第 1 次作业

1. $\mathbf{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)'=\begin{pmatrix}\boldsymbol{x}_1{}'\\ \boldsymbol{x}_2{}'\\ \vdots\\ \boldsymbol{x}_n'\end{pmatrix}=\begin{pmatrix}x_{11}&x_{12}&\cdots&x_{1p}\\ x_{21}&x_{22}&\cdots&x_{2p}\\ \vdots&\vdots&\ddots&\vdots\\ x_{n1}&x_{n2}&\cdots&x_{np}\end{pmatrix}_{n\times p}$，$\bar{\boldsymbol{x}}=\frac{1}{n}\sum_{i=1}^n\boldsymbol{x}_i=\frac{1}{n}\mathbf{X}'\mathbf{1}_n$，

其中 $\mathbf{1}_n=\begin{pmatrix}1\\1\\ \vdots\\1\end{pmatrix}_{p\times 1}$。$\hat{\boldsymbol{\Sigma}}=\frac{1}{n}\sum_{i=1}^n(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'=\mathbf{X}'\mathbf{A}\mathbf{X}$，问：$\mathbf{A}=?$

Solution. 首先要知道，$\sum_{i=1}^n\boldsymbol{x}_i\boldsymbol{x}_i'=\mathbf{X}'\mathbf{X}$，$\bar{\boldsymbol{x}}=\frac{1}{n}\mathbf{X}'\mathbf{1}_n$。

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}}&=\frac{1}{n}\sum_{i=1}^n(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'\\
&=\frac{1}{n}\sum_{i=1}^n(\boldsymbol{x}_i\boldsymbol{x}_i'-\boldsymbol{x}_i\bar{\boldsymbol{x}}'-\bar{\boldsymbol{x}}\boldsymbol{x}_i'+\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}')\\
&=\frac{1}{n}\left(\sum_{i=1}^n\boldsymbol{x}_i\boldsymbol{x}_i'-\left(\sum_{i=1}^n\boldsymbol{x}_i\right)\bar{\boldsymbol{x}}'-\bar{\boldsymbol{x}}\left(\sum_{i=1}^n\boldsymbol{x}_i'\right)+n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'\right)\\
&=\frac{1}{n}\left(\sum_{i=1}^n\boldsymbol{x}_i\boldsymbol{x}_i'-n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'-\bar{\boldsymbol{x}}\cdot n\bar{\boldsymbol{x}}'+n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'\right)\\
&=\frac{1}{n}\left(\sum_{i=1}^n\boldsymbol{x}_i\boldsymbol{x}_i'-n\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}'\right)\\
&=\frac{1}{n}\left(\mathbf{X}'\mathbf{X}-n\left(\frac{1}{n}\mathbf{X}'\mathbf{1}_n\right)\left(\frac{1}{n}\mathbf{1}_n'\mathbf{X}\right)\right)\\
&=\frac{1}{n}\mathbf{X}'\left(\mathbf{I}_n-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{X}\\
&\triangleq\mathbf{X}'\mathbf{A}\mathbf{X}\Rightarrow\mathbf{A}=\frac{1}{n}\left(\mathbf{I}_n-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)
\end{aligned}
$$

■

第 2 次作业

1. 若 $A>0$，$B>0$，$A-B>0$，则 $B^{-1}-A^{-1}>0$，且 $|A|>|B|$。

Proof. 思路 1：首先给出分块矩阵的逆，$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$，若 A 为可逆矩阵，A_{11} 为方阵

(1) 若 $|A_{11}|\neq 0$，则 $A^{-1}=\begin{pmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1} \\ -A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1} & A_{22\cdot 1}^{-1} \end{pmatrix}$

(2) 若 $|A_{22}|\neq 0$，则 $A^{-1}=\begin{pmatrix} A_{11\cdot 2}^{-1} & -A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} & A_{22}^{-1}+A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} \end{pmatrix}$

(3) 特别地，当 $|A_{11}|\neq 0$ 且 $|A_{22}|\neq 0$ 时

$$A_{11\cdot 2}^{-1}=\left(A_{11}-A_{12}A_{22}^{-1}A_{21}\right)^{-1}=A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1}$$
$$A_{22\cdot 1}^{-1}=\left(A_{22}-A_{21}A_{11}^{-1}A_{12}\right)^{-1}=A_{22}^{-1}+A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1}$$

$$\begin{aligned} A&=B-(B-A)=B-C \\ A^{-1}&=(B-C)^{-1}\triangleq\left(A_{11}-A_{12}A_{21}^{-1}A_{21}\right)^{-1} \quad (\text{其中} A_{11}=B, A_{12}=A_{21}=I, A_{22}=C^{-1}) \\ &=\left(A_{11}-A_{12}A_{22}^{-1}A_{21}\right)^{-1}=A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22\cdot 1}^{-1}A_{21}A_{11}^{-1} \\ &=B^{-1}+B^{-1}\left(A_{22}-A_{21}A_{11}^{-1}A_{12}\right)^{-1}B^{-1} \\ &=B^{-1}-B^{-1}\left((A-B)^{-1}+B^{-1}\right)^{-1}B^{-1} \end{aligned}$$

因此，$B^{-1}-A^{-1}=B^{-1}\left((A-B)^{-1}+B^{-1}\right)^{-1}B^{-1}>0$。

$A>B$，令 $A=(A-B)+B$，基于 weyl 不等式，$\lambda_i(A)\geqslant\lambda_i(B)\Rightarrow|A|>|B|$。 ■

Proof. 思路 2：

$$\begin{aligned} &A-B>0\Rightarrow B^{-\frac{1}{2}}(A-B)B^{-\frac{1}{2}}>0 \\ \Rightarrow& B^{-\frac{1}{2}}AB^{-\frac{1}{2}}>I \quad \left(\text{the eigenvalues of } B^{-\frac{1}{2}}AB^{-\frac{1}{2}}>1\right) \\ \Rightarrow& B^{\frac{1}{2}}A^{-1}B^{\frac{1}{2}}<I \\ \Rightarrow& B^{\frac{1}{2}}\left(A^{-1}-B^{-1}\right)B^{\frac{1}{2}}<0\Rightarrow A^{-1}<B^{-1}. \end{aligned}$$

已知

$$\begin{aligned} &\left|\lambda I-B^{-\frac{1}{2}}AB^{-\frac{1}{2}}\right| \\ =&\left|B^{-\frac{1}{2}}\right|\left|\lambda I-B^{-1}A\right|\left|B^{-\frac{1}{2}}\right| \\ =&\left|\lambda I-B^{-1}A\right|, \end{aligned}$$

则 the eigenvalues of $B^{-1}A > 1 \Rightarrow |A| = |B|\,|B^{-1}A| > |B|$。 ■

Proof. 思路 3：$A > B$，$\lambda_i \triangleq \lambda_i(A) \geqslant \lambda_i(B) \triangleq \gamma_i$，$B^{-1} - A^{-1}$ 的特征值为

$$\frac{1}{\gamma_i} - \frac{1}{\lambda_i} = \frac{\lambda_i - \gamma_i}{\gamma_i \lambda_i} > 0 \Rightarrow B^{-1} - A^{-1} > 0$$

■

2. 设 A 和 B 分别为 $p \times q$ 和 $q \times p$ 的矩阵，则 $|I_p + AB| = |I_q + BA|$。

Proof.

$$\begin{bmatrix} I_p & A \\ 0 & I_q \end{bmatrix} \begin{bmatrix} I_p & -A \\ B & I_q \end{bmatrix} = \begin{bmatrix} I_p + AB & 0 \\ B & I_q \end{bmatrix}$$

$$\begin{bmatrix} I_p & 0 \\ -B & I_q \end{bmatrix} \begin{bmatrix} I_p & -A \\ B & I_q \end{bmatrix} = \begin{bmatrix} I_p & -A \\ B & I_q + BA \end{bmatrix}.$$

两边同时取行列式得，$\begin{vmatrix} I_p & -A \\ B & I_q \end{vmatrix} = |I_p + AB| = |I_q + BA|$。直接利用分块矩阵的知识，令 $A_{11} = I_p$，$A_{12} = A$，$A_{21} = -B$，$A_{22} = I_q$，则

$$\begin{aligned} |A| =& |A_{22}||A_{11\cdot 2}| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}| = |I_p + AB| \\ =& |A_{11}||A_{22\cdot 1}| = |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}| = |I_q + BA| \end{aligned}$$

■

第 3 次作业

1. $X \sim N_p(\mu, \Sigma)$，A 和 B 是对称矩阵，证明 $\mathrm{Cov}(X'AX, X'BX) =?$

Proof. 首先给出一些结论：设 $\boldsymbol{Y} \sim N_p(\mathbf{0}, \mathbf{I}_p)$, 则有

a. 设 a 为 p 元向量, A 为对称矩阵, 则

$$\mathrm{Cov}(a'Y, Y'AY) = 0$$

Proof. $a=(a_1,\cdots,a_p)'$, $a'Y=\sum_{i=1}^p a_iY_i$,

$$
\begin{aligned}
Y'AY &= (Y_1,\cdots,Y_p)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}\begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} \\
&= \left(\sum_{i=1}^p a_{i1}Y_i, \sum_{i=1}^p a_{i2}Y_i, \cdots, \sum_{i=1}^p a_{ip}Y_i\right)\begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} \\
&= \sum_{i=1}^p \sum_{j=1}^p a_{ij}Y_iY_j
\end{aligned}
$$

$$
\begin{aligned}
\text{Cov}\left(a'Y, Y'AY\right) =& \text{Cov}\left(\sum_{i=1}^P a_iY_i, \sum_{i=j=1}^p \sum_{j=1}^p a_{ij}Y_iY_j\right) \\
&= \sum_{i=1}^P a_ia_{ii}E\left[\left(Y_i-EY_i\right)\left(Y_i^2-EY_i{}^2\right)\right] \\
&= \sum_{i=1}^P a_ia_{ii}\left[E\left(Y_i^3\right)-EY_iEY_i^2\right] \\
&= \sum_{i=1}^P a_ia_{ii}E\left(Y_i^3\right)=0
\end{aligned}
$$

■

b. 设 A,B 为对称矩阵，则

$$\text{Cov}\left(Y'AY, Y'BY\right)=2\text{tr}(AB)$$

Proof. 由于 $Y'AY = \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}Y_iY_j, Y'BY = \sum_{k=1}^{p}\sum_{l=1}^{p} b_{kl}Y_kY_l$,

$$\operatorname{Cov}(Y'AY, Y'BY) = E(Y'AYY'BY) - \underbrace{E(Y'AY)}_{tr(A)}\underbrace{E(Y'BY)}_{tr(B)}$$

$$E(Y'AYY'BY) = E\left[\sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}Y_iY_j \cdot \sum_{k=1}^{p}\sum_{l=1}^{p} b_{kl}Y_kY_l\right]$$

$$E(Y_iY_jY_kY_l) = \begin{cases} 3, & i=j=k=l \to EY^4 = \prod_{i=1}^{2}(2i-1) = 3 \\ 1, & i=j\neq k=l \\ & i=k\neq j=l \\ 0, & \text{其他} \end{cases}$$

