Chapter 4 Special Topics

Key-words

- g-inverse, Moore-Penrose inverse
- QR factorization, Cholesky decomposition
- Kronecker product, Vec operator

4.1 Multivariate normal distribution

$\underline{x} = (x_1, \ldots, x_p)^t \sim N_p(\underline{\mu}, \Sigma)$, $\Sigma > 0$

$$f(\underline{x}; \underline{\mu}, \Sigma) = |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^t \Sigma^{-1}(\underline{x}-\underline{\mu}\right)$$

where $E(\underline{x}) = \underline{\mu}$, $VAR(\underline{x}) \equiv E\left\{(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})^t\right\} = \Sigma$

1) $p=2$, Bivariate normal

$$\underline{x} = (x_1, x_2)^t \sim N_2(\underline{\mu}, \Sigma) \quad \text{where } \underline{\mu} = (\mu_1, \mu_2)^t \;,\; \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

2) $\underline{x} = \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} \sim N_p(\underline{\mu}, \Sigma)$ where $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

$$\Sigma_{12} = 0 \quad \Rightarrow \quad \underline{x}_1 = (x_{1,} \cdots, x_q)^t \sim N_q(\underline{\mu}_1, \Sigma_{11}) \perp \underline{x}_2 = (x_{q+1,} \cdots, x_p)^t \sim N_{r(}\underline{\mu}_1, \Sigma_{22}) \text{ with } r = p-q$$

Ref: Sec 2.6 and Sec 2.7

4.2 Generalized inverse

$\underline{y} = X\underline{b} + \underline{e}$

$\Longrightarrow \quad (X^t X)\underline{b} = X^t \underline{y}$

$X^t X_{p \times p}$: singular $\Longrightarrow \quad \nexists (X^t X)^{-1}$

$\Longleftrightarrow \quad rank(X^t X) \neq p$: not full-rank

■ Linear equations and the solution

$A\underline{b} = \underline{c} \ \wedge \ \exists \underline{b}$

$\Longleftrightarrow \quad rank(A : \underline{c}) = rank(A)$: consistent

Note

A : singular $\wedge$ $AGA = A$: condition for g-inverse

$\Rightarrow$ $A\underline{b} = \underline{c} \Rightarrow AGA\underline{b} = AG\underline{c}$

$\Rightarrow A\underline{b} = AG\underline{c}$

$\Leftrightarrow \underline{b} = G\underline{c}$

G : g-inverse of A

■ Generalized inverse matrix

$A_{p\times q}$ with $rank(A) = r(\leq p$ or $q)$ and $G_{q\times p}$: g-inverse of A

$\Rightarrow$ $\underset{q\times p}{G} = \underset{q\times q}{V} \underset{q\times q}{\begin{pmatrix} D_{11}^{-1} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}} \underset{q\times p}{U^t}$ where $A = UDV^t$, $D = \begin{pmatrix} D_{11} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix}$; $D_{11_{r\times r}}$: diagonal matrix

$U^tU = I$, $V^tV = I$, H_{12}, H_{21}, H_{22} : any matrices

Proof)

- $\underset{(p\times q)}{A} = \underset{(p\times q)}{U}\ \underset{(q\times q)}{D}\ \underset{(q\times q)}{V^t}$: SVD ($U^tU = I_q$, $V^tV = I_q$)

- $D = \begin{pmatrix} D_{11} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix}$ where $D_{11_{r\times r}}$: diagonal matrix

$$\Rightarrow \quad AGA = U\begin{pmatrix} D_{11} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} V^t \boxed{G}\, U\begin{pmatrix} D_{11} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} V^t$$

$$= U\begin{pmatrix} D_{11} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} V^t \boxed{V\begin{pmatrix} D_{11}^{-1} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} U^t}\, U\begin{pmatrix} D_{11} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} V^t$$

$$= UDV^t$$

$$= A$$

$$\therefore AGA = A$$

Moore-Penrose inverse

From $G = V\begin{pmatrix} D_{11}^{-1} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} U^t$, if $H_{12} = 0, H_{21} = 0, H_{22} = 0$,

- $A^- = V\begin{pmatrix} D_{11}^{-1} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} U^t$: Moore-Penrose inverse

Note Properties of A^-

i) $AA^-A = A,\ A^-AA^- = A^-$

ii) AA^- : symm. and A^-A : symm.

iii) $rank(A) = r\ \ (\le p, q)$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \ ;\ rank(A_{11}) = r,\ A_{11} : r \times r$$

$$\Rightarrow G_{q\times p} = \begin{pmatrix} A_{11}^{-1} & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix}$$

- ginv(A) in R and SAS/IML

4.3 QR factorization and Cholesky decomposition

Useful ways of expressing a given matrix in the form of a product of the matrices having some special structure.

