《多元统计分析》
课后作业参考答案

任课教师：肖磊

中国石油大学（北京）克拉玛依校区文理学院数学与统计系

Thursday 21st May, 2026

目录

Chapter 1

第 1 周作业参考答案

第 1 周作业截止时间： 2025 年 3 月 10 日 24:00.

1. [2 分] 最大值一定是异常值吗?

【解】 不一定. 例如下述代码生成的数据集 x，其最大值、最小值都不是异常值. 但是给数据集 x 添加一个数 10 之后的数据集 y，其最大值则为异常值. 结果如图1.1所示.

```
rm(list = ls(all = TRUE))
graphics.off()
set.seed(2022) # 产生随机数的种子
x = rnorm(20, 0, 1) # 生成标准正态分布的容量为20的随机数(样本)
options(digits = 3) # 保留三位有效数字
x
summary(x) # x 的五数总括 + 均值
y = c(x, 10)
y
summary(y) # y 的五数总括 + 均值
par(mfrow = c(1,  2))
boxplot(x, xlab = "Boxplot of data set x") # x 的箱线图
boxplot(y, xlab = "Boxplot of data set y") # y 的箱线图
```

2. [2 分] 均值或中位数是否有可能位于四分位数之外?

【解】 均值有可能位于两个四分位数之间，也有可能位于四分位数之外. 例如，上述数据集 x 的均值就位于两个四分位数之间，但是给数据集 x 添加一个数 50 之后的数据集 z，其均值则位于四分位数之外. 结果如图1.2所示.

由中位数的定义可知，中位数则总是位于两个四分位数之间.

```
graphics.off()
```

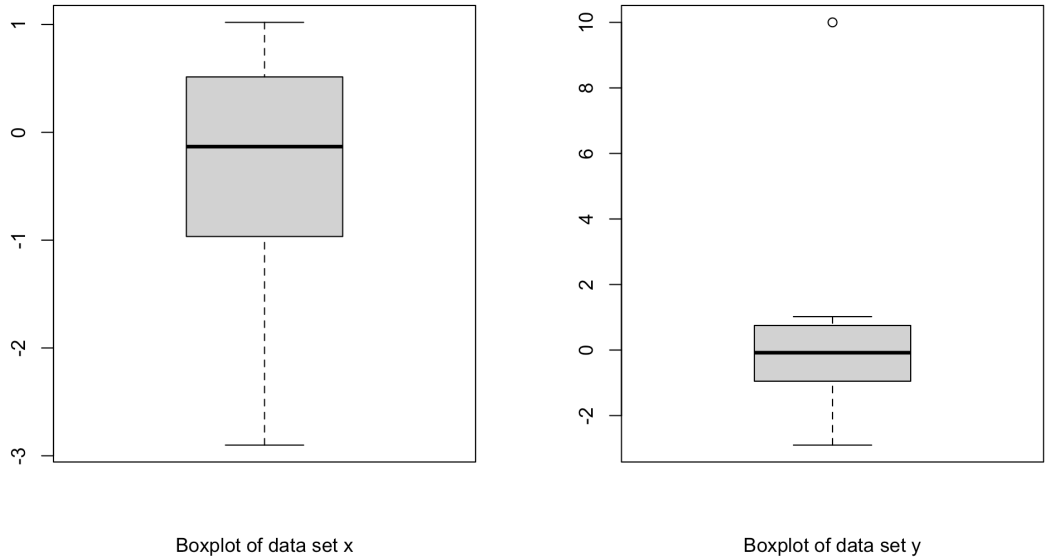

图 1.1: 最大值不一定是异常值.

```
z = c(x, 50)
par(mfrow = c(1,  2)) # 绘图区域设置为一行两列
boxplot(x, xlab = "Boxplot of data set x") # x 的箱线图
abline(h = mean(x), col = "red")
boxplot(z, xlab = "Boxplot of data set z") # z 的箱线图
abline(h = mean(z), col = "red")
```

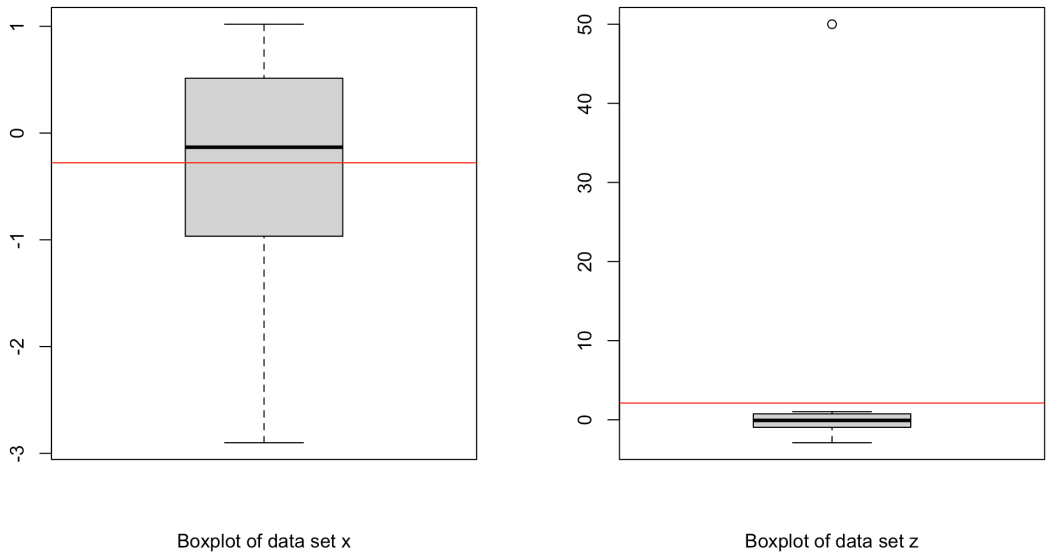

图 1.2: 最大值有可能位于四分位数之外.

3. [2 分] 假设数据来自标准正态分布 $N(0,1)$. 你预计会有百分之多少的数据可能是异常值呢?

【解】 异常值的定义: 小于"下四分位数 $-1.5\times$ 四分位差"的数据, 或者大于"上四分位数 $+1.5\times$ 四分位差"的数据. 如果数据来自正态分布 $N(0,1)$, 则会有 0.698% 的数据可能是异常值, 计算这一比例的 R 代码如下:

```
rm(list = ls(all = TRUE))
LowerQ = qnorm(0.25, 0, 1, lower.tail = TRUE) # N(0, 1)的下四分位数
UpperQ = qnorm(0.75, 0, 1, lower.tail = TRUE) # N(0, 1)的上四分位数
a = UpperQ - LowerQ # 四分位差
pnorm(LowerQ - 1.5 * a, 0, 1, lower.tail = TRUE) +
    pnorm(UpperQ + 1.5 * a, 0, 1, lower.tail = FALSE)
```

4. 关于五数总括中的五个数字.

(a) [2 分] 有没有可能五个数字全部相等呢?

【解】 有可能.

(b) [2 分] 如果可能的话, 会在什么情况下发生呢?

【解】 如果数据集当中所有的数据都相同时, 则五数总括中的五个数字全部相等. 验证这一现象的 R 代码如下:

```
rm(list = ls(all = TRUE))
x = rep(3.14, 30)
summary(x)
```

5. 对于瑞银纸币的对角线变量而言.

(a) [2 分] 使用带宽选择准则来计算对角线变量的最优选定带宽 h 并作核密度估计的图形.

【解】 选用 Gauss 核函数, 带宽的最优值为

$$h_G = 1.06\, n^{-1/5}\, \widehat{\sigma} = 1.06\, n^{-1/5} \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{1.1}$$

计算得最优带宽值为 $h_G \approx 0.138$, 相应的核密度估计图形如图1.3 (左图) 所示.

如果选用二次核函数, 则带宽的最优值为

$$h_Q = 2.62 \times h_G \tag{1.2}$$

计算得最优带宽值为 $h_G \approx 0.362$, 相应的核密度估计图形如图1.3 (右图) 所示. 计算及绘图的 R 代码如下:

```
rm(list = ls(all = TRUE))
graphics.off()
library(mclust)
data(banknote)
x = banknote$Length
n = dim(banknote)[1]
s = sqrt(sum((x - mean(x))^2) / n)
par(mfrow = c(1, 2))
## Gauss 核函数
(h = 1.06 * s * (n^(-1/5)))
plot(density(x, bw = h, kernel = "gaussian"), xlab='Bandwidth = 0.138',
     main = 'Gaussian Kernel', ylab = 'Diagonal',
     axes = FALSE, lwd = 2, col = "red")
axis(1, pos = 0); axis(2)
## 二次核函数
(hQ = 2.62 * h)
plot(density(x, bw = hQ, kernel = "biweight"), xlab = 'Bandwidth = 0.362',
     main = 'Quartic Kernel', ylab = 'Diagonal',
     axes = FALSE, lwd = 2, col = "red", ylim = c(0, 1.0))
axis(1, pos = 0); axis(2)
```

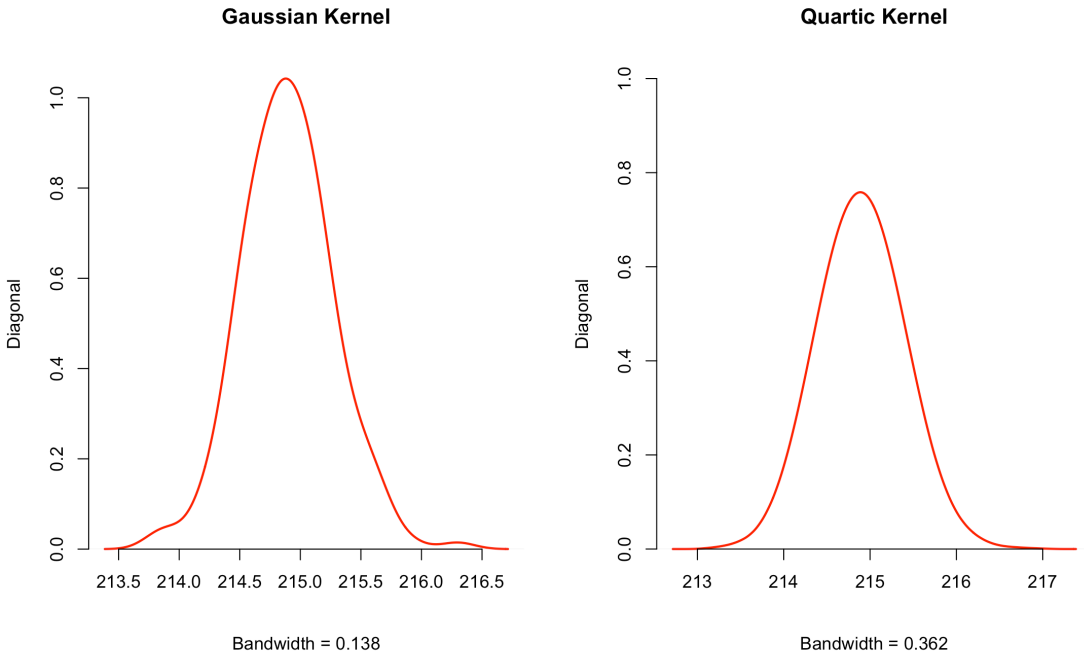

图 1.3: Gauss 核函数、二次核函数的核密度估计曲线.