则

$$\begin{aligned} E(Y'AYY'BY) =& E\left[\sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}Y_iY_j \cdot \sum_{k=1}^{p}\sum_{l=1}^{p} b_{kl}Y_kY_l\right] \\ =& 3\sum_{i=1}^{p} a_{ii}b_{ii} + \sum_{1\leq i\neq k\leq p} a_{ii}b_{kk} + \sum_{1\leq i\neq j\leq p} a_{ij}b_{ij} + \sum_{1\leq i\neq k\leq p} a_{ik}b_{ki} \\ =& 3\sum_{i=1}^{p} a_{ii}b_{ii} + \sum_{1\leq i\neq k\leq p} a_{ii}b_{kk} + 2\sum_{1\leq i\neq j\leq p} a_{ij}b_{ij} \\ =& 3\sum_{i=1}^{p} a_{ii}b_{ii} + \sum_{i=1}^{p}\sum_{k=1}^{p} a_{ii}b_{kk} - \sum_{i=1}^{p} a_{ii}b_{ii} + 2\sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}b_{ij} - 2\sum_{i=1}^{p} a_{ii}b_{ii} \\ =& \sum_{i=1}^{p}\sum_{k=1}^{p} a_{ii}b_{kk} + 2\sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}b_{ij} \\ =& \operatorname{tr}(A)\operatorname{tr}(B) + 2\operatorname{tr}(AB), \end{aligned}$$

则 $\operatorname{Cov}(Y'AY, Y'BY) = \operatorname{tr}(A)\operatorname{tr}(B) + 2\operatorname{tr}(AB) - \operatorname{tr}(A)\operatorname{tr}(B) = 2\operatorname{tr}(AB)$。 ■

设 $X=\mu+CY$，$Y\sim N_p(0,I_p)$，$CC'=\Sigma$

$$
\begin{aligned}
&\operatorname{Cov}\left(X'AX, X'BX\right)\\
=&\operatorname{Cov}\left((\mu+CY)'A(\mu+CY),(\mu+CY)'B(\mu+CY)\right)\\
=&\operatorname{Cov}\left(\mu'A\mu+\mu'ACY+Y'C'A\mu+Y'C'ACY,\mu'B\mu+\mu'BCY+Y'C'B\mu+Y'C'BCY\right)\\
=&\operatorname{Cov}\left(2\mu'ACY+Y'C'ACY,2\mu'BCY+Y'C'BCY\right)\\
=&\operatorname{Cov}\left(2\mu'ACY,2\mu'BCY\right)+\operatorname{Cov}\left(Y'\underbrace{C'AC}_{\text{symmetirc}}Y,Y'\underbrace{C'BC}_{\text{symmetirc}}Y\right)\\
=&4\mu'AC\operatorname{Var}(Y)C'B\mu+2\operatorname{tr}\left(C'ACC'BC\right)\\
=&4\mu'A\Sigma B\mu+2\operatorname{tr}\left(C'A\Sigma BC\right)\\
=&4\mu'A\Sigma B\mu+2\operatorname{tr}\left(A\Sigma BCC'\right)\\
=&4\mu'A\Sigma B\mu+2\operatorname{tr}(A\Sigma B\Sigma)
\end{aligned}
$$

■

2. 求 $\operatorname{Var}(S^2)=?$

Solution. 设 $X\sim N(0,1)$，一组大小为 n 的随机样本为 $x=(x_1,x_2,\ldots,x_n)'$

$$
\begin{aligned}
S^2=&\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\\
=&\frac{1}{n-1}x'\left(I_n-\frac{1}{n}1_n1_n'\right)x.
\end{aligned}
$$

已知，$\operatorname{Var}(X'AX)=2\operatorname{tr}(A\Sigma A\Sigma)+4\mu'A\Sigma A\mu$，将 $A=\dfrac{1}{n-1}\left(I_n-\dfrac{1}{n}1_n1_n'\right)$，$\Sigma=I_n$，$\mu=0_n$ 代入即可得 $\operatorname{Var}(S^2)$。 ■

3. 用特征函数证明定理 2.2.3：服从正态分布的随机向量的线性变换仍服从正态分布。

Proof. 设 $X\sim N_p(\mu,\Sigma), B$ 为 $q\times p$ 的常数矩阵, θ 为 $q\times 1$ 常向量, 令 $Z=BX+\theta$, 则 $Z\sim N_q\left(B\mu+\theta,B\Sigma B'\right)$。随机向量 X 的特征函数为

$$
\varphi_X(t)=\exp\left\{it'\mu-\frac{1}{2}t'\Sigma t\right\}
$$

由特征函数的性质, 得 $Z=BX+\theta$ 的特征函数为

$$
\begin{aligned}
\varphi_Z(s) &= \exp\left\{is'\theta\right\}\varphi_X\left(B's\right) \\
&= \exp\left\{is'\theta\right\}\exp\left\{is'B\mu-\frac{1}{2}s'B\Sigma B's\right\} \\
&= \exp\left\{is'(B\mu+\theta)-\frac{1}{2}s'B\Sigma B's\right\},
\end{aligned}
$$

则 $Z\sim N_q\left(B\mu+\theta, B\Sigma B'\right)$。 ■

第 4 次作业

1. 设 $\xi\sim N_p(\mu,\Sigma)$, 证明 $E\left(\xi_i-\mu_i\right)\left(\xi_j-\mu_j\right)\left(\xi_k-\mu_k\right)\left(\xi_l-\mu_l\right)=\sigma_{ij}\sigma_{kl}+\sigma_{ik}\sigma_{jl}+\sigma_{il}\sigma_{jk}$。

   参考链接：Isserlis 定理：如何计算多元正态分布的高阶矩？

2. 设 $y_i=x_i^\top\beta+\varepsilon_i$, $\varepsilon_i\sim N(0,\sigma^2)$, $i=1,2,\ldots,n$, 论证 $\hat\beta$ 与 $\hat\sigma^2$ 相互独立。

Proof. 回归系数 β 和方差 σ^2 的估计为:

$$
\begin{aligned}
\hat\beta = \left(X'X\right)^{-1}X'Y &= \left(X'X\right)^{-1}X'(X\beta+\varepsilon) \\
&= \beta+\left(X'X\right)^{-1}X'\varepsilon,
\end{aligned}
$$

$$
(n-p)\hat\sigma^2=\sum_{i=1}^{n}\left(y_i-\hat y_i\right)^2=(Y-\hat Y)'(Y-\hat Y),
$$

其中 $\hat Y=x\hat\beta=X\left(X'X\right)^{-1}X'Y\triangleq HY$，

$$
\begin{aligned}
Y-\hat Y = Y-HY = (I-H)Y &= (I-H)(X\beta+\varepsilon) \\
&= (I-H)\varepsilon,
\end{aligned}
$$

其中 $\varepsilon\sim N_n\left(0_n,\sigma^2 I_n\right)$，$H=X(X'X)^{-1}X'$ 为帽子/投影矩阵，是对称幂等矩阵，$I-H$ 也是对称幂等矩阵，则：

$$
\hat\sigma^2=\frac{1}{n-p}\varepsilon'(I-H)(I-H)\varepsilon=\frac{1}{n-p}\varepsilon'(I-H)\varepsilon.
$$

基于以下定理：

线性型和二次型的独立性

定理 2

设$\mathbf{X} \sim N_n(\boldsymbol{\mu}, \sigma^2 I_n)$, A是对称矩阵，B是$m \times n$矩阵，令$\xi = \mathbf{X}^T A \mathbf{X}$, $\mathbf{Z} = B\mathbf{X}$，若$BA = 0$，则ξ和$\mathbf{Z}$相互独立。

证明： 设$\text{rank}(A) = r > 0$，则存在正交矩阵Γ使得

$$\Gamma^T A \Gamma = \text{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0),$$

其中λ_i是A的非零特征值，$i = 1, \cdots, r$。注意到：

$$\begin{aligned} BA &= B\Gamma \text{diag}(\lambda_1, \cdots, \lambda_r, 0, \cdots, 0)\Gamma^T \\ &= (C_1 \ C_2) \begin{pmatrix} \Lambda_r & 0 \\ 0 & 0 \end{pmatrix} \Gamma^T \\ &= (C_1 \Lambda_r \ 0)\Gamma^T = 0. \end{aligned}$$

根据条件$BA = 0$，可得$C_1 = 0$。

线性型和二次型的独立性

正态分布的二次型和线性型
郭旭
两个例子
二次型的分布及其独立性
一般二次型的分布性质

证明： 令$\mathbf{Y} = \Gamma^T \mathbf{X}$，即$\mathbf{X} = \Gamma \mathbf{Y}$，则$\mathbf{Y} \sim N_n(\Gamma^T \boldsymbol{\mu}, \sigma^2 I_n)$，即$Y_1, \cdots, Y_n$相互独立。注意到：

$$\xi = \mathbf{X}^T A \mathbf{X} = \mathbf{Y}^T \Gamma^T A \Gamma \mathbf{Y} = \sum_{i=1}^{r} \lambda_i Y_i^2;$$

$$\mathbf{Z} = B\mathbf{X} = B\Gamma\mathbf{Y} = (C_1 \ C_2) \begin{pmatrix} Y_1 \\ \cdots \\ Y_n \end{pmatrix} = C_2 \begin{pmatrix} Y_{r+1} \\ \cdots \\ Y_n \end{pmatrix}.$$

由于$Y_1, \cdots, Y_r$和$Y_{r+1}, \cdots, Y_n$相互独立，因此ξ和$\mathbf{Z}$相互独立。

因为 $I-H$ 是对称矩阵，$(X'X)^{-1}X'(I-H)=(X'X)^{-1}X'-(X'X)^{-1}X'X\,(X'X)^{-1}X'=0$，则 $(X'X)^{-1}X'\varepsilon$ 与 $\varepsilon'(I-H)\varepsilon$ 相互独立，从而 $\hat\beta$ 与 $\hat\sigma^2$ 相互独立。

■

第 5 次作业

1. 设 $X\sim N_3(\mu,\Sigma)$, 其中

$$\mu=\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix},\Sigma=\begin{pmatrix}1&\rho&\rho\\ \rho&1&\rho\\ \rho&\rho&1\end{pmatrix},\quad 0<\rho<1$$

试求条件分布 $\begin{pmatrix}X_1 & X_2\end{pmatrix}\mid X_3$ 和 $X_1\mid\begin{pmatrix}X_2 & X_3\end{pmatrix}$。

Solution. 首先给出多元正态随机变量的条件分布：设 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$，$p\geq 2$，$\boldsymbol{\Sigma}>0$，对 $\boldsymbol{X}$，$\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 作如下剖分，

$$\boldsymbol{X}=\begin{bmatrix}\boldsymbol{X}^{(1)}\\ \boldsymbol{X}^{(2)}\end{bmatrix},\quad \boldsymbol{\mu}=\begin{bmatrix}\boldsymbol{\mu}^{(1)}\\ \boldsymbol{\mu}^{(2)}\end{bmatrix},\quad \boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{11}&\boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21}&\boldsymbol{\Sigma}_{22}\end{bmatrix}$$

其中

- $\boldsymbol{X}^{(1)}$ 和 $\boldsymbol{\mu}^{(1)}$ 为 $q\times 1$ 的向量
- $\boldsymbol{\Sigma}_{11}$ 为 $q\times q$ 矩阵
- $\boldsymbol{X}^{(2)}$ 和 $\boldsymbol{\mu}^{(2)}$ 为 $(p-q)\times 1$ 的向量
- $\boldsymbol{\Sigma}_{22}$ 为 $(p-q)\times(p-q)$ 矩阵
- $\boldsymbol{\Sigma}_{12}=\boldsymbol{\Sigma}_{21}'$ 为 $q\times(p-q)$ 矩阵

