MTL 766

08/08/2025

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix} \qquad \bar{X} = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{bmatrix}$$

$$S = [s_{ij}] \; ; \; s_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j)$$

(a) $E[\bar{X}] = \mu_{p \times 1}$

pf : $\underset{\sim}{\bar{X}} = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_p \end{bmatrix}$ $\qquad \bar{X}_i = \frac{1}{n} \sum_{i=1}^{n} X_{ki}$ sample is iid

$$E[\bar{X}_i] = \frac{1}{n} \sum_{i=1}^{n} E[X_{ki}]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mu_i \quad (\because \text{ sample is i.d.})$$

$$= \mu_i$$

$$E[\underset{\sim}{\bar{X}}] = \begin{bmatrix} E[\bar{X}_1] \\ \vdots \\ E[\bar{X}_p] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} = \mu_{p \times 1}$$

For any p, $\bar{X}$ is an unbiased estimator of $\mu \in \mathbb{R}^p$.

(b) $\text{cov}(\underset{\sim}{\bar{X}}) = \frac{\Sigma}{n}$

$$\underset{p\times 1}{\mathrm{cov}(\bar{X})} = E\Big[\underbrace{\underset{p\times 1}{(\bar{X}-E(\bar{X}))}\underset{1\times p}{(\bar{X}-E(\bar{X}))^T}}_{p\times p}\Big]$$

$$= E[(\bar{X}-\mu)(\bar{X}-\mu)^T] \quad \text{(using part (a))}$$

$$= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right)\left(\frac{1}{n}\sum_{j=1}^{n}X_i-\mu\right)^T\right]$$

$$= \frac{1}{n^2}E\left[\left(\sum_{i=1}^{n}(X_i-\mu)\right)\left(\sum_{i=1}^{n}(X_i-\mu)\right)^T\right]$$

$$= \cancel{\frac{1}{n^2}E\Big[}\ \frac{1}{n^2}E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\mu)(X_j-\mu)^T\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n}(X_i-\mu)(X_i-\mu)^T + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i-\mu)(X_j-\mu)^T\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n}E[(X_i-\mu)(X_i-\mu)^T]\right] + \underbrace{\frac{1}{n^2}\left[\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}E[(X_i-\mu)(X_j-\mu)^T]\right]}_{=0}$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n}E[(X_i-\mu)(X_i-\mu)^T]\right]$$

$\because$ X_i & X_j are two independent r.v. in $\mathbb{R}^p$, $\therefore$ $\mathrm{cov}(X_i, X_j) = 0$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\Sigma = \frac{\Sigma}{n} \quad \Sigma \rightarrow p\times p \text{ population var-cov. mtx.}$$

12/08/2025

c) $E[S] = \Sigma$, $S = [s_{ij}]_{n\times n}$

$$s_{ij} = \frac{1}{n-1}\sum_{r=1}^{n}(x_{ri}-\bar{x}_i)(x_{rj}-\bar{x}_j)$$

Proof: consider

$$(n-1)S = \sum_{k=1}^{n} (\underset{\sim}{X_k} - \underset{\sim}{\bar{X}})(\underset{\sim}{X_k} - \underset{\sim}{\bar{X}})^T$$

$$= \sum_{k=1}^{n} (\underbrace{\underset{\sim}{X_k} - \mu}_{} + \underbrace{\mu - \bar{X}}_{})(\underbrace{X_k - \mu}_{} + \underbrace{\mu - \bar{X}}_{})^T$$

$$= \sum_{k=1}^{n} (\underset{\sim}{X_k} - \mu)(\underset{\sim}{X_k} - \mu)^T + \sum_{k=1}^{n} (\underset{\sim}{X_k} - \mu)(\mu - \underset{\sim}{\bar{X}})^T$$

$$+ \sum_{k=1}^{n} (\mu - \underset{\sim}{\bar{X}})(\underset{\sim}{X_k} - \mu)^T + \sum_{k=1}^{n} (\mu - \bar{X})(\mu - \bar{X})^T$$