(b) [2 分] 为这两组 (真钞、假钞) 数据设置同一个带宽会更好吗?

【解】 选用 Gauss 核函数，利用两组 (真钞、假钞) 数据计算的核密度估计的带宽值分别为 0.188 (真钞) 与 0.234 (伪钞)，其结果有显著差异. 相应的核密度估计图形如图1.4 所示.

选用二次核函数，利用两组 (真钞、假钞) 数据计算的核密度估计的带宽值分别为 0.492 (真钞) 与 0.614 (伪钞)，其结果亦有显著差异. 相应的核密度估计图形如图1.5 所示.

结论：为这两组 (真钞、假钞) 数据设置同一个带宽不会更好. 计算及绘图的 R 代码如下：

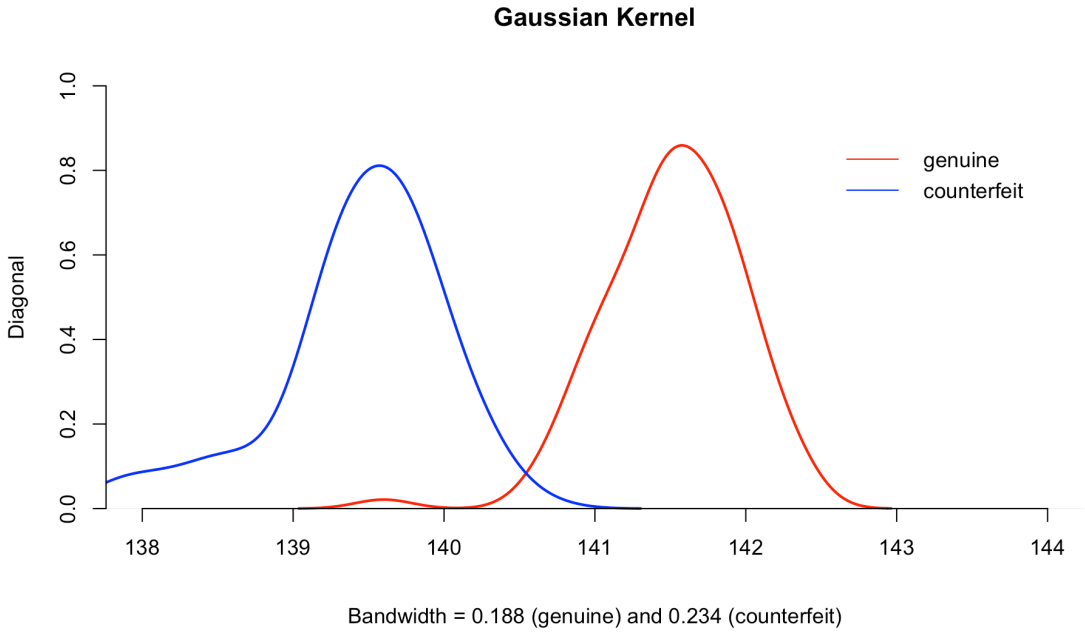

图 1.4: 两组数据 (真钞与伪钞) 的 Gauss 核函数的核密度估计曲线.

```
## 分类 (真钞、伪钞) 数据的核密度估计:Gauss 核函数
x = subset(banknote$Diagonal, banknote$Status == "genuine")
y = subset(banknote$Diagonal, banknote$Status == "counterfeit")
n = length(x)
sx = sqrt(sum((x - mean(x))^2) / n)
(hx = 1.06 * sx * (n^(-1/5)))
m = length(y)
sy = sqrt(sum((y - mean(y))^2) / m)
(hy = 1.06 * sy * (m^(-1/5)))
plot(density(x, bw = hx, kernel = "gaussian"),
     xlab='Bandwidth = 0.188 (genuine) and 0.234 (counterfeit)',
     main = 'Gaussian Kernel', ylab = 'Diagonal', xlim = c(138, 144),
     ylim = c(0, 1), axes = FALSE, lwd = 2, col = "red")
lines(density(y, bw = hy, kernel = "gaussian"), col = "blue", lwd = 2)
legend(142.5, 0.9, legend = c("genuine", "counterfeit"),
       col = c("red", "blue"), lty = 1, bty = "n")
```

```
axis(1, pos = 0); axis(2)
## 分类(真钞、伪钞)数据的核密度估计: 二次核函数
(hQx = 2.62 * hx)
(hQy = 2.62 * hy)
plot(density(x, bw = hQx, kernel = "biweight"),
     xlab='Bandwidth = 0.492 (genuine) and 0.614 (counterfeit)',
     main = 'Quartic Kernel', ylab = 'Diagonal', xlim = c(138, 144),
     ylim = c(0, 0.8), axes = FALSE, lwd = 2, col = "red")
lines(density(y, bw = hQy, kernel = "biweight"), col = "blue", lwd = 2)
legend(142.5, 0.7, legend = c("genuine", "counterfeit"),
       col = c("red", "blue"), lty = 1, bty = "n")
axis(1, pos = 0); axis(2)
```

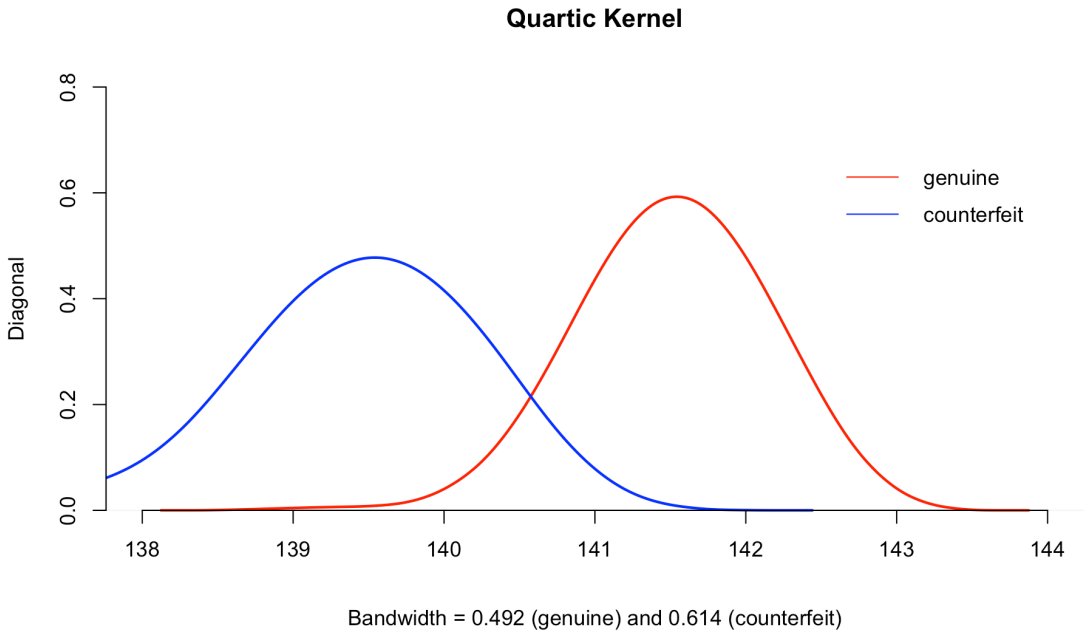

图 1.5: 两组数据 (真钞与伪钞) 的二次核函数的核密度估计曲线.

6. [2 分] 设 $|\mathcal{A}| = 0$. 问矩阵 $\mathcal{A}$ 的所有特征值都有可能是正数吗?

【解】 由《高等代数》的知识: 对于 n 阶方阵 $\mathcal{A}$, 其行列式 $|\mathcal{A}|$ 等于它的所有特征值 $\lambda_1, \lambda_2, \ldots, \lambda_n$ 的乘积, 即

$$|\mathcal{A}| = \lambda_1 \lambda_2 \cdots \lambda_n . \tag{1.3}$$

于是, 当 $|\mathcal{A}| = 0$ 时, 矩阵 $\mathcal{A}$ 的特征值至少有一个等于 0, 故其所有特征值不可能都是正数.

7. [2 分] 设矩阵 $\mathcal{A}$ (方阵) 的所有特征值都不为零. 问矩阵 $\mathcal{A}$ 是否一定可逆?

【解】 如果 n 阶方阵 $\mathcal{A}$ 的所有特征值 $\lambda_1, \lambda_2, \ldots, \lambda_n$ 都不为零, 则由 (1.3) 式知 $|\mathcal{A}| \neq 0$, 因此矩阵

$\mathcal{A}$ 一定可逆.

8. 设有矩阵 $\mathcal{A}$ 如下:

$$\mathcal{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix} \tag{1.4}$$

(a) [2 分] 利用 R 计算矩阵 $\mathcal{A}$ 的行列式 $|\mathcal{A}|$.

【解】 $\mathcal{A}$ 的行列式 $|\mathcal{A}| = 8$. 计算的 R 代码如下:

```
rm(list = ls(all = TRUE))
x = c(1, 2, 3, 2, 1, 2, 3, 2, 1)
A = matrix(x, nrow = 3, byrow = TRUE)
det(A) # 行列式值
```

(b) [2 分] 利用 R 求矩阵 $\mathcal{A}$ 的特征值与特征向量.