(1) 给定 $\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}$ 时 $\boldsymbol{X}^{(1)}$ 的条件分布服从 q 元正态分布, 即

$$\left(\boldsymbol{X}^{(1)}\mid\boldsymbol{X}^{(2)}=\boldsymbol{x}^{(2)}\right)\sim N_q\left(\boldsymbol{\mu}_{1\cdot 2},\boldsymbol{\Sigma}_{11\cdot 2}\right)$$

其中 $\boldsymbol{\mu}_{1\cdot 2}=\boldsymbol{\mu}^{(1)}+\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\boldsymbol{x}^{(2)}-\boldsymbol{\mu}^{(2)}\right),\boldsymbol{\Sigma}_{11\cdot 2}=\boldsymbol{\Sigma}_{11}-\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$。

(2) 给定 $\boldsymbol{X}^{(1)}=\boldsymbol{x}^{(1)}$ 时 $\boldsymbol{X}^{(2)}$ 的条件分布服从 $p-q$ 元正态分布, 即

$$\left(\boldsymbol{X}^{(2)}\mid\boldsymbol{X}^{(1)}=\boldsymbol{x}^{(1)}\right)\sim N_{p-q}\left(\boldsymbol{\mu}_{2\cdot 1},\boldsymbol{\Sigma}_{22\cdot 1}\right)$$

其中 $\boldsymbol{\mu}_{2\cdot 1}=\boldsymbol{\mu}^{(2)}+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{x}^{(1)}-\boldsymbol{\mu}^{(1)}\right),\boldsymbol{\Sigma}_{22\cdot 1}=\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$。

这里 $p=3$，$q=2$。设 $X=\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} =: \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$，$\mu=\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}$ 以及 $\Sigma=\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} =: \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$。

则 $(X_1,X_2) \mid X_3=x_3 =: X^{(1)} \mid X^{(2)}=x^{(2)} \sim N_2(\mu_{1\cdot 2},\Sigma_{11\cdot 2})$，其中

$$\begin{aligned}
\mu_{1\cdot 2} &= \mu_{11}+\Sigma_{12}\Sigma_{22}^{-1}\left(x^{(2)}-\mu^{(2)}\right) \\
&= \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \rho \\ \rho \end{pmatrix}(x_3-\mu_3) = \begin{pmatrix} \mu_1+\rho(x_3-\mu_3) \\ \mu_2+\rho(x_3-\mu_3) \end{pmatrix} \\
\Sigma_{11\cdot 2} &= \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\
&= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} \rho \\ \rho \end{pmatrix}\begin{pmatrix} \rho & \rho \end{pmatrix} = \begin{pmatrix} 1-\rho^2 & \rho-\rho^2 \\ \rho-\rho^2 & 1-\rho^2 \end{pmatrix}
\end{aligned}$$

从而得到

$$(X_1,X_2) \mid X_3=x_3 \sim N_2\left(\begin{pmatrix} \mu_1+\rho(x_3-\mu_3) \\ \mu_2+\rho(x_3-\mu_3) \end{pmatrix}, \begin{pmatrix} 1-\rho^2 & \rho-\rho^2 \\ \rho-\rho^2 & 1-\rho^2 \end{pmatrix}\right)$$

同理可得

$$X_1 \mid (X_2,X_3) \sim N\left(\mu_1+\frac{\rho}{1+\rho}(x_2+x_3-\mu_2-\mu_3), 1-\frac{2\rho^2}{1+\rho}\right)$$

■

2. 设 $X=\begin{pmatrix} X_1 & X_2 \end{pmatrix}' \sim N_2(0,I_2)$, 试求在 $\boldsymbol{X}_1+\boldsymbol{X}_2$ 给定下 $\boldsymbol{X}_1$ 的条件分布。

Solution. $X=\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(0,I_2)$，则

$$\begin{pmatrix} X_1 \\ X_1+X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right)$$

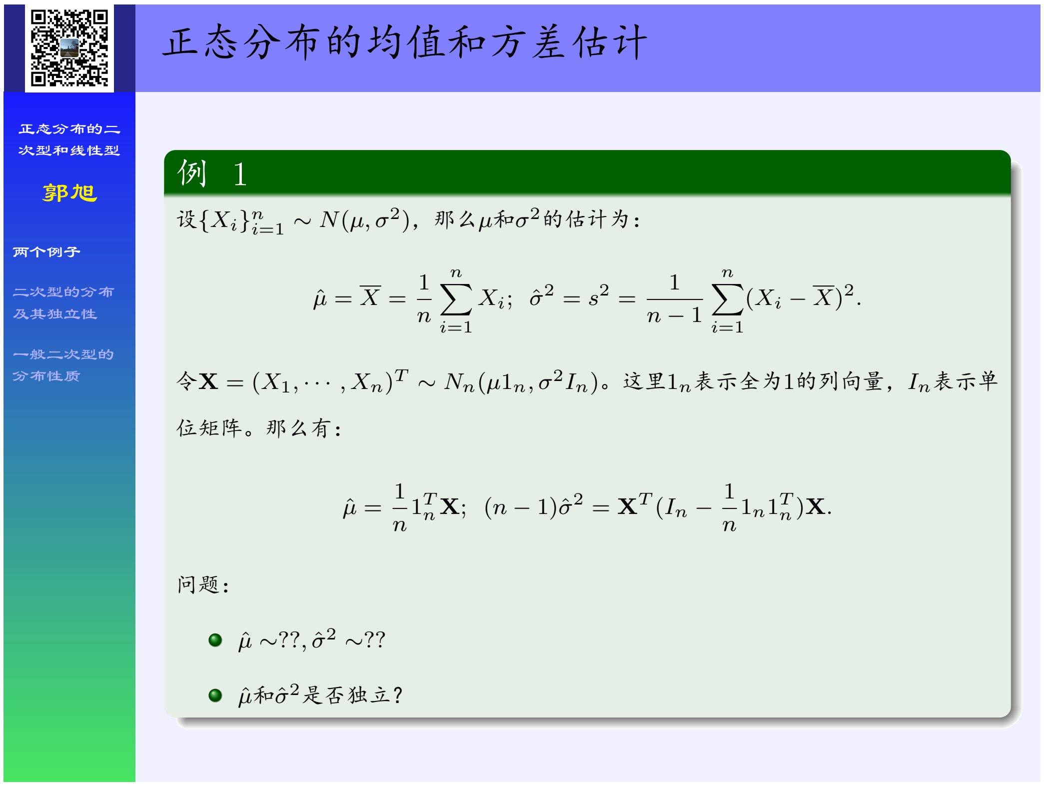
类似于第 1 题的 solution，可得

$$(X_1 \mid X_1+X_2=x_1+x_2) \sim N\left(\frac{1}{2}(x_1+x_2), \frac{1}{2}\right)$$

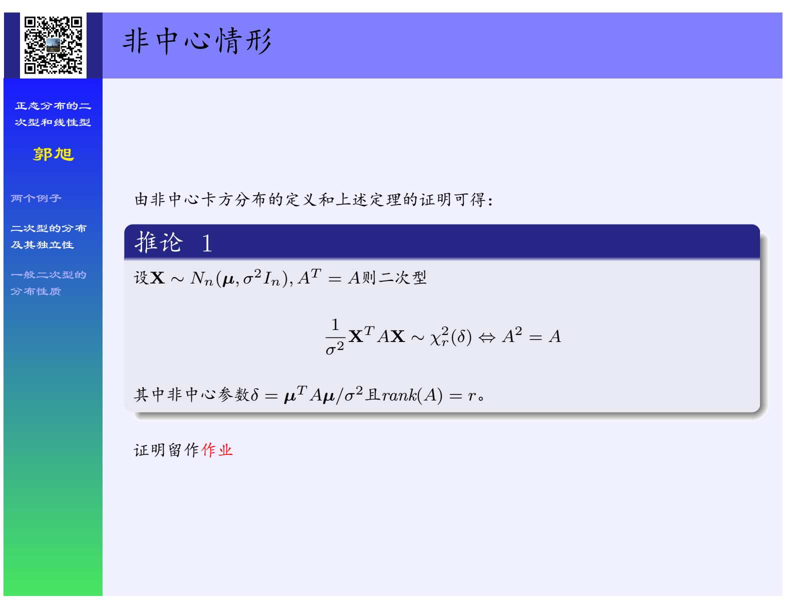
■

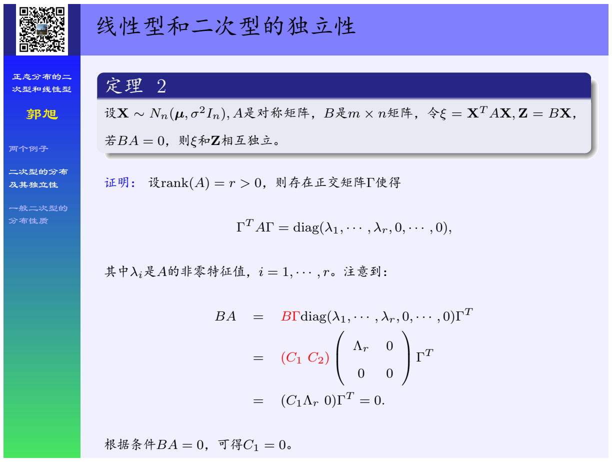
第 6 次作业

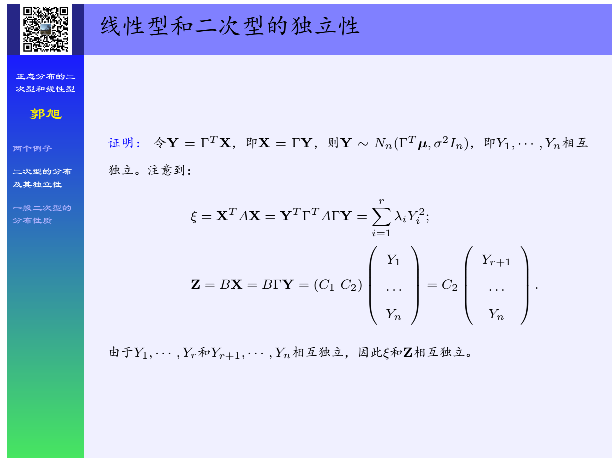
1. 利用正态分布二次型和线性型的知识证明以下问题。



Solution. 主要用到推论 1 和定理 1：







已知 $X = (X_1, \cdots, X_n)^T \sim N_n(\mu 1_n, \sigma^2 I_n)$，以及

$$\hat{\mu} = \frac{1}{n} 1_n^T X =: BX,$$
$$(n-1)\hat{\sigma}^2 = X^T \left(I_n - \frac{1}{n} 1_n 1_n^T \right) X =: X^T A X.$$