■ QR factorization

• $X_{n\times p}$ with $n \geq p$ • $R_{p\times p}$: <u>upper triangular matrix</u> $\Longrightarrow \underset{n\times p}{X} = \underset{n\times p}{Q}\ \underset{p\times p}{R}$ • $Q_{n\times p}$: $Q^t Q = I_p$	decomp<-qr(X) qr.Q(decomp) qr.R(decomp) qr.Q(decomp)%*%qr.R(decomp

<u>Corollary</u>

- $Q_{n\times n} \Longrightarrow Q^t Q = I,\ \ QQ^t = I$

- $X_{n\times p} = Q_{n\times n}\ R_{n\times p}$ where $R_{n\times p} = \begin{pmatrix} R_1 & p\times p \\ \underline{0} & (n-1)\times p \end{pmatrix}$

Example

$$X = \begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix} = \begin{bmatrix} 0.1961161 & -0.1694754 & 0.96582428 \\ -0.5883484 & -0.7676240 & -0.25416428 \\ 0.7844645 & -0.6180869 & 0.05083286 \end{bmatrix} \begin{bmatrix} -20.39608 & -57.854260 & 105.3143646 \\ 0.00000 & -3.858058 & -24.8530745 \\ 0.00000 & 0.000000 & 0.4574957 \end{bmatrix}$$

$$= QR$$

Application

- $\underset{n\times 1}{\underline{y}} = \underset{n\times p}{X} \underset{p\times 1}{\underline{b}} + \underset{n\times 1}{\underline{e}}$

$S = \| \underline{e} \|^2 = \| \underline{y} - X\underline{b} \|^2$

$\min_{\underline{b}} S \;\Rightarrow\; (X^t X)\underline{b} = X^t \underline{y}$: $\exists$ numerical errors !

In general, $rank(X) = full-rank$ and $X^t X$: non-singular $\Rightarrow$ $\hat{\underline{b}} = (X^t X)^{-1} X^t \underline{y}$

- $\| \underline{y} - X\underline{b} \|^2 = \| Q^t \underline{y} - Q^t X \underline{b} \|^2 \quad (\because QQ^t = I_n)$

$$= \| Q^t \underline{y} - R\underline{b} \|^2 \quad (\because Q^t X = R)$$

$$\underline{y}^* \stackrel{let}{=} Q^t \underline{y} = \begin{pmatrix} \underline{y}_1^* & p\times 1 \\ \underline{y}_2^* & (n-p)\times 1 \end{pmatrix}, \quad R_{n\times p} = \begin{pmatrix} R_1 & p\times p \\ \underline{0} & (n-p)\times p \end{pmatrix}$$

$$\| \underline{y} - X\underline{b} \|^2 = \| \underline{y}_1^* - R_1 \underline{b} \|^2 + \| \underline{y}_2^* \|^2 \geq \| \underline{y}_2^* \|^2$$

$\Longrightarrow$ Equality holds for $R_1 \underline{b} = \underline{y}_1^*$

■ Cholesky decomposition

• $A_{m\times m} \geq 0$ $\Longrightarrow \quad \exists\ T_{m\times m}$: lower triangular matrix with $t_{ii} \geq 0,\ i = 1, 2, \cdots, m$ s.t $A = TT^t$ • $A > 0 \quad \Longrightarrow \quad \exists!\ T$ with $t_{ii} > 0,\ i = 1, 2, \cdots, m$	Tt<-chol(A) Tt T<-t(Tt) T T%*%Tt

Note

- $A^{-1} = (TT^t)^{-1} = (T^t)^{-1} T^{-1}$

- T=ROOT(A) in SAS/IML

Example

$$X = \begin{pmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 6 & 1 & 0 \\ -8 & 5 & 3 \end{pmatrix} \begin{pmatrix} 2 & 6 & -8 \\ 0 & 1 & 5 \\ 0 & 0 & 3 \end{pmatrix} = TT^t$$