【解】 矩阵 $\mathcal{A}$ 的特征值为:

$$\lambda_1 = 5.702\,, \quad \lambda_2 = -0.702\,, \quad \lambda_3 = -2.000$$

对应的特征向量为:

$$\gamma_1 = \begin{pmatrix} -0.606 \\ -0.515 \\ -0.606 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 0.365 \\ -0.857 \\ 0.365 \end{pmatrix}, \quad \gamma_3 = \begin{pmatrix} 0.707 \\ 0.000 \\ -0.707 \end{pmatrix}.$$

计算的 R 代码如下:

```
eigen_A = eigen(A) # 矩阵 A 的谱分解
round(eigen_A$values, digits = 3) # A 的特征值
round(eigen_A$vectors, digits = 3) # 对应的特征向量
```

(c) [2 分] 利用 R 验证矩阵 $\mathcal{A}$ 的 Jordan 分解 (定理 2.1).

【解】 矩阵 $\mathcal{A}$ 的特征值构成的对角矩阵为

$$\Lambda = \begin{pmatrix} 5.702 & & \\ & -0.702 & \\ & & -2.000 \end{pmatrix}$$

对应的特征向量构成的矩阵为

$$\Gamma = \begin{pmatrix} -0.606 & 0.365 & 0.707 \\ -0.515 & -0.857 & 0.000 \\ -0.606 & 0.365 & -0.707 \end{pmatrix}$$

计算可得

$$\Gamma\Lambda\Gamma^{\mathrm{T}} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix} = \mathcal{A}$$

这就验证了定理 2.1 的结论. 计算的 R 代码如下:

```
Lambda = diag(eigen_A$values) # 特征值构成的对角矩阵
Gamma = eigen_A$vectors # 特征向量矩阵
Gamma %*% Lambda %*% t(Gamma) # Jordan 分解的验证
A # 矩阵 A
```

Chapter 2

第 2 周作业参考答案

1. 设 $\boldsymbol{a}$ 是一个 $(p\times 1)$ 向量，$\mathcal{A}=\mathcal{A}^{\mathrm{T}}$ 是一个对称的 $(p\times p)$ 矩阵.

(a) [2 分] 证明

$$\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}}{\partial \boldsymbol{x}}=\boldsymbol{a} \tag{2.1}$$

【证明】 对于函数 $f:\ \mathbb{R}^p\to\mathbb{R}$ 以及 $\boldsymbol{x}\in\mathbb{R}^p$，我们已经定义了

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\triangleq\begin{pmatrix}\dfrac{\partial f(\boldsymbol{x})}{\partial x_1}\\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_2}\\ \vdots\\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_p}\end{pmatrix},\quad \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}}\triangleq\left(\frac{\partial f(\boldsymbol{x})}{\partial x_1},\ \frac{\partial f(\boldsymbol{x})}{\partial x_2},\ \cdots,\ \frac{\partial f(\boldsymbol{x})}{\partial x_p}\right) \tag{2.2}$$

因此，对于

$$\boldsymbol{a}=\begin{pmatrix}a_1\\ a_2\\ \vdots\\ a_p\end{pmatrix}\in\mathbb{R}^p$$

因为 $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}=a_1x_1+a_2x_2+\cdots+a_px_p$，由上述定义 (2.2) 我们有

$$\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}}=\begin{pmatrix}\dfrac{\partial\left(a_1x_1+a_2x_2+\cdots+a_px_p\right)}{\partial x_1}\\ \dfrac{\partial\left(a_1x_1+a_2x_2+\cdots+a_px_p\right)}{\partial x_2}\\ \vdots\\ \dfrac{\partial\left(a_1x_1+a_2x_2+\cdots+a_px_p\right)}{\partial x_p}\end{pmatrix}=\begin{pmatrix}a_1\\ a_2\\ \vdots\\ a_p\end{pmatrix}=\boldsymbol{a}$$

同理，因为 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}=a_1x_1+a_2x_2+\cdots+a_px_p$，由上述定义 (2.2) 我们有

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}}{\partial \boldsymbol{x}}=\begin{pmatrix}\dfrac{\partial\left(a_1x_1+a_2x_2+\cdots+a_px_p\right)}{\partial x_1}\\ \dfrac{\partial\left(a_1x_1+a_2x_2+\cdots+a_px_p\right)}{\partial x_2}\\ \vdots\\ \dfrac{\partial\left(a_1x_1+a_2x_2+\cdots+a_px_p\right)}{\partial x_p}\end{pmatrix}=\begin{pmatrix}a_1\\ a_2\\ \vdots\\ a_p\end{pmatrix}=\boldsymbol{a}$$

(b) [2 分] 证明

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=2\mathcal{A}\boldsymbol{x} \tag{2.3}$$

【证明】 设

$$\mathcal{A}=\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1p}\\ a_{21}&a_{22}&\cdots&a_{2p}\\ \vdots&\vdots&&\vdots\\ a_{p1}&a_{p2}&\cdots&a_{pp}\end{pmatrix}$$

则

$$\begin{aligned}\boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}=&a_{11}x_1^2+a_{12}x_1x_2+\cdots+a_{1p}x_1x_p+\\ &a_{21}x_2x_1+a_{22}x_2^2+\cdots+a_{2p}x_2x_p+\\ &\cdots\cdots+\\ &a_{p1}x_px_1+a_{p2}x_px_2+\cdots+a_{pp}x_p^2\triangleq f\end{aligned}$$

因此，我们有

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\begin{pmatrix}\dfrac{\partial f}{\partial x_1}\\ \dfrac{\partial f}{\partial x_2}\\ \vdots\\ \dfrac{\partial f}{\partial x_p}\end{pmatrix}=\begin{pmatrix}2a_{11}x_1+\left(a_{12}+a_{21}\right)x_2+\cdots+\left(a_{1p}+a_{p1}\right)x_p\\ \left(a_{12}+a_{21}\right)x_1+2a_{22}x_2+\cdots+\left(a_{2p}+a_{p2}\right)x_p\\ \vdots\\ \left(a_{1p}+a_{p1}\right)x_1+\left(a_{2p}+a_{p2}\right)x_2+\cdots+2a_{pp}x_p\end{pmatrix}$$

因为 $\mathcal{A}=\mathcal{A}^{\mathrm{T}}$，所以 $a_{ij}=a_{ji}$，于是

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\begin{pmatrix}2a_{11}x_1+2a_{12}x_2+\cdots+2a_{1p}x_p\\ 2a_{21}x_1+2a_{22}x_2+\cdots+2a_{2p}x_p\\ \vdots\\ 2a_{p1}x_1+2a_{p2}x_2+\cdots+2a_{pp}x_p\end{pmatrix}=2\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1p}\\ a_{21}&a_{22}&\cdots&a_{2p}\\ \vdots&\vdots&&\vdots\\ a_{p1}&a_{p2}&\cdots&a_{pp}\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_p\end{pmatrix}=2\mathcal{A}\boldsymbol{x}$$

(c) [2 分] 证明二次型 $Q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}$ 的 Hessian 矩阵为

$$\frac{\partial^2 \boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}} = 2\mathcal{A} \tag{2.4}$$

【证明】 利用上述结论，我们看到

$$\frac{\partial^2 \boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}} = \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}\left(\frac{\partial \boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}}{\partial \boldsymbol{x}}\right) = \frac{\partial\left(2\mathcal{A}\boldsymbol{x}\right)}{\partial \boldsymbol{x}^{\mathrm{T}}} = \left(\frac{\partial\left(2\mathcal{A}\boldsymbol{x}\right)}{\partial x_1}, \frac{\partial\left(2\mathcal{A}\boldsymbol{x}\right)}{\partial x_2}, \cdots, \frac{\partial\left(2\mathcal{A}\boldsymbol{x}\right)}{\partial x_p}\right)$$

因为

$$\frac{\partial\left(2\mathcal{A}\boldsymbol{x}\right)}{\partial x_i} = \frac{\partial}{\partial x_i}\begin{pmatrix} 2a_{11}x_1 + 2a_{12}x_2 + \cdots + 2a_{1p}x_p \\ 2a_{21}x_1 + 2a_{22}x_2 + \cdots + 2a_{2p}x_p \\ \vdots \\ 2a_{p1}x_1 + 2a_{p2}x_2 + \cdots + 2a_{pp}x_p \end{pmatrix} = \begin{pmatrix} 2a_{1i} \\ 2a_{2i} \\ \vdots \\ 2a_{pi} \end{pmatrix}$$

于是，我们得到

$$\begin{aligned} \frac{\partial^2 \boldsymbol{x}^{\mathrm{T}}\mathcal{A}\boldsymbol{x}}{\partial \boldsymbol{x}\partial \boldsymbol{x}^{\mathrm{T}}} &= \left(\frac{\partial\left(2\mathcal{A}\boldsymbol{x}\right)}{\partial x_1}, \frac{\partial\left(2\mathcal{A}\boldsymbol{x}\right)}{\partial x_2}, \cdots, \frac{\partial\left(2\mathcal{A}\boldsymbol{x}\right)}{\partial x_p}\right) \\ &= \begin{pmatrix} 2a_{11} & 2a_{12} & \cdots & 2a_{1p} \\ 2a_{21} & 2a_{22} & \cdots & 2a_{2p} \\ \vdots & \vdots & & \vdots \\ 2a_{p1} & 2a_{p2} & \cdots & 2a_{pp} \end{pmatrix} = 2\mathcal{A} \end{aligned}$$

2. [2 分] 证明一个投影矩阵的特征值仅取值于集合 $\{0, 1\}$ 中.