4th term RHS: $n(\bar{X} - \mu)(\bar{X} - \mu)^T$

3rd ~~the~~ term RHS: $- \sum_{k=1}^{n} (\underset{\sim}{\bar{X}} - \mu)(\underset{\sim}{X_k} - \mu)^T$

$$= -(\underset{\sim}{\bar{X}} - \mu)\, n\, (\underset{\sim}{\bar{X}} - \mu)^T$$

$$= -n(\underset{\sim}{\bar{X}} - \mu)(\underset{\sim}{\bar{X}} - \mu)^T$$

~~3rd~~ 2nd term RHS: $-n(\underset{\sim}{\bar{X}} - \mu)(\underset{\sim}{\bar{X}} - \mu)^T$

2 + 3 + 4 $\equiv -n(\underset{\sim}{\bar{X}} - \mu)(\underset{\sim}{\bar{X}} - \mu)^T$

$$E[(n-1)S] = \sum_{k=1}^{n} \underbrace{E[(X_k - \mu)(X_k - \mu)^T]}_{\text{cov}(X_k) = \Sigma} - n \underbrace{E[(\bar{X} - \mu)(\bar{X} - \mu)^T]}_{\text{cov}(\bar{X})}$$

$$= \sum_{k=1}^{n} (\Sigma) - n\left(\frac{\Sigma}{n}\right) = (n-1)\Sigma$$

$$\Rightarrow E[S] = \Sigma \quad \left(\because \text{\# samples} = n > 2\right)$$

If we use n ~~it~~ (instead of $n-1$) in s_{ij}

formula and call the matrix by say

$$S_n = [s_{ij}]$$

$$s_{ij} = \frac{1}{n} \sum_{r=1}^{n} (x_{ri} - \bar{x}_i)(x_{rj} - \bar{x}_j)$$

$$E[S_n] = \frac{n-1}{n} \Sigma \qquad \lim_{n \to \infty} \left(\frac{n-1}{n}\right) = 1$$

bias created

$$\lim_{n \to \infty} E[S_n] = \Sigma$$

we have $\frac{p(p+1)}{2}$ many s_{ij}

As $p \uparrow$, finding variability in dataset can become computationally expensive

Instead of this, can we have some way of expressing variability by one real number.

$$\text{Trace}(S) = \sum_{i=1}^{p} s_{ii} \rightarrow \text{Total variation in data OR total variability in data}$$

$\downarrow$

A limitation is that it does not care of dependency in features

$$|S| = \begin{vmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{vmatrix} = s_{11} \cdot s_{22} - \underbrace{(s_{12})^2}_{\text{does take care of inter-dependency}}$$

$(= s_{12})$

$|S| \leftrightarrow$ sample generalized variance

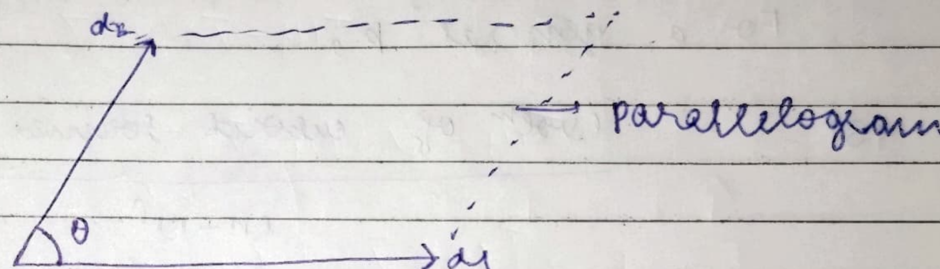
or generalized Variability in data

$p = 2$ (2 features)

$$\begin{array}{c} \quad X_1 \quad\; X_2 \\ \begin{bmatrix} \vdots & \vdots \\ \vdots & \vdots \end{bmatrix}_{n \times 2} \\ \overline{X}_1 \quad \overline{X}_2 \end{array} \qquad \begin{bmatrix} \vec{d_1} & \vec{d_2} \\ {\scriptstyle n\times 1} & {\scriptstyle n \times 1} \end{bmatrix}$$

$$\left.\begin{aligned} \vec{d_1} &= X_1 - \overline{X}_1 \\ \vec{d_2} &= X_2 - \overline{X}_2 \end{aligned}\right\} \text{deviation vectors}$$