Remark

- $A = VD_\lambda V^t \ (A \geq 0) \ \Rightarrow \ A = A^{1/2}A^{1/2}$ with $A^{1/2} = VD_{\sqrt{\lambda}} V^t$

- $\exists\, A^{1/2}$ s.t. $A = A^{1/2}(A^{1/2})^t$

  ($\exists\, Q_{m \times m}$ s.t. $QQ^t = Q^t Q = I$) $\Rightarrow$ $A^{1/2} = VD_{\sqrt{\lambda}} Q^t$

  ($\because\ A^{1/2}(A^{1/2})^t = VD_{\sqrt{\lambda}} Q^t Q D_{\sqrt{\lambda}} V^t = VD_\lambda V^t = A$)

- $(A^{1/2})^t = A^{1/2}$

- $A^{-1} = A^{-1/2}(A^{-1/2})^t$

- If $A^{1/2}$: a lower triangular matrix with $a_{ii} \geq 0,\ i = 1, \ldots, m$,
  then $A = A^{1/2}(A^{1/2})^t$: Cholesky decomposition of A

Example 1

$$\underline{x}_{m\times 1} \sim (\underline{\mu}, \Sigma),\ \ \Sigma > 0$$

$\exists\ \Sigma^{1/2}$ s.t. $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^t \equiv \Sigma^{1/2}\Sigma^{1/2}$

$\Longrightarrow\ \underline{z} = \Sigma^{-1/2}(\underline{x} - \underline{\mu})$ where $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$

$\Longrightarrow$ · $E(\underline{z}) = \Sigma^{-1/2}E(\underline{x} - \underline{\mu}) = \Sigma^{-1/2}(\underline{\mu} - \underline{\mu}) = \underline{0}$

· $Var(\underline{z}) = Var(\Sigma^{-1/2}(\underline{x} - \underline{\mu})) = \Sigma^{-1/2}\, Var(\underline{x} - \underline{\mu})\, (\Sigma^{-1/2})^t$

$$= \Sigma^{-1/2}\, Var(\underline{x})(\Sigma^{-1/2})^t$$

$$= \Sigma^{-1/2}\Sigma(\Sigma^{-1/2})^t$$

$$= \Sigma^{-1/2}(\Sigma^{1/2}(\Sigma^{1/2})^t)(\Sigma^{-1/2})^t$$

$$= I_m \times (\Sigma^{-1/2}\Sigma^{1/2})^t = I_m \times I_m = I_m$$

Example 2

$d(\underline{z}, \underline{0})$= *Euclidean distance* between $\underline{z}$ and $\underline{0}$

$$d_I(\underline{z}, \underline{0}) = \| \underline{z} - \underline{0} \| = [(\underline{z} - \underline{0})^t (\underline{z} - \underline{0})]^{1/2}$$

$$= (\underline{z}^t \underline{z})^{1/2}$$

$$= [(\underline{x} - \underline{\mu})^t (\Sigma^{-1/2})^t \Sigma^{-1/2} (\underline{x} - \underline{\mu})]^{1/2}$$

$$= [(\underline{x} - \underline{\mu})^t \Sigma^{-1} (\underline{x} - \underline{\mu})]^{1/2} \quad (\because \Sigma^{-1} = (\Sigma^{-1/2})^t \Sigma^{-1/2})$$

$$= d_\Sigma(\underline{x}, \underline{\mu})$$: *Mahalanobis distance*

4.4 Kronecker product and Vec Operator

Kronecker product

- $A_{m\times n}$ and $B_{p\times q}$

  $\Longrightarrow$ Kronecker Product of $A_{m\times n}$ and $B_{p\times q}$

$$A\otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} : mp\times nq$$

Examples : $A=[0\ \ 1\ \ 2]$, $B=\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

1) $A\otimes B = [0B\ \ 1B\ \ 2B] = \begin{bmatrix} 0 & 0 & 1 & 2 & 2 & 4 \\ 0 & 0 & 3 & 4 & 6 & 8 \end{bmatrix}$

2) $B\otimes A = \begin{bmatrix} 1A & 2A \\ 3A & 4A \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 & 0 & 2 & 4 \\ 0 & 3 & 6 & 0 & 4 & 8 \end{bmatrix}$

```
A<-matrix(c(0, 1, 2), byrow=T, nrow=3)
At<-t(A)
B<-matrix(c(1, 2,
            3, 4),  byrow=T, nrow=2)
K<-kronecker(At, B)
```