【证明】 若 $\mathcal{P}$ 是一个投影矩阵，则由投影矩阵的定义有

$$\mathcal{P} = \mathcal{P}^{\mathrm{T}} = \mathcal{P}^2$$

设 λ 是 $\mathcal{P}$ 的特征值，$\boldsymbol{\gamma} \neq \mathbf{0}$ 是对应的特征向量，则有 $\mathcal{P}\boldsymbol{\gamma} = \lambda\boldsymbol{\gamma}$. 两侧左乘投影矩阵 $\mathcal{P}$ 可得

$$\begin{aligned} \mathcal{P}\cdot\mathcal{P}\boldsymbol{\gamma} = \mathcal{P}\left(\lambda\boldsymbol{\gamma}\right) \quad &\Longrightarrow \quad \mathcal{P}^2\boldsymbol{\gamma} = \lambda\mathcal{P}\boldsymbol{\gamma} = \lambda^2\boldsymbol{\gamma} \\ &\Longrightarrow \quad \mathcal{P}\boldsymbol{\gamma} = \lambda^2\boldsymbol{\gamma} \\ &\Longrightarrow \quad \lambda\boldsymbol{\gamma} = \lambda^2\boldsymbol{\gamma} \\ &\Longrightarrow \quad \left(\lambda - \lambda^2\right)\boldsymbol{\gamma} = \mathbf{0} \end{aligned}$$

因为 $\boldsymbol{\gamma} \neq \mathbf{0}$，所以必有 $(\lambda - \lambda^2) = 0$，即 $\lambda(1 - \lambda) = 0$. 故 λ 只能取值 0 或 1.

3. [2 分] 作度量矩阵为 $\mathcal{A} = \Sigma^{-1}$ 的某个等距椭球体的图形，其中

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{2.5}$$

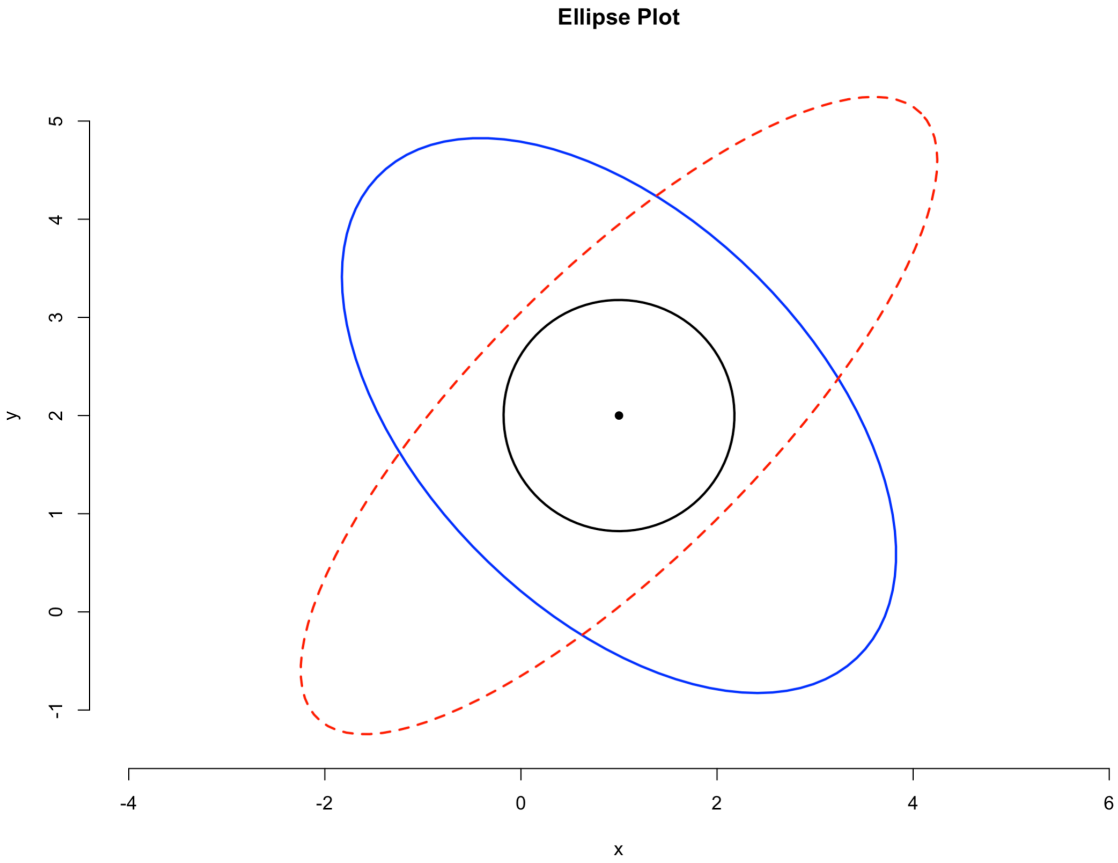

图 2.1: 以 $\mathcal{A}$ 为度量矩阵的等距椭球曲线.

【解】R 中的 ellipse 包提供了绘制椭圆的函数，以下利用 ellipse 包分别绘制 $\rho = \frac{1}{2}, -\frac{4}{5}, 0$ 时，以 $\mathcal{A}$ 为度量矩阵的等距椭球曲线，结果如图2.1 所示. 用到的 R 代码如下:

```
rm(list = ls(all = TRUE)) # 清除当前所有变量与对象
graphics.off() # 清除当前所有图形

library(ellipse) # 载入 ellipse 包
rho = 1/2
Sigma = matrix(c(1, rho, rho, 1), nrow = 2, byrow = TRUE)
A = solve(Sigma)
center = c(1, 2) # 椭球曲线的中心
Ellipse_Coords = ellipse(A, centre = center)
plot(Ellipse_Coords, type = "l", axes = FALSE, main = "Ellipse Plot",
```

```
     col = "blue", lwd = 2, xlim = c(-4, 6), asp = 1)
points(center[1], center[2], pch = 16)
axis(1); axis(2)

rho = -4/5
Sigma2 = matrix(c(1, rho, rho, 1), nrow = 2, byrow = TRUE)
B = solve(Sigma2)
Ellipse_Coords_2 = ellipse(B, centre = center, level = 0.85)
lines(Ellipse_Coords_2[, 1], Ellipse_Coords_2[, 2],
     col = "red", lwd = 2, lty = 2)

rho = 0
Sigma3 = matrix(c(1, rho, rho, 1), nrow = 2, byrow = TRUE)
C = solve(Sigma3)
Ellipse_Coords_3 = ellipse(C, centre = center, level = 0.5)
lines(Ellipse_Coords_3[, 1], Ellipse_Coords_3[, 2],
     col = "black", lwd = 2, lty = 1)
```

4. 对于课堂中讨论过的汽车数据集，

(a) [2 分] 计算变量 $X_2 = \text{miles per gallon}$ 与 $X_8 = \text{weight}$ 的协方差.

【解】 X_2 与 X_8 的协方差的无偏估计量为

$$\widehat{\text{Cov}}(X_2,\ X_8) = -3732.025$$

X_2 与 X_8 的协方差的有偏估计量为

$$\widehat{\text{Cov}}(X_2,\ X_8) = -3681.592$$

上述结果的 R 代码如下：

```
library(corrgram)
cov(auto$MPG, auto$Weight) # 无偏估计
((dim(auto)[1] - 1) / dim(auto)[1]) * cov(auto$MPG, auto$Weight) # 有偏估计
```

(b) [2 分] 你期待协方差的符号是正还是负，为什么?

【解】 【参考答案】 作 X_2 与 X_8 的散点图，由图 2.2 可见，随着 Weight 的增大 MPG 在降低，所以 X_2 与 X_8 的协方差的符号应为负. 散点图的 R 代码如下：

```
library(ggplot2)
```

```
library(plotly)
plot_ly(auto, x = ~Weight, y = ~MPG, name = "default") %>%
    add_markers(alpha = 0.9, name = "alpha")
```

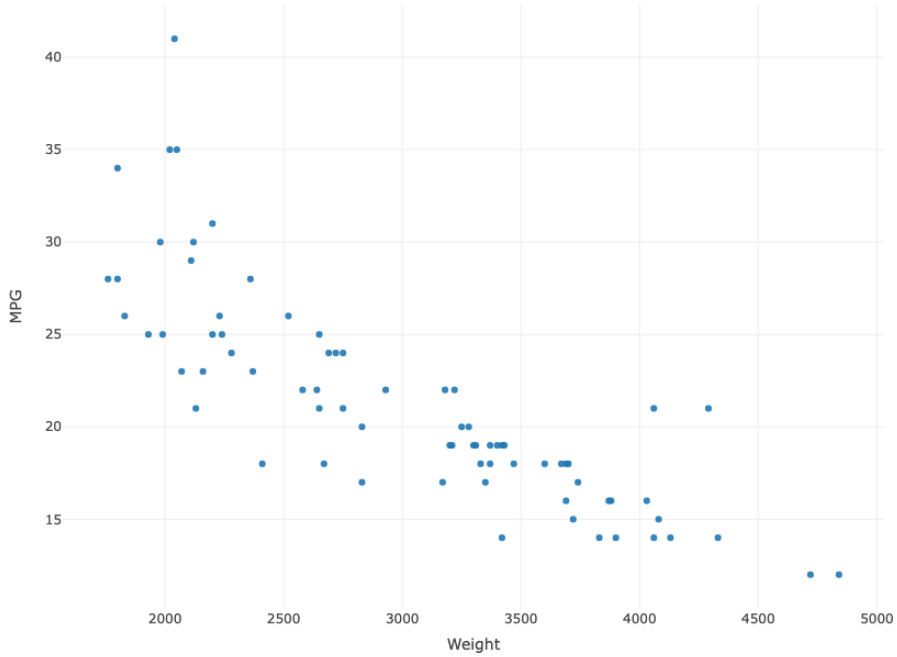

图 2.2: auto 数据集 MPG 与 Weight 的散点图.

5. 一位纺织店经理研究“经典蓝色”套头衫在 10 个不同时期的销售情况. 他调查了销量 (X_1)；价格的变化 (X_2)，单位：欧元；当地报纸的广告费用 (X_3)，单位：欧元；以及是否有促销员 (X_4)，促销员的时长，单位：小时. 所得观测数据矩阵如下：

$$\mathcal{X} = \begin{pmatrix} 230 & 125 & 200 & 109 \\ 181 & 99 & 55 & 107 \\ 165 & 97 & 105 & 98 \\ 150 & 115 & 85 & 71 \\ 97 & 120 & 0 & 82 \\ 192 & 100 & 150 & 103 \\ 181 & 80 & 85 & 111 \\ 189 & 90 & 120 & 93 \\ 172 & 95 & 110 & 86 \\ 170 & 125 & 130 & 78 \end{pmatrix} \tag{2.6}$$

(a) [2 分] 计算 $\mathcal{X}$ 的样本相关矩阵.