In R^n plane



Area of $\|$gm

$= |\vec{d_1} \times \vec{d_2}|$

$= \left| \, \|\vec{d_1}\| \, \|\vec{d_2}\| \sin\theta \, \hat{n} \right|$

Note: $\|\vec{d_j}\|^2 = (n-1) s_{jj}$

$\left| \sqrt{(n-1)s_{11}} \sqrt{(n-1)s_{22}} \sqrt{1 - \cos^2\theta} \right|$

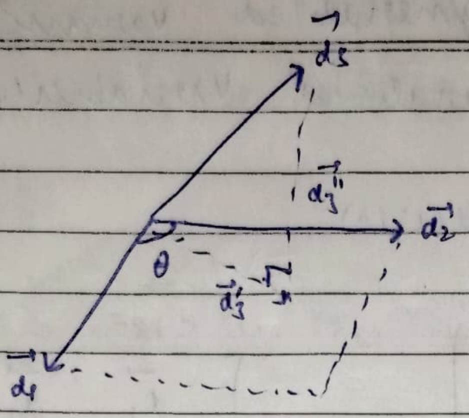
$= (n-1)\sqrt{s_{11}}\sqrt{s_{22}}\sqrt{1-\rho_{12}^2}$ $\qquad \rho_{12}$: correlation coeff. b/w $\vec{d_1}$ & $\vec{d_2}$

$= (n-1)\sqrt{s_{11}}\sqrt{s_{22}}\sqrt{1 - \left(\dfrac{s_{12}}{\sqrt{s_{11}}\sqrt{s_{22}}}\right)^2}$

$= (n-1)\sqrt{s_{11}s_{22} - s_{12}^2} = (n-1)|S|^{1/2}$

$\Rightarrow \det(S) = \dfrac{(\text{Area})^2}{(n-1)^2} \Rightarrow \det(S) \propto (\text{Area})^2$

$p=3$



A vol of a parallelopiped formed by d_1, d_2, d_3

$= \vec{d_3} \cdot (\vec{d_1} \times \vec{d_2})$

$= \vec{d_3}'' \cdot (\vec{d_1} \times \vec{d_2}) = |\vec{d_3}''||\vec{d_1}||\vec{d_2}||\sin\theta|$

For a general p,

$$|S| = \frac{(\text{vol}^n \text{ of cuboid formed by } \vec{d_1}, \dots, \vec{d_p})^2}{(n-1)^p}$$

Taking interpretation of $|S|$ in $\mathbb{R}^p$

vol^n of p-dimensional ellipsoid

$$\frac{2(\pi)^{p/2}}{p\,\Gamma\frac{p}{2}} \sqrt{|S|} \times C$$

C → appears because of radius factor

↓ constant

$\sqrt{|S|}$ → appears due to difference in structure of sphere and ellipse

$\frac{2(\pi)^{p/2}}{p\,\Gamma\frac{p}{2}}$ ↓ vol^n of a unit sphere in p-dim^n

Notations: Δ → deviation matrix

$\Delta = [\vec{d_1}\ \vec{d_2}\ \cdots\ \vec{d_p}]$

$$\frac{2(\pi)^{3/2}}{3 \times \frac{1}{2}\sqrt{\pi}} = \frac{4}{3}\pi$$

Observe that:

$$(n-1)\, S_{ij} = d_i^T d_j$$

$$\underset{p \times p}{(n-1)\, S} = \underbrace{\underset{p \times n}{\Delta^T}\ \underset{n \times p}{\Delta}}_{p \times p} \Rightarrow S = \frac{\Delta^T \Delta}{n-1}$$

Note: Sum of all rows of Δ matrix $= 0$

Note: $\Delta = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} - \begin{bmatrix} \bar{x}_1 & \cdots & \bar{x}_p \\ \vdots & \ddots & \vdots \\ \bar{x}_1 & \cdots & \bar{x}_p \end{bmatrix}$

$$= \underset{n \times p}{X} - \underset{n \times 1}{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}} \bar{x}^T_{1 \times p}$$

X ↓ sample data matrix

$$\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} \qquad \therefore \Delta = X - \vec{1} \cdot \bar{x}^T$$

$$\therefore S = \frac{1}{n-1} (X - \vec{1} \cdot \bar{x}^T)^T (X - \vec{1} \cdot \bar{x}^T)$$

Result: For some vector $\vec{a} \in \mathbb{R}^p$ $(\vec{a} \neq \vec{0})$

$$S\vec{a} = 0 \iff \Delta \vec{a} = 0$$

i.e. $|S| = 0 \iff$ cols. of Δ are L.D.