则 $\hat{\mu} \sim N(B\mu 1_n, B\sigma^2 I_n B^T) = N(\mu, \dfrac{\sigma^2}{n})$。

注意到：$A^2 = A$ 且 $A^T = A$, 故 A 是对称幂等矩阵, 可得：

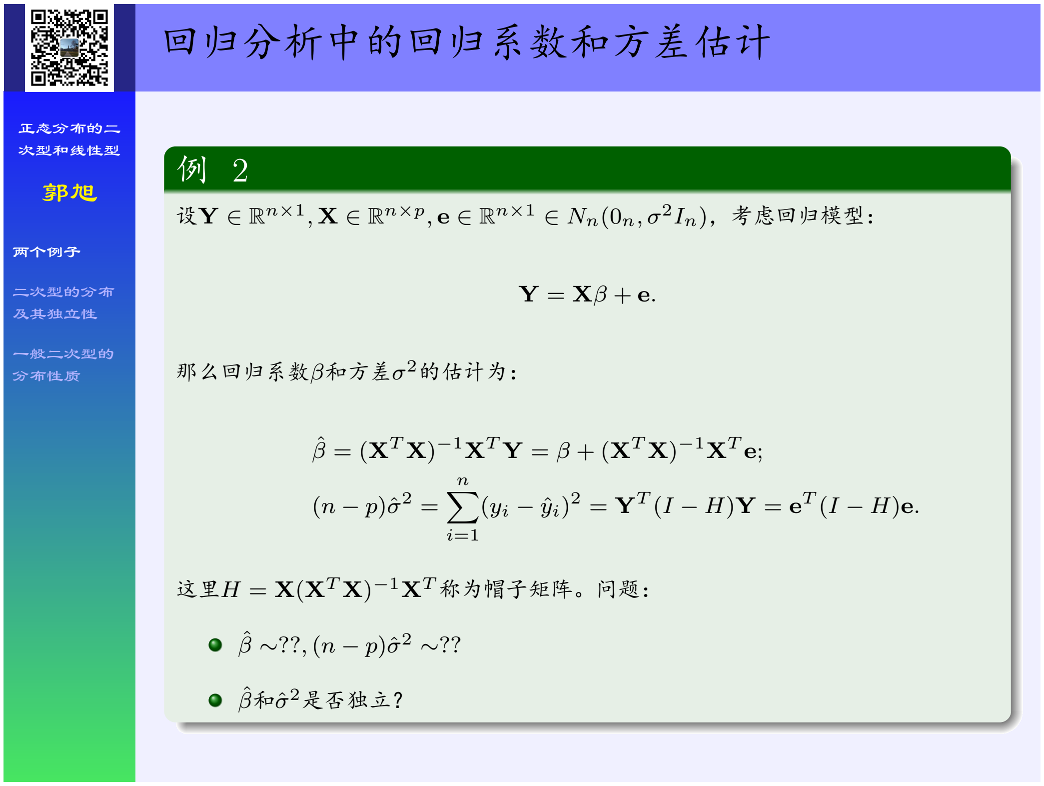
$$\mathrm{rank}(A) = \mathrm{tr}(A) = n - \frac{1}{n} 1_n^T 1_n = n - 1.$$

根据推论 1，我们有 $\dfrac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi_r^2(\delta)$，其中 $r = \mathrm{rank}(A) = n-1$，$\delta = \dfrac{\mu 1_n A (\mu 1_n)^T}{\sigma^2} = \dfrac{\mu^2 1_n^T \left(I_n - \frac{1}{n} 1_n 1_n^T\right) 1_n}{\sigma^2} = 0$，从而

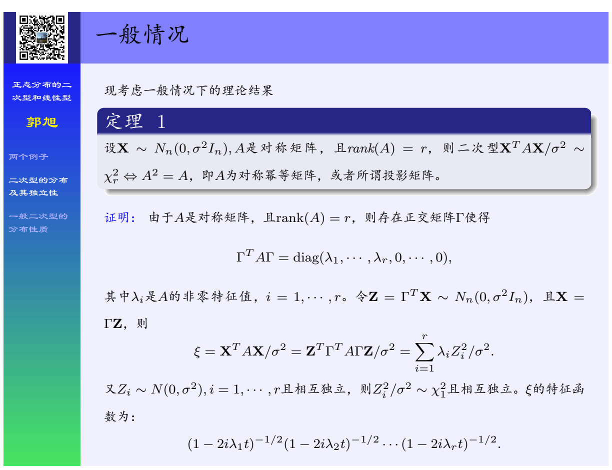
$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-1}^2.$$

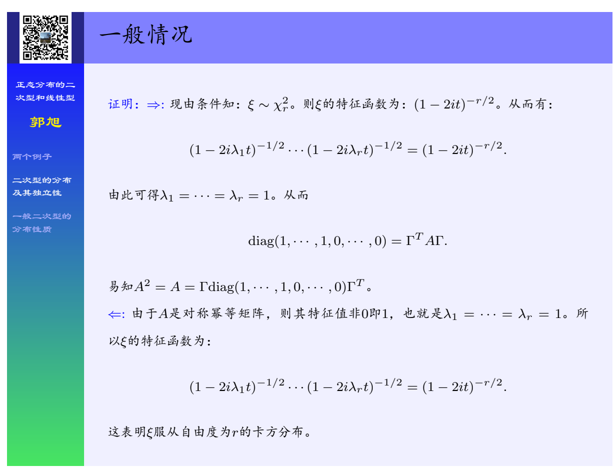
利用定理 2，因为 $BA = \dfrac{1}{n} 1_n^T \left(I_n - \dfrac{1}{n} 1_n 1_n^T \right) = 0$，则 $\hat{\mu}$ 与 $\hat{\sigma}^2$ 相互独立。 ■

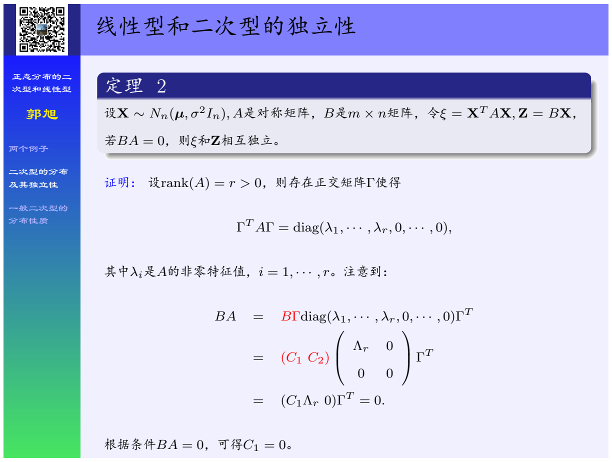
2. 利用正态分布二次型和线性型的知识证明以下问题。

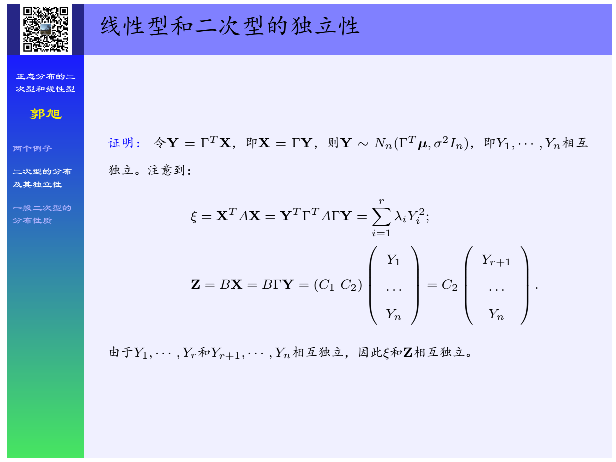


Solution. 主要用到定理 1 和定理 2：









已知 $e\sim N_n\left(0_n,\sigma^2I_n\right)$，以及

$$\hat\beta=\beta+\left(X^TX\right)^{-1}X^Te=:\beta+Be,$$
$$(n-p)\hat\sigma^2=e^T(I-H)e=:e^TAe.$$

则 $\hat\beta\sim N_n(\beta,B\sigma^2I_nB^T)=N_n(\beta,\sigma^2(X^TX)^{-1})$，又已知 A 是对称幂等矩阵，可得

$$\begin{aligned}\mathrm{rank}(A)=\mathrm{tr}(A)=&n-\mathrm{tr}(X(X^TX)^{-1}X^T)\\ =&n-\mathrm{tr}((X^TX)^{-1}X^TX)\\ =&n-\mathrm{tr}(I_p)=n-p.\end{aligned}$$

根据定理 1 得

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

利用定理 2，$BA=(X^TX)^{-1}X^T(I-H)=0$，则 $\hat\beta$ 与 $\hat\sigma^2$ 相互独立。 ■

第 7 次作业

1. 设 p 维随机向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)'\sim N_p(\boldsymbol{\mu},\mathbf{I}_p)$，$y=\boldsymbol{x}'\boldsymbol{x}\sim\chi^2(p,\lambda),\lambda=\boldsymbol{\mu}'\boldsymbol{\mu}$, 证明：

(1) $\mathrm{E}(y)=p+\lambda$

(2) $\mathrm{Var}(y)=2p+4\lambda$

Solution. 思路 1：因为 $\boldsymbol{x}=(x_1,x_2,\cdots,x_p)'\sim N_p(\boldsymbol{\mu},\mathbf{I}_p)$，则 $x_1,x_2,\cdots,x_p$ 相互独立，$x_i\sim N(\mu_i,1),i=1,2,\cdots,p$。从而可得

$$\mathrm{E}(y)=\mathrm{E}(\sum_{i=1}^{p}x_i^2)=\sum_{i=1}^{p}\mathrm{E}x_i^2=\sum_{i=1}^{p}(1+\mu_i^2)=p+\lambda$$

$$\begin{aligned}\mathrm{Var}(y)=\mathrm{Var}(\sum_{i=1}^{p}x_i^2)&=\sum_{i=1}^{p}\mathrm{Var}(x_i^2)\\&=\sum_{i=1}^{p}[\mathrm{E}x_i^4-(\mathrm{E}x_i^2)^2]\\&=\sum_{i=1}^{p}[(\mu_i^4+6\mu_i^2+3)-(1+\mu_i^2)^2]\\&=\sum_{i=1}^{p}(4\mu_i^2+2)=4\lambda+2p\end{aligned}$$

■

Solution. 思路 2：利用二次型的性质，设多元正态随机变量 $X\sim N_p(\mu,\Sigma)$，A 和 B 为 p 阶对称矩阵，则

- $\mathrm{E}(X'AX)=\mu'A\mu+\mathrm{tr}(A\Sigma)$
- $\mathrm{Cov}(X'AX,X'BX)=2\mathrm{tr}(A\Sigma B\Sigma)+4\mu'A\Sigma B\mu$

带入 $\Sigma=I_p$，$A=I_p$，$B=I_p$ 得

$$\begin{aligned}\mathrm{E}(y)&=\mathrm{E}(x'x)=\mu'\mu+\mathrm{tr}(I_p)=\lambda+p\\\mathrm{Var}(y)&=\mathrm{Var}(x'x)=2\mathrm{tr}(I_p)+4\mu'\mu=2p+4\lambda\end{aligned}$$

■

Solution. 思路 3：已知非中心 Γ 分布 $\Gamma(\alpha,\lambda,\delta)$ 的特征函数为

$$\varphi(t)=\left(1-\frac{it}{\lambda}\right)^{-\alpha}\exp\left\{\frac{it\delta/2}{\lambda-it}\right\}$$

当形状参数 $\alpha=n/2$，尺度参数 $\lambda=1/2$ 时非中心 Γ 分布为非中心 $\chi_n^2(\delta)$ 分布。 ■

第 8 次作业

1. 高惠璇老师书上的习题 3-11, 表 3.4 给出 15 名两周岁婴儿的身高 (X_1), 胸围 (X_2) 和上半臂围 (X_3) 的测量数据. 假设男婴的测量数据 $X_{(\alpha)}(\alpha = 1, \cdots, 6)$ 为来自总体 $N_3(\mu_1, \Sigma)$ 的随机样本; 女婴的测量数据 $Y_{(\alpha)}(\alpha = 1, \cdots, 9)$ 为来自总体 $N_3(\mu_2, \Sigma)$ 的随机样本. 试利用表 3.4 中的数据解决如下问题.