【解】 $\mathcal{X}$ 的样本相关矩阵为：

$$R = \begin{pmatrix} 1.0000 & -0.1676 & 0.8672 & 0.6329 \\ -0.1676 & 1.0000 & 0.1213 & -0.4638 \\ 0.8672 & 0.1213 & 1.0000 & 0.3083 \\ 0.6329 & -0.4638 & 0.3083 & 1.0000 \end{pmatrix} \tag{2.7}$$

计算样本相关矩阵的 R 代码如下：

```
x = c(230, 125, 200, 109, 181, 99, 55, 107, 165, 97, 105, 98, 150,
      115, 85, 71, 97, 120, 0, 82, 192, 100, 150, 103 , 181, 80,
      85, 111, 189, 90, 120, 93, 172, 95, 110, 86, 170, 125, 130, 78)
Pullover = matrix(x, ncol = 4, byrow = TRUE)
options(digits = 4)
cor(Pullover)
```

(b) [2 分] 就相关系数的符号进行说明.

【解】 相关系数大于零意味着正相关，相关系数小于零意味着负相关，其绝对值越接近于 1 意味着线性相关性越强，其绝对值越接近于 0 意味着线性相关性越弱. 例如，计算所得相关矩阵中 $\rho_{X_1X_3} = 0.8672$，说明广告费用 X_3 与销量 X_1 有较强的正相关 (如图 2.3). 再比如，相关矩阵中 $\rho_{X_2X_4} = -0.4638$，说明促销时长 X_4 与价格 X_2 呈现弱的负相关 (如图 2.4).

所用 R 代码如下：

```
pullover = as.data.frame(Pullover)
names(pullover) = c("sales", "price", "advertisement", "promoters")
p1 = plot_ly(pullover, x = ~advertisement, y = ~sales) %>%
    add_markers()
p1
p2 = plot_ly(pullover, x = ~promoters, y = ~price) %>%
    add_markers()
p2
```

(c) [2 分] 检验假设 $\rho_{X_1X_2} = 0$.

【解】 由相关矩阵知

$$\rho_{X_1X_2} = -0.1676$$

对 $\rho_{X_1X_2}$ 作 Fisher 的 Z 变换，计算可得

$$W = \frac{1}{2}\log\left(\frac{1+\rho_{X_1X_2}}{1-\rho_{X_1X_2}}\right) = -0.1692$$

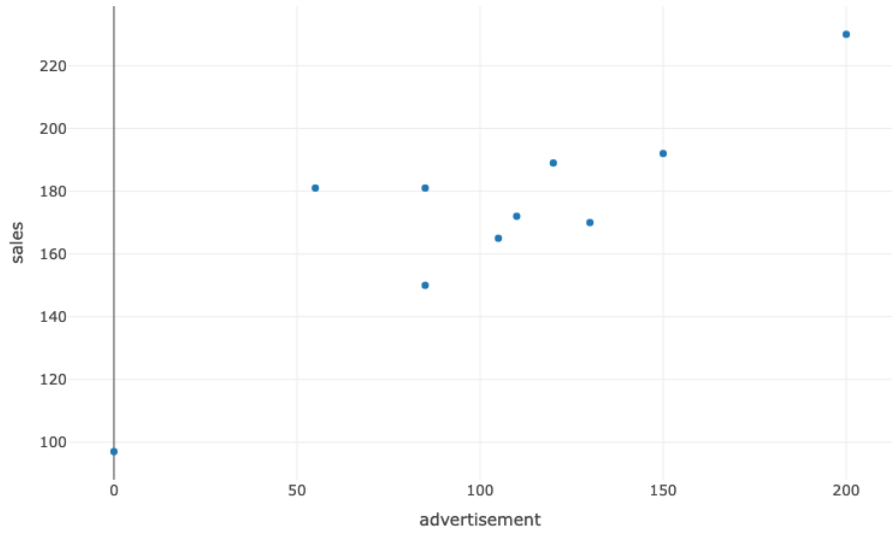

图 2.3: Pullover 数据集 sales 与 advertisement 的散点图.

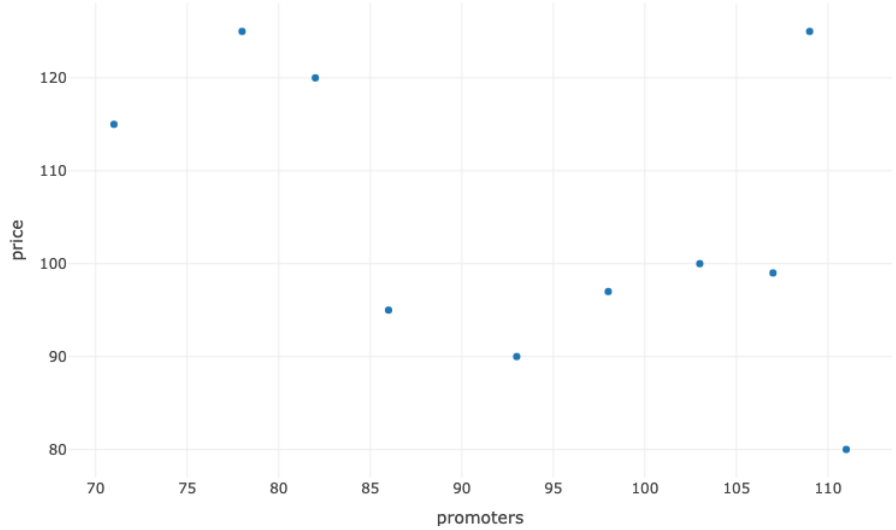

图 2.4: Pullover 数据集 price 与 promoters 的散点图.

由于样本容量 $n = 10$ 较小，所以进一步对 W 作改进的 Hotelling 变换，算得

$$W^* = W - \frac{3W + \tanh(W)}{4(n-1)} = -0.1504, \quad \text{且} \quad \mathbb{V}\mathrm{ar}\,(W^*) = \frac{1}{n-1} = \frac{1}{9}$$

对于假设 H_0 : $\rho_{x_1 x_2} = 0$，检验统计量的取值为

$$Z = \frac{W^* - 0}{\sqrt{\mathbb{V}\mathrm{ar}\,(W^*)}} = -0.4513$$

检验的 p 值为

$$2 \times \mathbb{P}(Z > 0.4513) = 0.6518$$

检验的显著水平 $\alpha = 0.05$ 时，因为该检验的 p 值 $0.6518 > 0.05$，所以我们接受假设 H_0 :

$\rho_{X_1X_2} = 0$.

计算所用的 R 代码如下：

```
r_12 = cor(pullover$sales, pullover$price)
r_12
w = (log((1 + r_12) / (1 - r_12))) / 2
w
w_star = w - (3 * w + tanh(w)) / (4 * (10 - 1))
w_star
z = (w_star - 0) / (sqrt(1/(10-1)))
z
p_value = 2 * pnorm(abs(z), 0, 1, lower.tail = FALSE)
p_value
```

6. [2 分] 证明 $\mathrm{rank}(\mathcal{H}) = \mathrm{tr}(\mathcal{H}) = n - 1$, 其中 $\mathcal{H} = \mathcal{I}_p - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$.

【证明】 因为

$$\mathcal{H} = \mathcal{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{pmatrix}$$

所以

$$\mathrm{tr}(\mathcal{H}) = \left(1-\frac{1}{n}\right) + \left(1-\frac{1}{n}\right) + \cdots + \left(1-\frac{1}{n}\right) = n-1$$

我们对矩阵 $\mathcal{H}$ 作以下的初等变换：

$$\mathcal{H} \xrightarrow{r_2-r_1,\ \cdots,\ r_n-r_1} \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -1 & 1 & 0 & \cdots & 0 \\ -1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -1 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

$$\xrightarrow{c_1+c_2,\ \cdots,\ c_1+c_n} \begin{pmatrix} 0 & -\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

由此易见 $\mathrm{rank}(\mathcal{H}) = n - 1$.

7. 设 $\mathcal{X}$ 表示课堂中讨论过的钞票数据集当中伪钞数据的观测矩阵.

(a) [2 分] 计算 $\mathcal{X}$ 的样本协方差矩阵 $\mathcal{S} = \mathrm{Cov}(\mathcal{X})$.

【解】 计算 $\mathcal{X}$ 的样本协方差矩阵的无偏估计的 R 代码为:

```
library(mclust)
data("banknote")
counterfeit = subset(banknote, Status == "counterfeit")
S.f = cov(counterfeit[, 2:7])
round(S.f, digits = 4)
```

所得结果如下:

$$\mathcal{S} = \mathrm{Cov}(\mathcal{X}) = \begin{pmatrix} 0.1240 & 0.0315 & 0.0240 & -0.1006 & 0.0194 & 0.0116 \\ 0.0315 & 0.0651 & 0.0468 & -0.0240 & -0.0119 & -0.0051 \\ 0.0240 & 0.0468 & 0.0889 & -0.0186 & 0.0001 & 0.0342 \\ -0.1006 & -0.0240 & -0.0186 & 1.2813 & -0.4902 & 0.2385 \\ 0.0194 & -0.0119 & 0.0001 & -0.4902 & 0.4045 & -0.0221 \\ 0.0116 & -0.0051 & 0.0342 & 0.2385 & -0.0221 & 0.3112 \end{pmatrix} \tag{2.8}$$

(b) [2 分] 作 $\mathcal{S}$ 的 Jordan 分解.