Application : Multivariate Linear Model

– (Univariate) Linear Model: $\underset{n\times 1}{\underline{y}} = \underset{n\times p}{X} \; \underset{p\times 1}{\underline{b}} + \underset{n\times 1}{\underline{e}}$, $\underline{e} \sim N_n(\mathbf{0}, \sigma^2 I_n)$ and σ^2 is unknown.

$$\hat{\underline{b}} = (X^t X)^{-1} X^t \underline{y} \quad \sim N_n(\underline{b},\, (X^t X)^{-1}\sigma^2)$$

– Multivariate Linear Model : $\underset{n\times m}{Y} = \underset{n\times p}{X} \; \underset{p\times m}{B} + \underset{n\times m}{E}$, $E \sim N(0, I_n \otimes \Sigma) \Rightarrow Y \sim N(XB, I_n \otimes \Sigma)$

$$\hat{B} = (X^t X)^{-1} X^t Y \sim N(B, (X^t X)^{-1} \otimes \Sigma)$$

Properies

■ Basic

1) $\alpha \otimes A = A \otimes \alpha$, for any scalar α.
2) $(\alpha A) \otimes (\beta B) = (\alpha\beta) A \otimes B$, for any scalars α and β.
3) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$
4) $(A+B) \otimes C = (A \otimes C) + (B \otimes C)$, if A and B are of the same size.
5) $A \otimes (B+C) = (A \otimes B) + (A \otimes C)$, if B and C are of the same size.
6) $(A \otimes B)^t = A^t \otimes B^t$
7) $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, if $A_{m\times h}$, $B_{p\times k}$, $C_{h\times n}$, $D_{k\times q}$.

Trace, Determinant and Rank if $A_{m \times m}$, $B_{p \times p}$.

8) $tr(A \otimes B) = tr(A)tr(B)$

9) $|A \otimes B| = |A|^p |B|^m$

10) $rank(A \otimes B) = rank(A)rank(B)$

Inverse and g-Inverse if $A_{m \times n}$, $B_{p \times q}$.

11) $(A \otimes B)^{-1}$ = $A^{-1} \otimes B^{-1}$, if $m = n, p = q$, and $A \otimes B$ is nonsingular.

12) $(A \otimes B)^{-}$ = $A^{-} \otimes B^{-}$

Vec operator

Def) Operator transforming a matrix to a vector

$$A_{m \times n} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n] \text{ with } \boldsymbol{a}_j = (a_{1j}, a_{2j}, \cdots, a_{mj})^t,$$

$$\Rightarrow vec(A) = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} : mn \times 1 \text{ vector}$$

Example	

$A = \begin{bmatrix} 1 & 2 & 5 \\ 3 & 4 & 6 \end{bmatrix} \quad \Rightarrow \; vec(A) = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \\ 5 \\ 6 \end{bmatrix}$	`vec <- function(x) t(t(as.vector(x)))`
	`A <- matrix(c(1,2,5,3,4,6), nrow=2, ncol=3, byrow=T)`
	`A`
	`vec(A)`

Application: $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_p)$ with $Cov(\boldsymbol{x}_j) = \Sigma,\ j = 1,..,p$

$$\Rightarrow \; vec(X) = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_p \end{bmatrix} \quad \Rightarrow \; Cov[vec(X)] = \begin{bmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Sigma \end{bmatrix} = I_n \otimes \Sigma \; : \; np \times np$$

Properies

■ Basic

$\boldsymbol{a}$ and $\boldsymbol{b}$: any vectors, A and B : same size

1) $vec(\boldsymbol{a}) = vec(\boldsymbol{a}^t) = \boldsymbol{a}$

2) $vec(\boldsymbol{a}\boldsymbol{b}^t) = \boldsymbol{b} \otimes \boldsymbol{a}$

3) $vec(\alpha A + \beta B) = \alpha\, vec(A) + \beta\, vec(B)$, where α and β are scalars.

■ Trace if $A_{m\times n},\ B_{n\times p},\ C_{p\times q},\ D_{q\times m}$

4) $tr(A^t B) = [vec(A)]^t vec(B)$

5) $vec(ABC) = (C^t \otimes A) vec(B)$

6) $tr(ABCD) = [vec(A^t)]^t (D^t \otimes B) vec(C)$