$\iff \vec{d_1}, \dots, \vec{d_p}$ are L.D. vectors in $\mathbb{R}^p$

Pf: $(\Leftarrow)$ $S\vec{a}$ $\quad \Delta\vec{a} = \vec{0} \Rightarrow \Delta^T\Delta\vec{a} = \vec{0}$

$\Rightarrow S\vec{a} = \vec{0}$

$(\Rightarrow)$ $S\vec{a} = \vec{0}$

$\Rightarrow \vec{a}^T S\vec{a} = 0 \Rightarrow \vec{a}^T\Delta^T\Delta\vec{a} = 0$

$\Rightarrow (\Delta\vec{a})^T(\Delta\vec{a}) = 0$

$\Rightarrow \|\Delta\vec{a}\|^2 = 0 \Rightarrow \Delta\vec{a} = \vec{0}$

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \underbrace{\frac{1}{n}\underset{p\times n}{X^T}\underset{n\times 1}{\vec{1}}}_{p\times 1}$$

$$\Delta = X - \vec{1}\left(\frac{1}{n}(X^T\vec{1})^T\right)$$

$$\underset{n\times p}{\Delta} = X - \frac{1}{n}\underset{n\times 1}{\vec{1}}\,\underset{1\times n}{\vec{1}^T}\,\underset{n\times p}{X} = \left(I - \frac{\vec{1}\vec{1}^T}{n}\right)X$$

$$S = \frac{1}{n-1}\Delta^T\Delta = \frac{1}{n-1}X^T\left(I - \frac{\vec{1}\vec{1}^T}{n}\right)^T\left(I - \frac{\vec{1}\vec{1}^T}{n}\right)X$$

$$\Rightarrow (n-1)S = X^T H X$$

where $H = I - \dfrac{\vec{1}\vec{1}^T}{n}$, $\vec{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1}$

verify: $H^T = H$; $H^2 = H \Rightarrow H^TH = H^2 = H$

symmetric idempotent matrix

called centering matrix

Multivariate Normal

$\underset{\sim}{X} \sim N_p(\mu, \Sigma) \Rightarrow a^T \underset{\sim}{X} \sim N(\vec{a}^T\mu, \vec{a}^T \Sigma \vec{a})$

$\downarrow$

$p \times 1$

$\underset{\sim}{Y} = A\underset{\sim}{X} + b$

$\Rightarrow \underset{\sim}{Y} \sim N_q(A\mu + b, A\Sigma A^T)$

By prop. of pdf: $\int_{\mathbb{R}^p} f_{\underset{\sim}{X}}(x)\, dx = 1$

using the spectral decomposition

$\Sigma = P \Lambda P^T, \quad PP^T = I = P^T P$

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p), \quad \lambda_i > 0 \ \forall i$

$\Sigma^{-1} = (P^T)^{-1} \Lambda^{-1} P^{-1}$

$\Rightarrow \Sigma^{-1} = P \Lambda^{-1} P^T$

$(x-\mu)^T \Sigma^{-1} (x-\mu) = \underline{(x-\mu)^T P}\, \Lambda^{-1} P^T (x-\mu)$

$= y^T \Lambda^{-1} y \quad ; \quad y = P^T(x-\mu)$

$y^T \begin{pmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_p} \end{pmatrix} y = \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i}$

$Py = x - \mu \Rightarrow \boxed{x = Py + \mu}$

$$dx = \underbrace{\text{abs}(|J|)}_{\text{Jacobian}} dy$$

$$= (\text{abs}\ |P|)\ dy = dy \quad (|P|^2 = 1)$$

$$\Rightarrow \int_{y \in \mathbb{R}^p} f(y)\, dy = 1$$

$$\int_{y \in \mathbb{R}^p} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} \right) dy = 1$$

$$|\Sigma| = |\Lambda| = \lambda_1 \cdots, \lambda_p$$

$$\int_{y \in \mathbb{R}^p} \frac{1}{(2\pi)^{p/2} (\lambda_1 \cdots \lambda_p)^{1/2}} \prod_{i=1}^{p} \exp\left(-\frac{1}{2} \frac{y_i^2}{\lambda_i} \right) dy = 1$$

$$\Rightarrow \prod_{i=1}^{p} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \lambda_i}} \exp\left(- \frac{y_i^2}{2\lambda_i} \right) dy_i \right) = 1$$