(1) 检验两个总体的均值向量是否相同, $H_0 : \mu_1 = \mu_2$ $(\alpha = 0.05)$;

(2) 对总体均值向量的差构造置信域;

(3) 对每个分量构造单独置信区间和联合置信区间，并比较它们的差异.

```r
# 输入样本
X1 <- c(78, 76, 92, 81, 81, 84, 80, 75, 78, 75, 79, 78, 75, 64, 80)
X2 <- c(60.6, 58.1, 63.2, 59, 60.8, 59.5, 58.4, 59.2, 60.3, 57.4, 59.5,
        58.1, 58, 55.5, 59.2)
X3 <- c(16.5, 12.5, 14.5, 14, 15.5, 14, 14, 15, 15, 13, 14, 14.5, 12.5,
        11, 12.5)
gender <- rep(c("male", "female"), c(6, 9))
baby.body <- data.frame(gender, X1, X2, X3)
baby.body
```

```
##    gender X1   X2   X3
## 1    male 78 60.6 16.5
## 2    male 76 58.1 12.5
## 3    male 92 63.2 14.5
## 4    male 81 59.0 14.0
## 5    male 81 60.8 15.5
## 6    male 84 59.5 14.0
## 7  female 80 58.4 14.0
## 8  female 75 59.2 15.0
## 9  female 78 60.3 15.0
## 10 female 75 57.4 13.0
## 11 female 79 59.5 14.0
```

```
## 12 female 78 58.1 14.5
## 13 female 75 58.0 12.5
## 14 female 64 55.5 11.0
## 15 female 80 59.2 12.5
```

Solution. (1) 两正态总体均值向量检验

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq 0$$

已知两正态总体的样本量分别为 m 和 n，V_1 和 V_2 分别为对应总体的样本离差阵，当两总体协方差阵 $\Sigma_1 = \Sigma_2 = \Sigma$ 未知时，检验统计量 T^2 如下，服从 Hotelling T^2 分布:

$$T^2 = \frac{mn(m+n-2)}{m+n}(\bar{x}-\bar{y})^T \left(V_1+V_2\right)^{-1} (\bar{x}-\bar{y}) \sim T^2(p, m+n-2)$$

其中

$$V_1 = \sum_{i=1}^{m} \left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T$$
$$V_2 = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)\left(y_i - \bar{y}\right)^T$$

将 Hotelling T^2 分布转化为 F 分布:

$$F = \frac{(m+n-p-1)T^2}{(m+n-2)p} = \frac{mn(m+n-p-1)}{(m+n)p}(\bar{x}-\bar{y})^T \left(V_1+V_2\right)^{-1} (\bar{x}-\bar{y}) \sim F_{p,m+n-p-1}$$

```r
diff_mutest_unknown <- function(data1, data2, alpha=0.05) {
  # H0: mu1 - mu2 = 0 when Sigma0 is unknown
  # This is a F testing

  #----------input----------#
  # data1 and data2: design matrix
  # mu1 and mu2: mu1 - mu2 = 0 for null hypothesis
  # alpha: the significant level, default value=0.05
  #----------output----------#
  # Reject.area: reject region
```

```r
  # p.value: p-value

  data1 <- as.matrix(data1) # 将数据框转化为矩阵
  data2 <- as.matrix(data2)
  n <- nrow(data1); m <- nrow(data2)
  p <- ncol(data1)
  X.bar <- colMeans(data1); Y.bar <- colMeans(data2)
  V1 <- (n-1)*cov(data1); V2 <- (m-1)*cov(data2)
  f.stat <- n*m*(n+m-p-1)/((n+m)*p)*t(X.bar-Y.bar)%*%solve(V1+V2)%*%(X.bar-Y.bar)
  low.q <- qf(1-alpha, p, n+m-p-1) # 求下侧分位点，上侧：lower.tail=FALSE
  reject <- matrix(c(f.stat, low.q), nrow=1) # 按行排
  rownames(reject) <- c("Reject") # 行名
  colnames(reject) <- c("Obs", paste0("> ", 1-alpha)) # 列名
  pval <- 1 - pf(f.stat, p, n+m-p-1)
  return(list(Reject.area=reject, p.value=pval))
}

male.body <- baby.body[baby.body$gender=="male",-1]
female.body <-baby.body[baby.body$gender=="female",-1]
diff_mutest_unknown(male.body, female.body, alpha=0.05)
```

```
## $Reject.area
##               Obs    > 0.95
## Reject 1.498179 3.587434
##
## $p.value
##           [,1]
## [1,] 0.2692616
```

```r
# Alternative using package ICSNP
library(ICSNP)
HotellingsT2(male.body, female.body, mu=rep(0,(ncol(baby.body)-1)))
```

```
##
##  Hotelling's two sample T2-test
##
## data:  male.body and female.body
## T.2 = 1.4982, df1 = 3, df2 = 11, p-value = 0.2693
## alternative hypothesis: true location difference is not equal to c(0,0,0)
```

结论: 从拒绝域来看，检验统计量没有落入拒绝域；而且 p-value < $\alpha = 0.05$，所以不拒绝原假设。我们认为两个总体的均值向量是相同的。

(2) 构造两总体均值向量差的置信域

$$F = \frac{(m+n-p-1)T^2}{(m+n-2)p} = \frac{mn(m+n-p-1)}{(m+n)p}(\bar{x}-\bar{y})^T\left(V_1+V_2\right)^{-1}(\bar{x}-\bar{y}) \sim F_{p,m+n-p-1}$$

则 $\mu_1 - \mu_2$ 的 $1-\alpha$ 置信域为

$$\begin{aligned} D &= \{\mu_1 - \mu_2 \in \mathbb{R}^p \mid F \le c_\alpha\} \\ &= \left\{\mu_1 - \mu_2 \in \mathbb{R}^p \mid \left[(\mu_1-\mu_2)-(\bar{x}-\bar{y})\right]^T\left(V_1+V_2\right)^{-1}\left[(\mu_1-\mu_2)-(\bar{x}-\bar{y})\right] \le \frac{(m+n)p}{mn(m+n-p-1)}c_\alpha\right\} \end{aligned}$$

其中 c_α 为 $F_{p,m+n-p-1}$ 分布的上侧 α 分位点，$P(F_{p,m+n-p-1} > c_\alpha) = \alpha$。

```r
X.bar <- colMeans(male.body); Y.bar <- colMeans(female.body)
X.bar - Y.bar
```

```
##  X1  X2  X3
## 6.0 1.8 1.0
```

```r
m <- nrow(male.body); n <- nrow(female.body)
p <- ncol(female.body); alpha <- 0.05
V1 <- (m-1)*cov(male.body)
V2 <- (n-1)*cov(female.body)
solve(V1+V2)
```

```
##                X1           X2            X3
```

```
## X1  0.008844188 -0.02841209  0.007911044
## X2 -0.028412087  0.14739623 -0.067973510
## X3  0.007911044 -0.06797351  0.081017053
```

```
(m+n)*p/(m*n*(m+n-p-1))*qf(1-alpha, p, m+n-p-1)
```

```
## [1] 0.2717753
```

结论: 置信域为

$$\{\mu_1-\mu_2\in\mathbb{R}^3 \mid \left[(\mu_1-\mu_2)-\begin{pmatrix}6.0\\1.8\\1.0\end{pmatrix}\right]^T\begin{bmatrix}0.0088 & -0.0284 & 0.0079\\ -0.0284 & 0.1474 & -0.0680\\ 0.0079 & -0.0680 & 0.0810\end{bmatrix}^{-1}\left[(\mu_1-\mu_2)-\begin{pmatrix}6.0\\1.8\\1.0\end{pmatrix}\right]\leq 0.2717753\right\}$$

(3) a. 构造联合置信区间

对于任意的 a，考虑均值向量 μ 的线性组合 $a^T\mu$ 的置信区间便能够得到想要的联合置信区间。若取 $a=e_i=(0,\cdots,1,\cdots,0)'$，我们便同时得到 $\mu_i(i=1,\cdots,p)$ 的置信度均为 $1-\alpha$ 的 $T^2=n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)\leq c^2$ 区间

$$\bar{x}_i-c\sqrt{\frac{s_{ii}}{n}}\leqslant\mu_i\leqslant\bar{x}_i+c\sqrt{\frac{s_{ii}}{n}}$$

其中 $c=\sqrt{\frac{(n-1)p}{(n-p)}F_\alpha}$，$s_{ii}$ 为样本协方差阵 S 的第 i 个对角元素.

b. 构造单独置信区间

单个分量的置信度为 $1-\alpha$ 的置信区间为

$$\bar{x}_i-t_{a/2}\sqrt{\frac{s_{ii}}{n}}\leq\mu_i\leq\bar{x}_i+t_{a/2}\sqrt{\frac{s_{ii}}{n}}$$

```
comb_CI <- function(data, alpha, a, seperate=FALSE) {
  data <- as.matrix(data)
  n <- nrow(data)
  p <- ncol(data)
  if (p!=length(a)) {
```

```r
    stop("the length of a is not matching to the column number of data")
}
  if(seperate == TRUE) {
  # 单独置信区间
  if (length(which(a==1))==1 & length(which(a==0))==p-1) {
    k <- qt(1-alpha/2,n-1)
  } else {
    stop("when seperate is TRUE, this must be only one 1 in vector a")
    }
  } else {
    k <- sqrt(((n-1)*p*qf(1-alpha,p,n-p))/(n-p))
  }
  S <- cov(data)
  Xbar <- as.matrix(colMeans(data))
  lowBound <- t(a) %*% Xbar - k * sqrt((t(a) %*% S %*% a)/n)
  upBound <- t(a) %*% Xbar + k * sqrt((t(a) %*% S %*% a)/n)
  interval <- c(lowBound, upBound)
  inter.len <- upBound - lowBound
  return(list(interval=interval, inter.len=inter.len))
}
```

