3 \\ 5/3 \end{bmatrix}$$

$$= 3$$