$$\Rightarrow y_i \sim N(0, \lambda_i) \quad \forall\ 1 \le i \le p$$

Note: ① $\underset{\sim}{X} \sim N_p(\mu, \Sigma)$

Then $Y_{p \times 1} = P^T(\underset{\sim}{X} - \mu)$

with $y_i \sim N(0, \lambda_i)$

and pdf of $\underset{\sim}{Y} = (y_1, \dots, y_p)^T$

= pdt. of pdf of y_i ; $1 \le i \le p$

$\underset{\sim}{Y}_1, \dots, \underset{\sim}{Y}_p$ are independent

r.vs and $y_i \sim N(0, \lambda_i)$

② $Y_i \sim N(0, \lambda_i) \Rightarrow \dfrac{y_i}{\sqrt{\lambda_i}} \sim N(0,1)$

$\Rightarrow \dfrac{y_i^2}{\lambda_i} \sim \chi^2_{(1)}$, $1 \leq i \leq p$ and y_i are iid

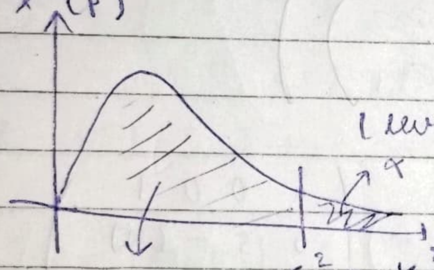
$\Rightarrow \sum\limits_{i=1}^{p} \dfrac{y_i^2}{\lambda_i} \sim \chi^2_{(p)}$

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

$P(\underbrace{d(X, \mu)}_{\text{statistical distance}} \leq c^2) \Rightarrow P(\underbrace{(X-\mu)^T \Sigma^{-1} (X-\mu)}_{\chi^2(p)} \leq c^2)$

$\Rightarrow P(\underbrace{(X-\mu)^T \Sigma^{-1} (X-\mu) \leq \chi^2_{(p),\alpha}}_{\text{confidence ellipse}}) \leq 1-\alpha$

$\chi^2(p)$



(level of significance) α

$1-\alpha$ → level of confidence

$c^2 = \chi^2_{(p),\alpha}$

confidence ellipse α ~~ee~~

③ $\because \underset{\sim}{X} \sim N_p(\mu, \Sigma)$

$A_{q\times p} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 0 \end{bmatrix}$

$\underset{\sim}{Y} = \underset{q\times p}{A} \, \underset{p\times 1}{\underset{\sim}{X}} + \underset{q\times 1}{b}$

$b = 0 \in \mathbb{R}^q$

$Y \sim N_q(A\mu + b, A\Sigma A^T)$

$AX = \begin{pmatrix} X_1 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}$

$X = (X_1, \dots, X_p)^T_{p\times 1}$

$\Rightarrow X_1 \sim N(\mu_1, \sigma_{11} = \sigma_1^2)$

$\vdots$

$X_p \sim N(\mu_p, \sigma_{pp} = \sigma_p^2)$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim N_4(\mu_{4\times1}, \Sigma_{4\times1})$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}_{2\times4} \qquad b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$Y = AX = \begin{bmatrix} X_2 \\ X_3 \end{bmatrix} \sim N_2(A\mu, A\Sigma A^T)$$

$$A\mu = \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \qquad A\Sigma A^T = \begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} X_2 \\ X_3 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{pmatrix} \right)$$

__Example__ $X \sim N_4(\mu, \Sigma)$

$$\Sigma = \begin{pmatrix} 4 & 0 & 0 & 1 \\ 0 & 5 & -3 & 0 \\ 0 & -3 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\mu = (-1, 1, 2, -3)^T$$

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix} \right)$$

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

$$\underset{\sim}{X} \sim N_p\left(\begin{pmatrix} \mu_1 \rightarrow k\times1 \\ \hline \mu_2 \end{pmatrix}_{p\times1}, \begin{pmatrix} \overset{k\times k}{\Sigma_{11}} & \Sigma_{12} \\ \hline \Sigma_{21} & \underset{(p-k)\times(p-k)}{\Sigma_{22}} \end{pmatrix} \right)$$

$$\underset{\sim}{X}_{p\times1} = \begin{pmatrix} \underset{\sim}{X_1} \rightarrow k\times1 \\ \hline \underset{\sim}{X_2} \rightarrow (p-k)\times1 \end{pmatrix}$$

$$\Sigma_{21} = \Sigma_{12}^T$$

Result: A random vector $\underset{\sim}{X}_1$ $(k \times 1)$ is independent of the random vector $X_2 : (p-k) \times 1$ iff

$$\Sigma_{12} = O_{k \times (p-k)}$$

Holds both ways in case of normal distribution

Pf: let $\tilde{X}_1$ and $\tilde{X}_2$ be two independent r.v.s.