比较单独置信区间和联合置信区间的区别

```r
compare_fun <- function(data, gender) {
  feature <- c(" 身高", " 胸围", " 上半臂围")
  for (i in 1:ncol(data)){
  a <- rep(0, 3); a[i] <- 1
  single.out <- comb_CI(data, 0.05, a, seperate=TRUE)
  single.CI <- single.out$interval
  single.len <- single.out$inter.len
  comb.out <- comb_CI(data, 0.05, a, seperate=FALSE)
  comb.CI <- comb.out$interval
```

```r
  comb.len <- comb.out$inter.len

  cat(" ", gender, feature[i], " 单独 95% 的置信区间为 (",
      single.CI[1], ",", single.CI[2],")",
      " 区间长度为", single.len,
      "\n ",
      gender, feature[i], " 联合 95% 的置信区间为 (",
      comb.CI[1], ",", comb.CI[2],")",
      " 区间长度为", comb.len)
      cat("\n")
  }
}
```

```r
compare_fun(male.body, " 男生")
```

```
##   男生 身高 单独95%的置信区间为( 76.10072 , 87.89928 ) 区间长度为 11.79857
##   男生 身高 联合95%的置信区间为( 66.3704 , 97.6296 ) 区间长度为 31.25921
##   男生 胸围 单独95%的置信区间为( 58.33094 , 62.06906 ) 区间长度为 3.738113
##   男生 胸围 联合95%的置信区间为( 55.24811 , 65.15189 ) 区间长度为 9.903781
##   男生 上半臂围 单独95%的置信区间为( 13.05345 , 15.94655 ) 区间长度为 2.893094
##   男生 上半臂围 联合95%的置信区间为( 10.66751 , 18.33249 ) 区间长度为 7.664984
```

```r
compare_fun(female.body, " 女生")
```

```
##   女生 身高 单独95%的置信区间为( 72.19529 , 79.80471 ) 区间长度为 7.609425
##   女生 身高 联合95%的置信区间为( 68.80284 , 83.19716 ) 区间长度为 14.39432
##   女生 胸围 单独95%的置信区间为( 57.32112 , 59.47888 ) 区间长度为 2.157754
##   女生 胸围 联合95%的置信区间为( 56.35915 , 60.44085 ) 区间长度为 4.081702
##   女生 上半臂围 单独95%的置信区间为( 12.46515 , 14.53485 ) 区间长度为 2.069702
##   女生 上半臂围 联合95%的置信区间为( 11.54243 , 15.45757 ) 区间长度为 3.915139
```

可以看出，男婴/女婴数据的 95% 联合置信区间长度比 95% 单独置信区间长度更长。对于每个特征，区间下限都更低，上限都更高。说明，单独置信区间是更加严格的。

第 9 次作业

考虑三个总体的 Bayes 的判别问题, 先验概率、误判损失以及密度函数在新的样本点处的值如下表, 要求

(1) 根据错判平均损失最小的规则, 应把 x_0 判给哪一个总体?

(2) 若所有错判损失相等, 此时 x_0 应判入哪个总体? 并计算其后验概率.

- 错判损失: $L(1/1) = 0$, $L(1/2) = 500$, $L(1/3) = 100$; $L(2/1) = 10$, $L(2/2) = 0$, $L(2/3) = 50$; $L(3/1) = 50$, $L(3/2) = 200$, $L(3/3) = 0$;
- 先验概率: $q_1 = 0.2$, $q_2 = 0.3$, $q_3 = 0.5$;
- 密度函数: $f_1(x_0) = 0.01$, $f_2(x_0) = 0.85$, $f_3(x_0) = 0.01$.

Solution. 基于错判平均损失最小的规则,

$$h_1(x_0) = \sum_{i=1}^{3} q_i L(1|i) f_i(x_0) = 0.3 \times 500 \times 0.85 + 0.5 \times 100 \times 0.01 = 128$$
$$h_2(x_0) = \sum_{i=1}^{3} q_i L(2|i) f_i(x_0) = 0.2 \times 10 \times 0.01 + 0.5 \times 50 \times 0.01 = 0.27$$
$$h_3(x_0) = \sum_{i=1}^{3} q_i L(3|i) f_i(x_0) = 0.2 \times 50 \times 0.01 + 0.3 \times 200 \times 0.85 = 51.1$$

$h_2(x_0) < h_3(x_0) < h_1(x_0)$, 因此 x_0 应该判给第二个总体 $G2$.

当所有错判损失相等时, Bayes 的判别问题等价于后验概率最大的判别, 后验概率分别为

$$P(G_1|x_0) = \frac{q_1 f_1(x_0)}{\sum_{i=1}^{3} q_i f_i(x_0)} = \frac{2}{262}$$
$$P(G_2|x_0) = \frac{q_2 f_2(x_0)}{\sum_{i=1}^{3} q_i f_i(x_0)} = \frac{255}{262}$$
$$P(G_3|x_0) = \frac{q_3 f_3(x_0)}{\sum_{i=1}^{3} q_i f_i(x_0)} = \frac{5}{262}$$

$P(G_2|x_0) > P(G_3|x_0) > P(G_1|x_0)$, 因此 x_0 应该判给第二个总体 $G2$. ■

第 10 次作业

(ESL, Ex. 4.2) Suppose we have features $x \in \mathbb{R}^p$, a two-class response, with class sizes N_1, N_2, and the target coded as $-N/N_1, N/N_2$.

(a) Show that the LDA rule classifies to class 2 if

$$x^T \hat{\mathbf{\Sigma}}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) > \frac{1}{2} (\hat{\mu}_2 + \hat{\mu}_1)^T \hat{\mathbf{\Sigma}}^{-1} (\hat{\mu}_2 - \hat{\mu}_1) - \log (N_2/N_1)$$

and class 1 otherwise.

(b) Consider minimization of the least squares criterion

$$\sum_{i=1}^{N} \left(y_i - \beta_0 - x_i^T \beta\right)^2$$

Show that the solution $\hat{\beta}$ satisfies

$$\left[(N-2)\hat{\mathbf{\Sigma}} + N\hat{\mathbf{\Sigma}}_B\right] \beta = N (\hat{\mu}_2 - \hat{\mu}_1)$$

(after simplification), where $\hat{\mathbf{\Sigma}}_B = \frac{N_1 N_2}{N^2} (\hat{\mu}_2 - \hat{\mu}_1) (\hat{\mu}_2 - \hat{\mu}_1)^T$.

(c) Hence show that $\hat{\mathbf{\Sigma}}_B \beta$ is in the direction $(\hat{\mu}_2 - \hat{\mu}_1)$ and thus

$$\hat{\beta} \propto \hat{\mathbf{\Sigma}}^{-1} (\hat{\mu}_2 - \hat{\mu}_1)$$

Therefore the least-squares regression coefficient is identical to the LDA coefficient, up to a scalar multiple.

Solution. Part (a): Under zero-one classification loss, for each class ω_k the Bayes' discriminant functions $\delta_k(x)$ take the following form

$$\delta_k(x) = \ln (p (x \mid \omega_k)) + \ln (\pi_k) \tag{1}$$

If our conditional density $p (x \mid \omega_k)$ is given by a multidimensional normal then its function form is given by

$$p (x \mid \omega_k) = \mathcal{N} (x; \mu_k, \Sigma_k) \equiv \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) \right\}$$

Taking the logarithm of this expression as required by Equation (1) we find

$$\ln (p (x \mid \omega_k)) = -\frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) - \frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln (|\Sigma_k|)$$

so that our discriminant function in the case when $p\left(x \mid \omega_k\right)$ is a multidimensional Gaussian is given by

$$\delta_k(x) = -\frac{1}{2}\left(x-\mu_k\right)^T \Sigma_k^{-1}\left(x-\mu_k\right) - \frac{p}{2}\ln(2\pi) - \frac{1}{2}\ln\left(|\Sigma_k|\right) + \ln\left(\pi_k\right). \quad (2)$$

We will now consider some specializations of this expression for various possible values of Σ_k and how these assumptions modify the expressions for $\delta_k(x)$. Since linear discriminant analysis (LDA) corresponds to the case of equal covariance matrices our decision boundaries (given by Equation (2)), but with equal covariances ($\Sigma_k = \Sigma$). For decision purposes we can drop the two terms $-\frac{p}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma|)$ and use a discriminant $\delta_k(x)$ given by

$$\delta_k(x) = -\frac{1}{2}\left(x-\mu_k\right)^T \Sigma^{-1}\left(x-\mu_k\right) + \ln\left(\pi_k\right)$$

Expanding the quadratic in the above expression we get

$$\delta_k(x) = -\frac{1}{2}\left(x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_k - \mu_k^T\Sigma^{-1}x + \mu_k^T\Sigma^{-1}\mu_k\right) + \ln\left(\pi_k\right)$$

Since $x^T\Sigma^{-1}x$ is a common term with the same value in all discriminant functions we can drop it and just consider the discriminant given by

$$\delta_k(x) = \frac{1}{2}x^T\Sigma^{-1}\mu_k + \frac{1}{2}\mu_k^T\Sigma^{-1}x - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \ln\left(\pi_k\right).$$

Since $x^T\Sigma^{-1}\mu_k$ is a scalar, its value is equal to the value of its transpose so

$$x^T\Sigma^{-1}\mu_k = \left(x^T\Sigma^{-1}\mu_k\right)^T = \mu_k^T\left(\Sigma^{-1}\right)^T x = \mu_k^T\Sigma^{-1}x$$

since Σ^{-1} is symmetric. Thus the two linear terms in the above combine and we are left with

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + \ln\left(\pi_k\right).$$

Next we can estimate π_k from data using $\pi_i = \frac{N_i}{N}$ for $i = 1, 2$ and we pick class 2 as the classification outcome if $\delta_2(x) > \delta_1(x)$ (and class 1 otherwise). This inequality can be written as

$$x^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \ln\left(\frac{N_2}{N}\right) > x^T\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \ln\left(\frac{N_1}{N}\right)$$

or moving all the x terms to one side

$$x^T\Sigma^{-1}\left(\mu_2-\mu_1\right) > \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \ln\left(\frac{N_1}{N}\right) - \ln\left(\frac{N_2}{N}\right)$$

as we were to show.

Part (b): To minimize the expression $\sum_{i=1}^{N}\left(y_i - \beta_0 - \beta^T x_i\right)^2$ over $(\beta_0, \beta)'$ we know that the solution $\left(\hat{\beta}_0, \hat{\beta}\right)'$ must satisfy the normal equations which in this case is given by

$$X^T X \left[\begin{array}{c} \beta_0 \\ \beta \end{array}\right] = X^T \mathbf{y}$$

Our normal equations have a block matrix $X^T X$ on the left-hand-side given by

$$\left[\begin{array}{cccccccc} 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_{N_1} & x_{N_1+1} & x_{N_1+2} & \cdots & x_{N_1+N_2} \end{array}\right] \left[\begin{array}{cc} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \\ 1 & x_{N_1}^T \\ 1 & x_{N_1+1}^T \\ 1 & x_{N_1+2}^T \\ \vdots & \\ 1 & x_{N_1+N_2}^T \end{array}\right].$$

When we take the product of these two matrices we find

$$\left[\begin{array}{cc} N & \sum_{i=1}^{N} x_i^T \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i x_i^T \end{array}\right] \tag{3}$$

For the case where we code our response as $-\frac{N}{N_1}$ for the first class and $+\frac{N}{N_2}$ for the second class (where $N = N_1 + N_2$), the right-hand-side or $X^T y$ of the normal equations becomes When we take the product of these two matrices we get

$$\left[\begin{array}{c} N_1\left(-\frac{N}{N_1}\right) + N_2\left(\frac{N}{N_2}\right) \\ \left(\sum_{i=1}^{N_1} x_i\right)\left(-\frac{N}{N_1}\right) + \left(\sum_{i=N_1+1}^{N} x_i\right)\left(\frac{N}{N_2}\right) \end{array}\right] = \left[\begin{array}{c} 0 \\ -N\mu_1 + N\mu_2 \end{array}\right]$$