【解】 协方差矩阵 $\mathcal{S}$ 的特征值如下:

$$\lambda_1 = 1.5493\ ,\ \lambda_2 = 0.3192\ ,\ \lambda_3 = 0.1937\ ,\ \lambda_4 = 0.1039\ ,\ \lambda_5 = 0.0844\ ,\ \lambda_6 = 0.0245$$

对应的特征向量构成的矩阵为:

$$\begin{aligned} \Gamma &= (\gamma_1\ ,\ \gamma_2\ ,\ \gamma_3\ ,\ \gamma_4\ ,\ \gamma_5\ ,\ \gamma_6) \\ &= \begin{pmatrix} -0.0679 & 0.0877 & -0.5297 & -0.2943 & 0.7741 & 0.1461 \\ -0.0138 & -0.0101 & -0.3633 & -0.4184 & -0.2583 & -0.7912 \\ -0.0088 & 0.1301 & -0.4001 & -0.4149 & -0.5571 & 0.5835 \\ 0.9004 & 0.0511 & 0.2402 & -0.3433 & 0.1030 & 0.0212 \\ -0.3900 & 0.4934 & 0.5634 & -0.5268 & 0.0972 & -0.0114 \\ 0.1796 & 0.8540 & -0.2288 & 0.4134 & -0.0601 & -0.1080 \end{pmatrix} \end{aligned}$$

令

$$\Lambda = \mathrm{diag}\,(\lambda_1,\ \lambda_2,\ \lambda_3,\ \lambda_4,\ \lambda_5,\ \lambda_6) = \begin{pmatrix} 1.5493 & & & & & \\ & 0.3192 & & & & \\ & & 0.1937 & & & \\ & & & 0.1039 & & \\ & & & & 0.0844 & \\ & & & & & 0.0245 \end{pmatrix}$$

则有

$$\mathcal{S} = \Gamma\,\Lambda\,\Gamma^{\mathrm{T}}$$

计算特征值、特征向量以及结果验证的 R 代码如下:

```
X = eigen(S.f)
round(X$values, digits = 4)
Lambda.S = round(diag(X$values), digits = 4)
Lambda.S
Gamma.S = round(as.matrix(X$vectors), digits = 4)
Gamma.S
round(Gamma.S %*% Lambda.S %*% t(Gamma.S), digits = 4)
```

(c) [2 分] 为什么所有的特征值均为正?

【解】 由于矩阵 $\mathcal{S}$ 是一个正定矩阵，即 $\mathcal{S} > 0$，所以它的所有特征值均为正.

Chapter 3

第 3 周作业参考答案

1. 设 $\boldsymbol{X}=(X_1\,,\,X_2)^{\mathrm{T}}$ 是二维随机向量，且

$$\mathbb{E}(\boldsymbol{X})=\begin{pmatrix}0\\0\end{pmatrix}\,,\quad \mathbb{V}\mathrm{ar}(\boldsymbol{X})=\begin{pmatrix}1&0\\0&2\end{pmatrix}$$

(a) [2 分] 定义 $Y=X_1+X_2$，则 Y 是 $\boldsymbol{X}$ 的一个线性变换，写出变换矩阵 $\mathcal{A}$.

【解】 因为

$$Y=X_1+X_2=\begin{pmatrix}1&1\end{pmatrix}\begin{pmatrix}X_1\\X_2\end{pmatrix}$$

所以变换矩阵为

$$\mathcal{A}=\begin{pmatrix}1&1\end{pmatrix}$$

(b) [2 分] 计算 $\mathbb{V}\mathrm{ar}(Y)$.

【解】 因为 $Y=\mathcal{A}\boldsymbol{X}$，所以

$$\mathbb{V}\mathrm{ar}(Y)=\mathcal{A}\,\mathbb{V}\mathrm{ar}(\boldsymbol{X})\,\mathcal{A}^{\mathrm{T}}=\begin{pmatrix}1&1\end{pmatrix}\begin{pmatrix}1&0\\0&2\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}=3$$

2. [2 分] 设 $\boldsymbol{X}=(X_1\,,\,X_2)^{\mathrm{T}}$ 的联合概率密度函数为

$$f\,(x_1\,,\,x_2)=\begin{cases}\mathrm{e}^{-(x_1+x_2)}\,, & x_1>0\,,\,x_2>0\\ 0\,, & \text{其它}\end{cases}\tag{3.1}$$

令 $U_1=X_1+X_2$，$U_2=X_1-X_2$，求 $\boldsymbol{U}=(U_1\,,\,U_2)^{\mathrm{T}}$ 的联合概率密度函数.

【解】 因为

$$\begin{cases} u_1 = x_1 + x_2 \\ u_2 = x_1 - x_2 \end{cases} \Longrightarrow \begin{cases} x_1 = \dfrac{1}{2}u_1 + \dfrac{1}{2}u_2 \\ x_2 = \dfrac{1}{2}u_1 - \dfrac{1}{2}u_2 \end{cases}$$

该变换的 Jacobian 行列式为

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial u_1} & \dfrac{\partial x_1}{\partial u_2} \\ \dfrac{\partial x_2}{\partial u_1} & \dfrac{\partial x_2}{\partial u_2} \end{vmatrix} = \begin{vmatrix} \dfrac{1}{2} & \dfrac{1}{2} \\ \dfrac{1}{2} & -\dfrac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

于是，U_1 和 U_2 的联合概率密度函数为

$$f_{U_1,U_2}(u_1, u_2) = f_{X_1,X_2}(x_1(u_1, u_2), x_2(u_1, u_2)) \cdot |J| = \mathrm{e}^{-u_1} \cdot \left|-\frac{1}{2}\right| = \frac{1}{2}\mathrm{e}^{-u_1}$$

其中

$$\begin{cases} x_1 = \dfrac{1}{2}u_1 + \dfrac{1}{2}u_2 > 0 \\ x_2 = \dfrac{1}{2}u_1 - \dfrac{1}{2}u_2 > 0 \end{cases} \Longrightarrow \quad u_1 > |u_2|, -\infty < u_2 < \infty$$

即

$$f(u_1, u_2) = \frac{1}{2}\mathrm{e}^{-u_1}, \quad u_1 > |u_2|, -\infty < u_2 < \infty$$

3. 假设

$$f(x_1, x_2, x_3) = \begin{cases} k(x_1 + x_2 x_3), & 0 < x_1, x_2, x_3 < 1 \\ 0, & \text{其它} \end{cases} \tag{3.2}$$

(a) [2 分] 确定 k 的值，使得 f 是 $\boldsymbol{X} = (X_1, X_2, X_3)^{\mathrm{T}}$ 的概率密度函数.

【解】 因为

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty}\!\!\!\!\!\!\iiint f(x_1, x_2, x_3)\,\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \\ &= k \iiint_{0<x_1, x_2, x_3<1} x_1\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 + k \iiint_{0<x_1, x_2, x_3<1} x_2 x_3\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 \\ &= k\int_0^1 x_1\,\mathrm{d}x_1 + k\int_0^1 x_2\,\mathrm{d}x_2 \int_0^1 x_3\,\mathrm{d}x_3 \\ &= \frac{1}{2}k + \frac{1}{4}k = \frac{3}{4}k \end{aligned}$$

所以, $k=\dfrac{4}{3}$, 故

$$f\left(x_1,\ x_2,\ x_3\right)=\begin{cases}\dfrac{4}{3}\left(x_1+x_2x_3\right), & 0<x_1,\ x_2,\ x_3<1\\ 0, & \text{其它}\end{cases}$$

(b) [2 分] 计算 $\Sigma_{\boldsymbol{X}}=\mathbb{V}\mathrm{ar}(\boldsymbol{X})$.

【解】 首先计算边际密度函数如下

$$f(x_1,\ x_2)=\int_0^1 f\left(x_1,\ x_2,\ x_3\right)\mathrm{d}x_3=\frac{4}{3}x_1+\frac{2}{3}x_2\ ,\quad 0<x_1,\ x_2<1$$

$$f(x_1,\ x_3)=\int_0^1 f\left(x_1,\ x_2,\ x_3\right)\mathrm{d}x_2=\frac{4}{3}x_1+\frac{2}{3}x_3\ ,\quad 0<x_1,\ x_3<1$$

$$f(x_2,\ x_3)=\int_0^1 f\left(x_1,\ x_2,\ x_3\right)\mathrm{d}x_1=\frac{2}{3}+\frac{4}{3}x_2x_3\ ,\quad 0<x_2,\ x_3<1$$

$$f(x_1)=\int_0^1 f\left(x_1,\ x_2\right)\mathrm{d}x_2=\frac{4}{3}x_1+\frac{1}{3}\ ,\quad 0<x_1<1$$

$$f(x_2)=\int_0^1 f\left(x_1,\ x_2\right)\mathrm{d}x_1=\frac{2}{3}+\frac{2}{3}x_2\ ,\quad 0<x_2<1$$

$$f(x_3)=\int_0^1 f\left(x_1,\ x_3\right)\mathrm{d}x_1=\frac{2}{3}+\frac{2}{3}x_3\ ,\quad 0<x_3<1$$

于是我们有

$$\mathbb{E}\left(X_1\right)=\int_0^1 x_1\left(\frac{4}{3}x_1+\frac{1}{3}\right)\mathrm{d}x_1=\frac{11}{18}\ ,\qquad \mathbb{E}\left(X_1^2\right)=\int_0^1 x_1^2\left(\frac{4}{3}x_1+\frac{1}{3}\right)\mathrm{d}x_1=\frac{4}{9}$$

$$\mathbb{E}\left(X_2\right)=\int_0^1 x_2\left(\frac{2}{3}+\frac{2}{3}x_2\right)\mathrm{d}x_2=\frac{5}{9}\ ,\qquad \mathbb{E}\left(X_2^2\right)=\int_0^1 x_2^2\left(\frac{2}{3}+\frac{2}{3}x_2\right)\mathrm{d}x_2=\frac{7}{18}$$

$$\mathbb{E}\left(X_3\right)=\int_0^1 x_3\left(\frac{2}{3}+\frac{2}{3}x_3\right)\mathrm{d}x_3=\frac{5}{9}\ ,\qquad \mathbb{E}\left(X_3^2\right)=\int_0^1 x_3^2\left(\frac{2}{3}+\frac{2}{3}x_3\right)\mathrm{d}x_3=\frac{7}{18}$$