$\Rightarrow$ Any i^{th} r.v. from X_1 and any j^{th} r.v. from X_2 must be independent.

$\Rightarrow \operatorname{cov}(X_1^i, X_2^j) = 0$

(Note: we have not used normal dist. in this part)

Converse: let $\Sigma_{12} = 0 \Rightarrow \Sigma_{21} = \Sigma_{12}^T = 0$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \Rightarrow \Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}$$

and $|\Sigma| = |\Sigma_{11}||\Sigma_{22}|$ ($\because$ this is a block diagonal mtx.)

$$f_{\underset{\sim}{X}}(x_1, x_2) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \cdot \exp\left(-\frac{1}{2}(\underset{\wedge}{X}-\mu)^T \Sigma^{-1} (\underset{\wedge}{X}-\mu)\right)$$

$$= \frac{1}{(2\pi)^{\frac{k+p-k}{2}}(|\Sigma_{11}||\Sigma_{22}|)^{1/2}} \exp\left(-\frac{1}{2}(X_1-\mu_1 \vdots X_2-\mu_2)^T \begin{pmatrix} \Sigma_{11}^{-1} & \vdots \\ \cdots & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} X_1-\mu_1 \\ \cdots \\ X_2-\mu_2 \end{pmatrix}\right)$$

$$= (\,/\,/\,) \exp\left(-\frac{1}{2}(X_1-\mu_1)^T \Sigma_{11}^{-1}(X_1-\mu_1) - \frac{1}{2}(X_2-\mu_2)^T \Sigma_{22}^{-1}(X_2-\mu_2)\right)$$

$$= \left(\frac{1}{(2\pi)^{k/2}|\Sigma_{11}|^{1/2}} \exp\left(\frac{-1}{2}(X_1-\mu_1)^T \Sigma_{11}^{-1}(X_1-\mu_1)\right)\right) \times$$

$$\left(\frac{1}{(2\pi)^{p-k/2}|\Sigma_{22}|^{1/2}} \exp\left(\frac{-1}{2}(X_2-\mu_2)^T \Sigma_{22}^{-1}(X_2-\mu_2)\right)\right)$$

$$= f_{X_1}(x_1) \times f_{X_2}(x_2)$$

($\because X_1$ and X_2 are both normal)

$\Rightarrow X_1$ and X_2 are ind. r. vectors.

---

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

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \begin{matrix} \to k \times 1 \\ \to (p-k) \times 1 \end{matrix}$$

conditional distribution of $\underset{\sim}{X}_1$ given that $\underset{\sim}{X}_2 = x_2$ is called regression of $\underset{\sim}{X}_2$ on $\underset{\sim}{X}_1$

$$X_1 \mid X_2 = x_2 \sim N_k\left(\mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2) \right.$$
$$\left. , \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \right)$$

Pf: Define

$$C_{p \times p} = \left(\begin{array}{c|c} I_k & \cancel{O_{k \times p-k}} - \Sigma_{12} \Sigma_{22}^{-1} \\ \hline O_{p-k \times k} & I_{p-k} \end{array} \right)$$

let $Y = CX \Rightarrow \underset{\sim}{Y} \sim N_p(C\mu, C\Sigma C^T)$

$$C\mu = \left(\begin{array}{c} \mu_1 - \Sigma_{12} \Sigma_{22}^{-1} \mu_2 \\ \hline \mu_2 \end{array} \right)_{p \times 1}$$

$$C \Sigma C^T = \left(\begin{array}{c|c} I & -\Sigma_{12} \Sigma_{22}^{-1} \\ 0 & I_{p-k} \end{array} \right) \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right) \left(\begin{array}{c|c} I & 0 \\ \hline -\Sigma_{22}^{-1} \Sigma_{21} & I_{p-k} \end{array} \right)$$

$$= \left(\begin{array}{c|c} I & -\Sigma_{12} \Sigma_{22}^{-1} \\ \hline 0 & I_{p-k} \end{array} \right) \left(\begin{array}{c|c} \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} & \Sigma_{12} \\ \hline \Sigma_{21} - \Sigma_{21} = 0 & \Sigma_{22} \end{array} \right)$$

$$= \left(\begin{array}{c|c} \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} & 0 \\ \hline 0 & \Sigma_{22} \end{array} \right)$$

$\Rightarrow X_1$ and X_2 are ind. r. vectors.