Note that we can simplify the $(1, 2)$ and the $(2, 1)$ elements in the block coefficient matrix $X^T X$ in Equation (3) by introducing the class specific means (denoted by μ_1 and μ_2) as

$$\sum_{i=1}^{N} x_i = \sum_{i=1}^{N_1} x_i + \sum_{i=N_1+1}^{N} x_i = N_1\mu_1 + N_2\mu_2$$

Also if we pool all of the samples for this two class problem $(K = 2)$ together we can estimate the pooled covariance matrix $\hat{\Sigma}$ (see the section in the book on linear discriminant analysis) as

$$\hat{\Sigma} = \frac{1}{N-K}\sum_{k=1}^{K}\sum_{i:g_i=k}\left(x_i - \mu_k\right)\left(x_i - \mu_k\right)^T$$

When $K = 2$ this is

$$\begin{aligned}
\hat{\Sigma} &= \frac{1}{N-2}\left[\sum_{i:g_i=1}(x_i-\mu_1)(x_i-\mu_1)^T + \sum_{i:g_i=2}(x_i-\mu_2)(x_i-\mu_2)^T\right] \\
&= \frac{1}{N-2}\left[\sum_{i:g_i=1} x_i x_i^T - N_1\mu_1\mu_1^T + \sum_{i:g_i=1} x_i x_i^T - N_2\mu_2\mu_2^T\right]
\end{aligned}$$

From which we see that the sum $\sum_{i=1}^N x_i x_i^T$ found in the $(2,2)$ element in the matrix from Equation (3) can be written as

$$\sum_{i=1}^N x_i x_i^T = (N-2)\hat{\Sigma} + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T$$

Now that we have evaluated both sides of the normal equations we can write them down again as a linear system. We get

$$\left[\begin{array}{cc} N & N_1\mu_1^T + N_2\mu_2^T \\ N_1\mu_1 + N_2\mu_2 & (N-2)\hat{\Sigma} + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T \end{array}\right]\left[\begin{array}{c}\beta_0 \\ \beta\end{array}\right] = \left[\begin{array}{c} 0 \\ -N\mu_1 + N\mu_2\end{array}\right] \tag{4}$$

In more detail we can write out the first equation in the above system as

$$N\beta_0 + \left(N_1\mu_1^T + N_2\mu_2^T\right)\beta = 0$$

or solving for β_0, in terms of β, we get

$$\beta_0 = \left(-\frac{N_1}{N}\mu_1^T - \frac{N_2}{N}\mu_2^T\right)\beta$$

When we put this value of β_0 into the second equation in Equation (4) we find the total equation for β then looks like

$$(N_1\mu_1 + N_2\mu_2)\left(-\frac{N_1}{N}\mu_1^T - \frac{N_2}{N}\mu_2^T\right)\beta + \left((N-2)\hat{\Sigma} + N_1\mu_1\mu_1^T + N_2\mu_2\mu_2^T\right)\beta = N(\mu_2 - \mu_1)$$

Consider the terms that are outer products of the vectors μ_i (namely terms like $\mu_i\mu_j^T$) we see that taken together they look like

$$\begin{aligned}
\text{Outer Product Terms} &= -\frac{N_1^2}{N}\mu_1\mu_1^T - \frac{2N_1N_2}{N}\mu_1\mu_2^T - \frac{N_2^2}{N}\mu_2\mu_2^T + N_1\mu_1\mu_2^T + N_2\mu_2\mu_2^T \\
&= \left(-\frac{N_1^2}{N} + N_1\right)\mu_1\mu_1^T - \frac{2N_1N_2}{N}\mu_1\mu_2^T + \left(-\frac{N_2^2}{N} + N_2\right)\mu_2\mu_2^T \\
&= \frac{N_1}{N}(-N_1 + N)\mu_1\mu_1^T - \frac{2N_1N_2}{N}\mu_1\mu_2^T + \frac{N_2}{N}(-N_2 + N)\mu_2\mu_2^T \\
&= \frac{N_1N_2}{N}\mu_1\mu_1^T - \frac{2N_1N_2}{N}\mu_1\mu_2^T + \frac{N_2N_1}{N}\mu_2\mu_2^T \\
&= \frac{N_1N_2}{N}\left(\mu_1\mu_1^T - 2\mu_1\mu_2 - \mu_2\mu_2\right) = \frac{N_1N_2}{N}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T
\end{aligned}$$

Here we have used the fact that $N_1 + N_2 = N$. If we introduce the matrix $\hat{\Sigma}_B$ as

$$\hat{\Sigma}_B \equiv (\mu_2 - \mu_1)(\mu_2 - \mu_1)^T$$

we get that the equation for β looks like

$$\left[(N-2)\hat{\Sigma} + \frac{N_1 N_2}{N}\hat{\Sigma}_B\right]\beta = N(\mu_2 - \mu_1) \tag{5}$$

as we were to show.

Part (c): Note that $\hat{\Sigma}_B \beta$ is $(\mu_2 - \mu_1)(\mu_2 - \mu_1)^T \beta$, and the product $(\mu_2 - \mu_1)^T \beta$ is a scalar. Therefore the vector direction of $\hat{\Sigma}_B \beta$ is given by $\mu_2 - \mu_1$. Thus in Equation (5) as both the right-hand-side and the term $\frac{N_1 N_2}{N}\hat{\Sigma}_B$ are in the direction of $\mu_2 - \mu_1$ the solution β must be in the direction (i.e. proportional to) $\hat{\Sigma}^{-1}(\mu_2 - \mu_1)$. ■

第 11 次作业

1. 验证最短距离法的单调性;

2. 推导 Ward 法的距离迭代公式.

1. 记第 L 步的并类距离为 D_L，假设 $D_L = D_{pq}^{(L-1)}$，即合并 G_p, G_q 为新类 G_r，所以 $D_{pq}^{(L-1)} = \min D_{ij}^{(L-1)}$

新类 G_r 与其它类 G_k 的距离为

$$D_{rk}^{(L)} = \min\{D_{pk}^{(L-1)}, D_{qk}^{(L-1)}\} \geqslant D_{pq}^{(L-1)} = D_L \quad (k \neq p, q)$$

$\forall\, ij \neq r, p, q$，$D_{ij}^{(L)} = D_{ij}^{(L-1)} \geqslant D_{pq}^{(L-1)} = D_L$

因此第 $L+1$ 步的并类距离 $D_{L+1} = \min D_{ij}^{(L)} \geqslant D_L$

2、Ward法定义 $D_{pq}^2 = W_r - (W_p + W_q)$

假设将n个样本分为了k类，记为$G_1, G_2, \cdots, G_k$，

样本个数为n_t，$\overline{X}^{(t)}$表为G_t的重心，$X_{(i)}^{(t)}$表示G_t中的第i个样本，$t=1,\cdots,k$；$i=1,\cdots,n_t$

设G_p与G_q合并为G_r，则$n_p+n_q=n_r$，$\overline{X}^{(r)} = \frac{1}{n_r}(n_p\overline{X}^{(p)}+n_q\overline{X}^{(q)})$

$$W_r = \sum_{t=1}^{n_r}(X_{(t)}^{(r)} - \overline{X}^{(r)})^T(X_{(t)}^{(r)} - \overline{X}^{(r)}) \quad \leftarrow G_r\text{中的离差平方和}$$

$$= \sum_{t=1}^{n_p}(X_{(t)}^{(p)} - \overline{X}^{(r)})^T(X_{(t)}^{(p)} - \overline{X}^{(r)}) + \sum_{t=1}^{n_q}(X_{(t)}^{(q)} - \overline{X}^{(r)})^T(X_{(t)}^{(q)} - \overline{X}^{(r)})$$

$$= \sum_{t=1}^{n_p}(X_{(t)}^{(p)} - \overline{X}^{(p)} + \overline{X}^{(p)} - \overline{X}^{(r)})^T(X_{(t)}^{(p)} - \overline{X}^{(p)} + \overline{X}^{(p)} - \overline{X}^{(r)}) +$$

$$\sum_{t=1}^{n_q}(X_{(t)}^{(q)} - \overline{X}^{(q)} + \overline{X}^{(q)} - \overline{X}^{(r)})^T(X_{(t)}^{(q)} - \overline{X}^{(q)} + \overline{X}^{(q)} - \overline{X}^{(r)})$$

其中 $\sum_{t=1}^{n_p}(X_{(t)}^{(p)} - \overline{X}^{(p)})^T(\overline{X}^{(p)} - \overline{X}^{(r)})$

$$= \sum_{t=1}^{n_p}\left(X_{(t)}^{(p)T}X^{(p)} - X_{(t)}^{(p)T}\overline{X}^{(r)} - \overline{X}^{(p)T}\overline{X}^{(p)} + \overline{X}^{(p)T}\overline{X}^{(r)}\right)$$

$$= n_p\overline{X}^{(p)T}X^{(p)} - n_p\overline{X}^{(p)T}\overline{X}^{(r)} - n_p\overline{X}^{(p)T}\overline{X}^{(p)} + n_p\overline{X}^{(p)T}\overline{X}^{(r)}$$

$$= 0$$

同理可证其余交叉项，得

$$W_r = \|X^{(p)} - \overline{X}^{(p)} + \overline{X}^{(p)} - \overline{X}^{(r)}\|^2 + \|X^{(q)} - \overline{X}^{(q)} + \overline{X}^{(q)} - \overline{X}^{(r)}\|^2$$

$$= \underbrace{\|X^{(p)} - \overline{X}^{(p)}\|^2}_{W_p} + \|\overline{X}^{(p)} - \overline{X}^{(r)}\|^2 + \underbrace{\|X^{(q)} - \overline{X}^{(q)}\|^2}_{W_q} + \|\overline{X}^{(q)} - \overline{X}^{(r)}\|^2$$

代入 $\overline{X}^{(r)} = \frac{1}{n_r}(n_p\overline{X}^{(p)}+n_q\overline{X}^{(q)})$ 得 $\overline{X}^{(p)} - \overline{X}^{(r)} = \frac{n_q}{n_r}(\overline{X}^{(p)} - \overline{X}^{(q)})$

$\overline{X}^{(q)} - \overline{X}^{(r)} = \frac{n_p}{n_r}(\overline{X}^{(q)} - \overline{X}^{(p)})$

$$W_r = W_p + W_q + \left(\frac{n_q^2}{n_r^2}\cdot n_p + \frac{n_p^2}{n_r^2}\cdot n_q\right)(\overline{X}^{(p)} - \overline{X}^{(q)})^T(\overline{X}^{(p)} - \overline{X}^{(q)})$$

$$= W_p + W_q + \frac{n_p n_q}{n_p+n_q}(\overline{X}^{(p)} - \overline{X}^{(q)})^T(\overline{X}^{(p)} - \overline{X}^{(q)}) \Rightarrow D_{pq}^2 = \frac{n_p n_q}{n_p+n_q}(\overline{X}^{(p)} - \overline{X}^{(q)})^T(\overline{X}^{(p)} - \overline{X}^{(q)})$$

第 12 次作业

1. (高惠璇 7-3) 设 p 元总体 X 的协方差阵为

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} \quad (0 < \rho \leqslant 1).$$

   (1) 试证明总体的第一主成分 $Z_1 = \frac{1}{\sqrt{p}}(X_1 + X_2 + \cdots + X_p)$; (2) 试求第一主成分的贡献率.