因此

$$\mathbb{V}\mathrm{ar}\left(X_1\right)=\mathbb{E}\left(X_1^2\right)-\left[\mathbb{E}\left(X_1\right)\right]^2=\frac{4}{9}-\left(\frac{11}{18}\right)^2=\frac{23}{324}$$

$$\mathbb{V}\mathrm{ar}\left(X_2\right)=\mathbb{E}\left(X_2^2\right)-\left[\mathbb{E}\left(X_2\right)\right]^2=\frac{7}{18}-\left(\frac{5}{9}\right)^2=\frac{13}{162}$$

$$\mathbb{V}\mathrm{ar}\left(X_3\right)=\mathbb{E}\left(X_3^2\right)-\left[\mathbb{E}\left(X_3\right)\right]^2=\frac{7}{18}-\left(\frac{5}{9}\right)^2=\frac{13}{162}$$

再计算得

$$\begin{aligned}\operatorname{Cov}(X_1,\ X_2) &= \int_0^1\int_0^1 x_1x_2f(x_1,\ x_2)\mathrm{d}x_1\mathrm{d}x_2 - \mathbb{E}(X_1)\,\mathbb{E}(X_2)\\ &= \int_0^1\int_0^1 x_1x_2\left(\frac{4}{3}x_1+\frac{2}{3}x_2\right)\mathrm{d}x_1\mathrm{d}x_2 - \frac{11}{18}\cdot\frac{5}{9}\\ &= \frac{1}{3}-\frac{55}{162} = -\frac{1}{162}\end{aligned}$$

$$\begin{aligned}\operatorname{Cov}(X_1,\ X_3) &= \int_0^1\int_0^1 x_1x_3f(x_1,\ x_3)\mathrm{d}x_1\mathrm{d}x_3 - \mathbb{E}(X_1)\,\mathbb{E}(X_3)\\ &= \int_0^1\int_0^1 x_1x_3\left(\frac{4}{3}x_1+\frac{2}{3}x3\right)\mathrm{d}x_1\mathrm{d}x_3 - \frac{11}{18}\cdot\frac{5}{9}\\ &= \frac{1}{3}-\frac{55}{162} = -\frac{1}{162}\end{aligned}$$

$$\begin{aligned}\operatorname{Cov}(X_2,\ X_3) &= \int_0^1\int_0^1 x_2x_3f(x_2,\ x_3)\mathrm{d}x_2\mathrm{d}x_3 - \mathbb{E}(X_2)\,\mathbb{E}(X_3)\\ &= \int_0^1\int_0^1 x_2x_3\left(\frac{2}{3}+\frac{4}{3}x_2x_3\right)\mathrm{d}x_2\mathrm{d}x_3 - \frac{5}{9}\cdot\frac{5}{9}\\ &= \frac{17}{54}-\frac{49}{324} = \frac{1}{162}\end{aligned}$$

因此

$$\Sigma = \begin{pmatrix} \frac{23}{324} & -\frac{1}{162} & -\frac{1}{162}\\ -\frac{1}{162} & \frac{13}{162} & \frac{1}{162}\\ -\frac{1}{162} & \frac{1}{162} & \frac{13}{162}\end{pmatrix}$$

(c) [2 分] 给定 $X_1 = x_1$ 时，计算 $(X_2\,,\,X_3)$ 的条件协方差矩阵.

【解】 当 $0 < x_1 < 1$ 时，因为

$$f(x_2,\ x_3\,|x_1) = \frac{f(x_1,\ x_2,\ x_3)}{f(x_1)} = \frac{4(x_1+x_2x_3)}{4x_1+1}\,,\quad 0<x_2,\ x_3<1$$

$$f(x_2\,|x_1) = \frac{f(x_1,\ x_2)}{f(x_1)} = \frac{4x_1+2x_2}{4x_1+1}\,,\quad 0<x_2<1$$

$$f(x_3\,|x_1) = \frac{f(x_1,\ x_3)}{f(x_1)} = \frac{4x_1+2x_3}{4x_1+1}\,,\quad 0<x_3<1$$

于是，我们有

$$\mathbb{E}\left(X_2\,|X_1\right)=\int_0^1 x_2 f\left(x_2\,|x_1\right)\mathrm{d}x_2=\int_0^1 x_2\frac{4x_1+2x_2}{4x_1+1}\mathrm{d}x_2=\frac{6x_1+2}{3\left(4x_1+1\right)}$$

$$\mathbb{E}\left(X_2^2\,|X_1\right)=\int_0^1 x_2^2 f\left(x_2\,|x_1\right)\mathrm{d}x_2=\int_0^1 x_2^2\frac{4x_1+2x_2}{4x_1+1}\mathrm{d}x_2=\frac{8x_1+3}{6\left(4x_1+1\right)}$$

$$\mathbb{E}\left(X_3\,|X_1\right)=\int_0^1 x_3 f\left(x_3\,|x_1\right)\mathrm{d}x_3=\int_0^1 x_3\frac{4x_1+2x_3}{4x_1+1}\mathrm{d}x_3=\frac{6x_1+2}{3\left(4x_1+1\right)}$$

$$\mathbb{E}\left(X_3^2\,|X_1\right)=\int_0^1 x_3^2 f\left(x_3\,|x_1\right)\mathrm{d}x_3=\int_0^1 x_3^2\frac{4x_1+2x_3}{4x_1+1}\mathrm{d}x_3=\frac{8x_1+3}{6\left(4x_1+1\right)}$$

从而可得

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(X_2\,|X_1\right)&=\mathbb{E}\left(X_2^2\,|X_1\right)-\left[\mathbb{E}\left(X_2\,|X_1\right)\right]^2=\frac{8x_1+3}{6\left(4x_1+1\right)}-\left[\frac{6x_1+2}{3\left(4x_1+1\right)}\right]^2\\&=\frac{24x_1^2+12x_1+1}{18\left(4x_1+1\right)^2}\end{aligned}$$

$$\mathbb{V}\mathrm{ar}\left(X_3\,|X_1\right)=\mathbb{E}\left(X_3^2\,|X_1\right)-\left[\mathbb{E}\left(X_3\,|X_1\right)\right]^2=\frac{24x_1^2+12x_1+1}{18\left(4x_1+1\right)^2}$$

$$\begin{aligned}\mathrm{Cov}\left(X_2,\ X_3\,|X_1\right)&=\int_0^1\int_0^1 x_2x_3 f\left(x_2,\ x_3\,|x_1\right)\mathrm{d}x_2\mathrm{d}x_3-\mathbb{E}\left(X_2\,|X_1\right)\mathbb{E}\left(X_3\,|X_1\right)\\&=\int_0^1\int_0^1 x_2x_3\frac{4\left(x_1+x_2x_3\right)}{4x_1+1}\mathrm{d}x_2\mathrm{d}x_3-\frac{6x_1+2}{3\left(4x_1+1\right)}\cdot\frac{6x_1+2}{3\left(4x_1+1\right)}\\&=\frac{1}{9\left(4x_1+1\right)^2}\end{aligned}$$

所以，给定 $X_1=x_1$ 时，$(X_2,\ X_3)$ 的条件协方差矩阵为

$$\Sigma_{(X_2,\ X_3)|X_1=x_1}=\begin{pmatrix}\dfrac{24x_1^2+12x_1+1}{18\left(4x_1+1\right)^2} & \dfrac{1}{9\left(4x_1+1\right)^2}\\ \dfrac{1}{9\left(4x_1+1\right)^2} & \dfrac{24x_1^2+12x_1+1}{18\left(4x_1+1\right)^2}\end{pmatrix}$$

4. 设有概率密度函数

$$f\left(x_1\,,\,x_2\right)=\begin{cases}\dfrac{1}{2}\,\mathrm{e}^{-x_1}\,, & x_1>|x_2|\\ 0\,, & \text{其它}\end{cases}\tag{3.3}$$

(a) [2 分] 计算 $\mathbb{E}(\boldsymbol{X})$ 与 $\mathbb{V}\mathrm{ar}(\boldsymbol{X})$.

【解】 如图3.1，因为

$$\mathbb{E}\left(X_1\right)=\iint\limits_{x_1>|x_2|} x_1\cdot f\left(x_1,\ x_2\right)\mathrm{d}x_1\mathrm{d}x_2=\iint\limits_{x_1>|x_2|} x_1\cdot\frac{1}{2}\mathrm{e}^{-x_1}\mathrm{d}x_1\mathrm{d}x_2$$

$$=\frac{1}{2}\int_0^{\infty}x_1\mathrm{e}^{-x_1}\left(\int_{-x_1}^{x_1}\mathrm{d}x_2\right)\mathrm{d}x_1=2$$

$$\mathbb{E}\left(X_2\right)=\iint\limits_{x_1>|x_2|} x_2\cdot f\left(x_1,\ x_2\right)\mathrm{d}x_1\mathrm{d}x_2=\iint\limits_{x_1>|x_2|} x_2\cdot\frac{1}{2}\mathrm{e}^{-x_1}\mathrm{d}x_1\mathrm{d}x_2=0$$

所以

$$\mathbb{E}(\boldsymbol{X})=\begin{pmatrix}2\\0\end{pmatrix}$$

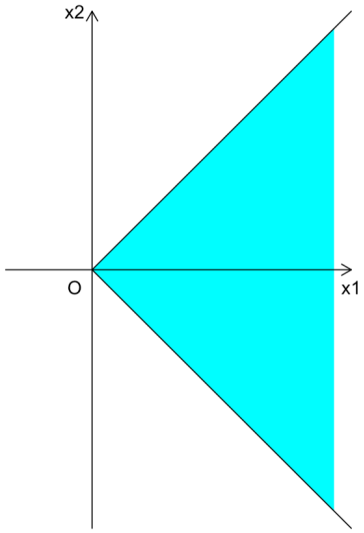

图 3.1: 第 4 题图.