---

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

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \begin{matrix} \to k \times 1 \\ \to (p-k) \times 1 \end{matrix}$$

conditional distribution of $\underset{\sim}{X_1}$ given that $\underset{\sim}{X_2} = x_2$ is called regression of $\underset{\sim}{X_2}$ on $\underset{\sim}{X_1}$

$$X_1 \mid X_2 = x_2 \sim N_k\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

Pf: Define

$$C_{p \times p} = \left(\begin{array}{c|c} I_k & \cancel{O_{k \times p-k}}\ -\Sigma_{12}\Sigma_{22}^{-1} \\ \hline O_{p-k \times k} & I_{p-k} \end{array}\right)$$

let $Y = CX \Rightarrow \underset{\sim}{Y} \sim N_p(C\mu, C\Sigma C^T)$

$$C\mu = \left(\begin{array}{c} \mu_1 + - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \hline \mu_2 \end{array}\right)_{p \times 1}$$

$$C\Sigma C^T = \left(\begin{array}{c|c} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ \hline 0 & I_{p-k} \end{array}\right)\left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right)\left(\begin{array}{c|c} I & 0 \\ \hline -\Sigma_{22}^{-1}\Sigma_{21} & I_{p-k} \end{array}\right)$$

$$= \left(\begin{array}{c|c} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ \hline 0 & I_{p-k} \end{array}\right)\left(\begin{array}{c|c} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{12} \\ \hline \Sigma_{21} - \Sigma_{21} = 0 & \Sigma_{22} \end{array}\right)$$

$$= \left(\begin{array}{c|c} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ \hline 0 & \Sigma_{22} \end{array}\right)$$

Also: $\underset{\sim}{Y} = \begin{pmatrix} X_1 - \Sigma_{12}\Sigma_{22}^{-1} X_2 \\ \hdashline X_2 \end{pmatrix} \begin{matrix} \to k \times 1 \\ \to (p-k)\times 1 \end{matrix}$, $\begin{pmatrix} Y_1 \\ \hline X_2 \end{pmatrix}$

Y_1 and X_2 are independent (prev thm)

$\Rightarrow$ conditional distribution of $Y_1 = X_1 - \Sigma_{12}\Sigma_{22}^{-1} X_2$ given $X_2 = x_2$ is same as the distribution of $Y_1 = X_1 - \Sigma_{12}\Sigma_{22}^{-1} x_2$

$$\underset{\sim}{X_1} - \Sigma_{12}\Sigma_{22}^{-1} x_2 \sim \mathbb{N}\left(\left(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$$

$$\Rightarrow \underset{\sim}{X_1} \sim \mathbb{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$$

e.g. — let $\vec{a} = (a_1, a_2, a_3)^T \in \mathbb{R}^3$

$$\underset{\sim}{X} \sim \mathbb{N}_4\left(\begin{pmatrix} -1 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 & 0 & 0 & 1 \\ 0 & 5 & -3 & 0 \\ 0 & -3 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \right)$$

Find $\vec{a}$ s.t. $\begin{pmatrix} X_1 - X_2 \\ X_2 - X_3 \end{pmatrix}$ and $(a_1X_1 + a_2X_2 + a_3X_3)$ are independent random vectors.

$$\begin{pmatrix} X_1 - X_2 \\ X_2 - X_3 \\ \hline a_1X_1 + a_2X_2 + a_3X_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ a_1 & a_2 & a_3 \end{pmatrix}}_{C} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim \mathbb{N}_3\left(C\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, C\Sigma_{123}C^T \right)$$

$$\Sigma_{123} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 5 & -3 \\ 0 & -3 & 2 \end{pmatrix} \qquad C\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \\ a_1\mu_1 + a_2\mu_2 + a_3\mu_3 \end{pmatrix}$$

$$C\Sigma_{123}C^T = \begin{pmatrix} 9 & -8 & \vdots & 4a_1 - 5a_2 + 3a_3 \\ -8 & 13 & \vdots & 8a_2 - 5a_3 \\ \hline \text{trans.} & & & (\dots) \end{pmatrix}_{3\times 3}$$