2. (高惠璇 7-6) 设三元总体 X 的协方差阵为 $\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & \rho\sigma^2 & \sigma^2 \end{bmatrix}$, 试求总体主成分, 并计算每个主成分解释的方差比例 ($|\rho| \leqslant 1/\sqrt{2}$).

7-3 (1) 设$A=\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$，$0<\rho\le 1$，求A的特征值和特征向量

$$|\lambda E-A|=\begin{vmatrix} \lambda-1 & -\rho & \cdots & -\rho \\ -\rho & \lambda-1 & \cdots & -\rho \\ \vdots & \vdots & \ddots & \vdots \\ -\rho & -\rho & \cdots & \lambda-1 \end{vmatrix}=[\lambda-1-\rho(p-1)]\cdot(\lambda-1+\rho)^{p-1}=0$$

$$\downarrow$$

$$\begin{vmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{vmatrix}_{p\times p}=\begin{vmatrix} a+(p-1)b & a+(p-1)b & \cdots & a+(p-1)b \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{vmatrix}$$

$$=(a+(p-1)b)\begin{vmatrix} 1 & 1 & \cdots & 1 \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{vmatrix}$$

$$=(a+(p-1)b)\begin{vmatrix} 1 & 1 & \cdots & 1 \\ b & a-b & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b & 0 & \cdots & a-b \end{vmatrix}=(a+(p-1)b)(a-b)^{p-1}$$

则特征根为 $\lambda_1=1+\rho(p-1)$，$\lambda_2=1-\rho$ ($p-1$重根)，

因为 $0<\rho\le 1$，所以 $\lambda_1=1+\rho(p-1)>1-\rho$ 为最大特征根，对应的特征向量为

$$\lambda_1 E-A\rightarrow\begin{pmatrix} 1 & 1-p & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1-p \\ 0 & 0 & \cdots & 0 \end{pmatrix}\rightarrow\begin{pmatrix} 0 & -p & \cdots & p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1-p \\ 0 & 0 & \cdots & 0 \end{pmatrix}\Rightarrow\begin{cases} -px_2+px_p=0 \\ -px_3+px_p=0 \\ \vdots \\ x_1+x_2+\cdots+(1-p)x_p=0 \end{cases}$$

$$\Rightarrow \eta_1=\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\ \text{单位化后得}\ \eta_1=\begin{pmatrix} \frac{1}{\sqrt{p}} \\ \vdots \\ \frac{1}{\sqrt{p}} \end{pmatrix}$$

因此矩阵 $\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$ 对应的第一大特征根为 $\lambda_1 = \sigma^2 + \sigma^2\rho(p-1)$，

对应的特征向量为 $a_1 = \begin{pmatrix} \frac{1}{\sqrt{p}} \\ \vdots \\ \frac{1}{\sqrt{p}} \end{pmatrix}$

第一主成分 $Z_1 = a_1^T x = \frac{1}{\sqrt{p}}(x_1 + x_2 + \cdots + x_p)$

(2) Z_1 的贡献率为 $\dfrac{\sigma^2 + \sigma^2\rho(p-1)}{\sigma^2 + \sigma^2\rho(p-1) + \sigma^2(1-\rho)(p-1)} = \dfrac{1+\rho(p-1)}{p}$

7-6 (1) 设 $A = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}$，求A的特征值和特征向量

$$|\lambda E - A| = \begin{vmatrix} \lambda-1 & -\rho & 0 \\ -\rho & \lambda-1 & -\rho \\ 0 & -\rho & \lambda-1 \end{vmatrix} = (\lambda-1)[(\lambda-1)^2 - \rho^2] - \rho^2(\lambda-1)$$
$$= (\lambda-1)[(\lambda-1)^2 - 2\rho^2]$$

得 $\lambda_1 = 1$，$\lambda_2 = 1+\sqrt{2}\rho$，$\lambda_3 = 1-\sqrt{2}\rho$，从大到小排序得 $(|\rho| \le \frac{1}{\sqrt{2}})$

$\lambda_1 = 1+\sqrt{2}\rho$，$\lambda_2 = 1$，$\lambda_3 = 1-\sqrt{2}\rho$，对应的特征向量为

$$\lambda_1 E - A = \begin{pmatrix} \sqrt{2}\rho & -\rho & 0 \\ -\rho & \sqrt{2}\rho & -\rho \\ 0 & -\rho & \sqrt{2}\rho \end{pmatrix} \rightarrow \begin{pmatrix} \sqrt{2} & -1 & 0 \\ -1 & \sqrt{2} & -1 \\ 0 & -1 & \sqrt{2} \end{pmatrix} \quad \begin{matrix} \sqrt{2}x_1 - x_2 = 0 \\ -x_1 + \sqrt{2}x_2 - x_3 = 0 \\ -x_2 + \sqrt{2}x_3 = 0 \end{matrix}$$

$a_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{pmatrix}$，同理得 $a_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$，$a_3 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{\sqrt{2}} \\ \frac{1}{2} \end{pmatrix}$

可得总体主成分以及对应的解释方差比例为

$$Z_1 = \frac{x_1 + \sqrt{2}x_2 + x_3}{2} \rightarrow Var(Z_1) = \lambda_1 = \sigma^2(1+\sqrt{2}\rho) \rightarrow \frac{1+\sqrt{2}\rho}{3}$$

$$Z_2 = \frac{\sqrt{2}}{2}(x_1 - x_3) \rightarrow Var(Z_2) = \lambda_2 = \sigma^2 \rightarrow \frac{1}{3}$$

$$Z_3 = \frac{x_1 - \sqrt{2}x_2 + x_3}{2} \rightarrow Var(Z_3) = \lambda_3 = \sigma^2(1-\sqrt{2}\rho) \rightarrow \frac{1-\sqrt{2}\rho}{3}$$

第 13 次作业

1. 主分量法的因子得分为

$$\hat{F} = (A'A)^{-1} A'x = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} l_1' x \\ \vdots \\ \frac{1}{\sqrt{\lambda_m}} l_m' x \end{pmatrix}$$

2. (高惠璇 8-2) 设标准化变量 X_1, X_2, X_3 的协方差阵 (即相关阵) 为

$$R = \begin{bmatrix} 1.00 & 0.63 & 0.45 \\ 0.63 & 1.00 & 0.35 \\ 0.45 & 0.35 & 1.00 \end{bmatrix},$$

R 的特征向量为

$$\begin{aligned} \lambda_1 &= 1.9633, \quad l_1 = (0.6250, 0.5932, 0.5075)', \\ \lambda_2 &= 0.6795, \quad l_2 = (-0.2186, -0.4911, 0.8432)', \\ \lambda_3 &= 0.3672, \quad l_3 = (0.7494, -0.6379, -0.1772)'. \end{aligned}$$

(1) 取公共因子个数 $m = 1$ 时, 求因子模型的主成分解, 并计算误差平方和 $Q(1)$; (2) 取公共因子个数 $m = 2$ 时, 求因子模型的主成分解, 并计算误差平方和 $Q(2)$; (3) 试求误差平方和 $Q(m) < 0.1$ 的主成分解.

Proof. 1. 当使用主分量法估计因子载荷矩阵时，$A = (\sqrt{\lambda_1} l_1, \cdots, \sqrt{\lambda_m} l_m)$，此时极小化

$$\sum_{i=1}^{p} \varepsilon_i^2 = \varepsilon^T \varepsilon = (x - AF)^\top (x - AF) = \varphi(F)$$

使得 $\varphi(\hat{F}) = \min \varphi(F)$，即 $\hat{F}$ 满足 $\frac{\partial \varphi(F)}{\partial F} = 0$

$$\begin{aligned} &\frac{\partial \left(x^T x - x^T AF - F^T A^T x + F^T A^T AF\right)}{\partial F} = -2A^T x + 2A^T AF = 0 \\ \Rightarrow &\hat{F} = \left(A^T A\right)^{-1} A^T x = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} l_1' x \\ \vdots \\ \frac{1}{\sqrt{\lambda_m}} l_m' x \end{pmatrix} \end{aligned}$$

■

Solution. 2. (1) $m = 1$ 时,

$$A = \left(\sqrt{\lambda_1} \ell_1\right) = \sqrt{1.9633} \begin{pmatrix} 0.6250 \\ 0.5932 \\ 0.5075 \end{pmatrix} = \begin{pmatrix} 0.8757 \\ 0.8312 \\ 0.7111 \end{pmatrix}$$

$$D=\begin{pmatrix} 1.00-0.8757^2 & 0 & 0 \\ 0 & 1.00 & -0.8312^2 \\ 0 & 0 & 0 \end{pmatrix}=\begin{pmatrix} 0.2331 & 0 & 0 \\ 0 & 0.3091 & 0 \\ 0 & 0 & 0.4943 \end{pmatrix}$$

$$\begin{aligned} R-\left(AA^T+D\right) &= \begin{pmatrix} 1.00 & 0.63 & 0.45 \\ 0.63 & 1.00 & 0.35 \\ 0.45 & 0.35 & 1.00 \end{pmatrix} - \begin{pmatrix} 1.00 & 0.7279 & 0.6227 \\ 0.7279 & 1.00 & 0.5911 \\ 0.6227 & 0.5911 & 1.00 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -0.0979 & -0.1727 \\ -0.0979 & 0 & -0.2411 \\ -0.1727 & -0.2411 & 0 \end{pmatrix} \end{aligned}$$

$$Q(1)=\sum_{i=1}^{3}\sum_{j=1}^{3}\varepsilon_{ij}^2=2\times\left(0.0979^2+0.1727^2+0.2411^2\right)=0.1951$$

(2) $m=2$ 时,

$$A=\left(\sqrt{\lambda_1}l_1,\sqrt{\lambda_2}l_2\right)=\begin{pmatrix} 0.8757 & -0.1802 \\ 0.8312 & -0.4048 \\ 0.7111 & 0.6950 \end{pmatrix}$$

$$\begin{aligned} D &= \begin{pmatrix} 1.00-0.8757^2-0.1802^2 & 0 & 0 \\ 0 & 1.00-0.8312^2-0.4048^2 & 0 \\ 0 & 0 & 1.00-0.711^2-0.6950^2 \end{pmatrix} \\ &= \begin{pmatrix} 0.2007 & 0 & 0 \\ 0 & 0.1452 & 0 \\ 0 & 0 & 0.01131 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} R-\left(AA^\top+D\right) &= \begin{pmatrix} 1.00 & 0.63 & 0.45 \\ 0.63 & 1.00 & 0.35 \\ 0.45 & 0.35 & 1.00 \end{pmatrix} - \begin{pmatrix} 1.00 & 0.8008 & 0.4975 \\ 0.8008 & 1.00 & 0.3097 \\ 0.4975 & 0.3097 & 1.00 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -0.1708 & -0.0475 \\ -0.1708 & 0 & 0.0403 \\ -0.0475 & 0.0403 & 0 \end{pmatrix} \end{aligned}$$

$$Q(2)=\sum_{i=1}^{3}\sum_{j=1}^{3}\varepsilon_{ij}^2=2\times\left(0.1708^2+0.0475^2+0.0403^2\right)=0.06611$$

(3) $m=2$ 时, $Q(2)=0.06611<0.1$，$m=2$ 的主成分解满足条件。 ■