For independence,

$$4a_1 - 5a_2 + 3a_2 = 0$$

$$8a_2 - 5a_3 = 0$$

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 20 \\ 32 \end{pmatrix}$$

Q: Find the conditional distribution of $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ given the value of $\begin{pmatrix} X_3 \\ X_4 \end{pmatrix} = \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}$.

$$\Sigma = \left(\begin{array}{cc|cc} \overbrace{4 \quad 0}^{\Sigma_{11}} & & \overbrace{0 \quad 1}^{\Sigma_{12}} & \\ 0 & 5 & -3 & 0 \\ \hline 0 & -3 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right)$$

$$\Sigma_{21} = \Sigma_{12}^T \qquad \Sigma_{22}$$

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \Bigg| \begin{pmatrix} X_3 = x_4 \\ x_4 = x_4 \end{pmatrix} \sim N_2\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ -3 & 0 \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right.$$

$$\left. + \begin{pmatrix} 0 & 1 \\ -3 & 0 \end{pmatrix} \begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ -3 & 0 \end{pmatrix} \begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -3 \\ 1 & 0 \end{pmatrix} \right)$$

Note: suppose we have l ind. r.v. $\underset{p \times 1}{\underset{\sim}{X_1}}, \ldots, \underset{p \times 1}{\underset{\sim}{X_l}}$

with each $X_i \sim N_p(\mu_i, \Sigma)$

let $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X_1} \\ \vdots \\ \underset{\sim}{X_l} \end{pmatrix}_{lp \times 1}$ $\quad E[X] = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_l \end{pmatrix}_{lp \times 1}$

$$\mathrm{cov}(\underset{\sim}{X}) = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma \end{pmatrix}_{lp \times lp}$$

let $\underset{\sim}{V_1} = c_1 \underset{\sim}{X_1} + \dots + c_l \underset{\sim}{X_l}$

$\underset{\sim}{V_2} = b_1 \underset{\sim}{X_1} + \dots + b_l X_l$

$$V = \begin{pmatrix} \underset{\sim}{V_1} \\ \underset{\sim}{V_2} \end{pmatrix}_{2p \times 1} \qquad V \sim N_{2p}(\mu, \Sigma)$$

$$\mu = \begin{pmatrix} \sum_{i=1}^{l} c_i \mu_i \\ \hline \sum_{i=1}^{l} b_i \mu_i \end{pmatrix} \qquad \Sigma_{2p \times 2p} = \left(\begin{array}{c|c} \mathrm{cov}(V_1, V_1) & \mathrm{cov}(V_1, V_2) \\ \hline \text{transpose} & \mathrm{cov}(V_2, V_2) \end{array}\right)$$

$$\mathrm{cov}(V_1, V_1) = E\left[(V_1 - E(V_1))(V_1 - E(V_1))^T\right]$$

$$= E\left[\left(\sum_{i=1}^{l} c_i (X_i - \mu_i)\right)\left(\sum_{j=1}^{l} c_j (X_j - \mu_j)^T\right)\right]$$

$$= E\left[\sum_{i=1}^{l} \sum_{j=1}^{l} c_i c_j (X_i - \mu_i)(X_j - \mu_j)^T\right]$$

$$= \sum_{i=1}^{l} c_i^2 \, E\left[(X_i - \mu_i)(X_i - \mu_i)^T\right] \quad (i \neq j \text{ terms are } 0)$$

$$= \left(\sum_{i=1}^{l} c_i^2\right) \Sigma$$

||^{ly}; $\mathrm{cov}(V_2, V_2) = \left(\sum_{i=1}^{l} b_i^2\right) \Sigma$

$$\mathrm{cov}(V_1, V_2) = \mathrm{cov}\left(\sum_{i=1}^{l} c_i X_i, \sum_{i=1}^{l} b_i X_i\right)$$

$$= \left(\sum_{i=1}^{l} c_i b_i\right) \Sigma$$