又因为

$$\mathbb{E}\left(X_1^2\right)=\iint\limits_{x_1>|x_2|} x_1^2\cdot f\left(x_1,\ x_2\right)\mathrm{d}x_1\mathrm{d}x_2=\iint\limits_{x_1>|x_2|} x_1^2\cdot\frac{1}{2}\mathrm{e}^{-x_1}\mathrm{d}x_1\mathrm{d}x_2$$

$$=\frac{1}{2}\int_0^{\infty}x_1^2\mathrm{e}^{-x_1}\left(\int_{-x_1}^{x_1}\mathrm{d}x_2\right)\mathrm{d}x_1=6$$

$$\mathbb{E}\left(X_2^2\right)=\iint\limits_{x_1>|x_2|} x_2^2\cdot f\left(x_1,\ x_2\right)\mathrm{d}x_1\mathrm{d}x_2=\iint\limits_{x_1>|x_2|} x_2^2\cdot\frac{1}{2}\mathrm{e}^{-x_1}\mathrm{d}x_1\mathrm{d}x_2$$

$$=\frac{1}{2}\int_0^{\infty}\mathrm{e}^{-x_1}\left(\int_{-x_1}^{x_1}x_2^2\mathrm{d}x_2\right)\mathrm{d}x_1=2$$

所以

$$\mathbb{V}\mathrm{ar}\left(X_1\right)=\mathbb{E}\left(X_1^2\right)-\left[\mathbb{E}\left(X_1\right)\right]^2=6-2^2=2$$
$$\mathbb{V}\mathrm{ar}\left(X_2\right)=\mathbb{E}\left(X_2^2\right)-\left[\mathbb{E}\left(X_2\right)\right]^2=2-0^2=2$$

由于

$$\mathrm{Cov}\left(X_1,\ X_2\right)=\mathbb{E}\left(X_1X_2\right)-\mathbb{E}\left(X_1\right)\mathbb{E}\left(X_2\right)=\iint\limits_{x_1>|x_2|}x_1x_2\frac{1}{2}\mathrm{e}^{-x_1}\mathrm{d}x_1\mathrm{d}x_2-2\times0=0$$

因此

$$\mathbb{V}\mathrm{ar}(\boldsymbol{X})=\begin{pmatrix}2&0\\0&2\end{pmatrix}.$$

(b) [2 分] 计算 $\mathbb{E}\left(X_1\,|X_2\right)$ 与 $\mathbb{E}\left(X_2\,|X_1\right)$.

【解】 首先计算 X_1 与 X_2 的边际概率密度函数

$$f_{X_1}\left(x_1\right)=\int f\left(x_1,\ x_2\right)\mathrm{d}x_2=\int_{-x_1}^{x_1}\frac{1}{2}\mathrm{e}^{-x_1}\mathrm{d}x_2=x_1\mathrm{e}^{-x_1}\ ,\quad x_1>0$$
$$f_{X_2}\left(x_2\right)=\int f\left(x_1,\ x_2\right)\mathrm{d}x_1=\int_{|x_2|}^{\infty}\frac{1}{2}\mathrm{e}^{-x_1}\mathrm{d}x_1=\frac{1}{2}\mathrm{e}^{-|x_2|}\ ,\quad -\infty<x_2<\infty$$

对于给定的 $x_2\in(-\infty,\infty)$，$\left(X_1\,|X_2\right)$ 的条件密度为

$$f_{X_1|X_2}\left(x_1\,|x_2\right)=\frac{f\left(x_1,\ x_2\right)}{f_{X_2}\left(x_2\right)}=\mathrm{e}^{|x_2|-x_1}\ ,\quad x_1>|x_2|$$

对于给定的 $x_1>0$，$\left(X_2\,|X_1\right)$ 的条件密度为

$$f_{X_2|X_1}\left(x_2\,|x_1\right)=\frac{f\left(x_1,\ x_2\right)}{f_{X_1}\left(x_1\right)}=\frac{1}{2x_1}\ ,\quad |x_2|<x_1$$

因此

$$\mathbb{E}\left(X_1\,|X_2\right)=\int x_1\cdot f_{X_1|X_2}\left(x_1\,|x_2\right)\mathrm{d}x_1=\int_{|x_2|}^{\infty}x_1\cdot\mathrm{e}^{|x_2|-x_1}\mathrm{d}x_1=1+|x_2|$$
$$\mathbb{E}\left(X_2\,|X_1\right)=\int x_2\cdot f_{X_2|X_1}\left(x_1\,|x_2\right)\mathrm{d}x_2=\int_{-x_1}^{x_1}x_2\cdot\frac{1}{2x_1}\mathrm{d}x_2=0$$

(c) [2 分] 计算 $\mathbb{V}\mathrm{ar}\left(X_1\,|X_2\right)$ 与 $\mathbb{V}\mathrm{ar}\left(X_2\,|X_1\right)$.

【解】 因为

$$\mathbb{E}\left(X_1^2\,|X_2\right)=\int x_1^2\cdot f_{X_1|X_2}\left(x_1\,|x_2\right)\mathrm{d}x_1=\int_{|x_2|}^{\infty}x_1^2\cdot\mathrm{e}^{|x_2|-x_1}\mathrm{d}x_1=|x_2|^2+2\left(|x_2|+1\right)$$

$$\mathbb{E}\left(X_2^2\,|X_1\right)=\int x_2^2\cdot f_{X_2|X_1}\left(x_2\,|x_1\right)\mathrm{d}x_2=\int_{-x_1}^{x_1}x_2^2\cdot\frac{1}{2x_1}\mathrm{d}x_2=\frac{x_1^2}{3}$$

所以

$$\mathbb{V}\mathrm{ar}\left(X_1\,|X_2\right)=\mathbb{E}\left(X_1^2\,|X_2\right)-\left[\mathbb{E}\left(X_1\,|X_2\right)\right]^2=1$$

$$\mathbb{V}\mathrm{ar}\left(X_2\,|X_1\right)=\mathbb{E}\left(X_2^2\,|X_1\right)-\left[\mathbb{E}\left(X_2\,|X_1\right)\right]^2=\frac{x_1^2}{3}$$

5. 设有概率密度函数

$$f\left(x_1\,,\,x_2\right)=\begin{cases}\dfrac{3}{4}\,x_1^{-\frac{1}{2}}\,, & 0<x_1<x_2<1\\ 0\,, & \text{其它}\end{cases}\tag{3.4}$$

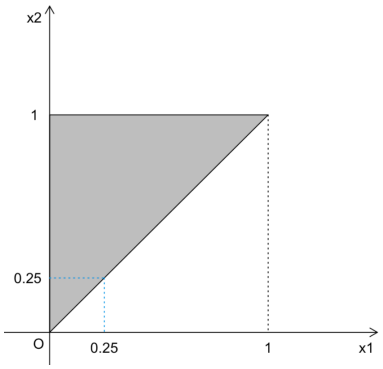

图 3.2: 第 4 题图.

(a) [2 分] 计算 $\mathbb{P}\left(X_1<0.25\right)$.

【解】 如图 3.2，X_1 与 X_2 的边际概率密度函数为

$$f_{X_1}\left(x_1\right)=\int f\left(x_1,\,x_2\right)\mathrm{d}x_2=\int_{x_1}^{1}\frac{3}{4}x_1^{-\frac{1}{2}}\mathrm{d}x_2=\frac{3}{4}\left(x_1^{-\frac{1}{2}}-x_1^{\frac{1}{2}}\right)\,,\quad 0<x_1<1$$

$$f_{X_2}\left(x_2\right)=\int f\left(x_1,\,x_2\right)\mathrm{d}x_1=\int_{0}^{x_2}\frac{3}{4}x_1^{-\frac{1}{2}}\mathrm{d}x_1=\frac{3}{2}x_2^{\frac{1}{2}}\,,\quad 0<x_2<1$$

所以

$$\mathbb{P}\left(X_1<0.25\right)=\int_0^{0.25}\frac{3}{4}\left(x_1^{-\frac{1}{2}}-x_1^{\frac{1}{2}}\right)\mathrm{d}x_1=\frac{11}{16}$$

(b) [2 分] 计算 $\mathbb{P}\left(X_2<0.25\right)$.

【解】 类似地，

$$\mathbb{P}\left(X_2<0.25\right)=\int_0^{0.25}\frac{3}{2}x_2^{\frac{1}{2}}\mathrm{d}x_2=\frac{1}{8}$$

(c) [2 分] 计算 $\mathbb{P}(X_2 < 0.25 \,|\, X_1 < 0.25)$.

【解】

$$\begin{aligned}
\mathbb{P}(X_2 < 0.25 | X_1 < 0.25) &= \frac{\mathbb{P}(X_1 < 0.25,\ X_2 < 0.25)}{P(X_1 < 0.25)} \\
&= \frac{16}{11} \times \iint\limits_{X_1<0.25,\ X_2<0.25} f(x_1,\ x_2)\,\mathrm{d}x_1\mathrm{d}x_2 \\
&= \frac{16}{11} \times \int_0^{0.25} \frac{3}{4} x_1^{-\frac{1}{2}} \left(\int_{x_1}^{0.25} \mathrm{d}x_2 \right) \mathrm{d}x_1 \\
&= \frac{16}{11} \times \frac{1}{8} = \frac{2}{11}
\end{aligned}$$

6. 设 $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$，其中

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 2 & a \\ a & 2 \end{pmatrix}. \tag{3.5}$$

(a) [2 分] 当 $a = 0, -\frac{1}{2}, +\frac{1}{2}, 1$ 时，分别作 $\boldsymbol{X}$ 的密度曲面的等值线椭圆的图形. 注意：要给出代码以及对应的图形！

【解】 利用 R 中的 ellipse 包中的 ellipse() 函数，可作二元正态密度曲面的等值线图，所用代码如下，结果如图3.3所示.

```
rm(list = ls(all = TRUE)) # 清除当前所有变量与对象
graphics.off() # 清除当前所有图形
library(ellipse)
f = function(mu = c(1, 2), a = 0) {
  S = matrix(c(2, a, a, 2), nrow = 2, byrow = TRUE)
  plot(ellipse(S, center = mu), type = "l", asp = 1, axes = FALSE,
       xlab = '', ylab = '', lty = 2, xlim = c(-3, 6), ylim = c(-2, 7))
  p = seq(0, 1.0, by = 0.05)
  n = length(p)
  for (i in 1:n) {
    lines(ellipse(S, center = mu, level = p[i]), col = 'blue', lty = 1)
  }
  points(mu[1], mu[2], pch = 16)
  arrows(-3, 0, 6, 0, length = 0.1)
  arrows(0, -2, 0, 7, length = 0.1)
}
par(mfrow = c(2, 2))
f()
```

```
title(sub = "a = 0")
f(a = -0.5)
title(sub = "a = -0.5")
f(a = 0.5)
title(sub = "a = 0.5")
f(a = 1)
title(sub = "a = 1")
```

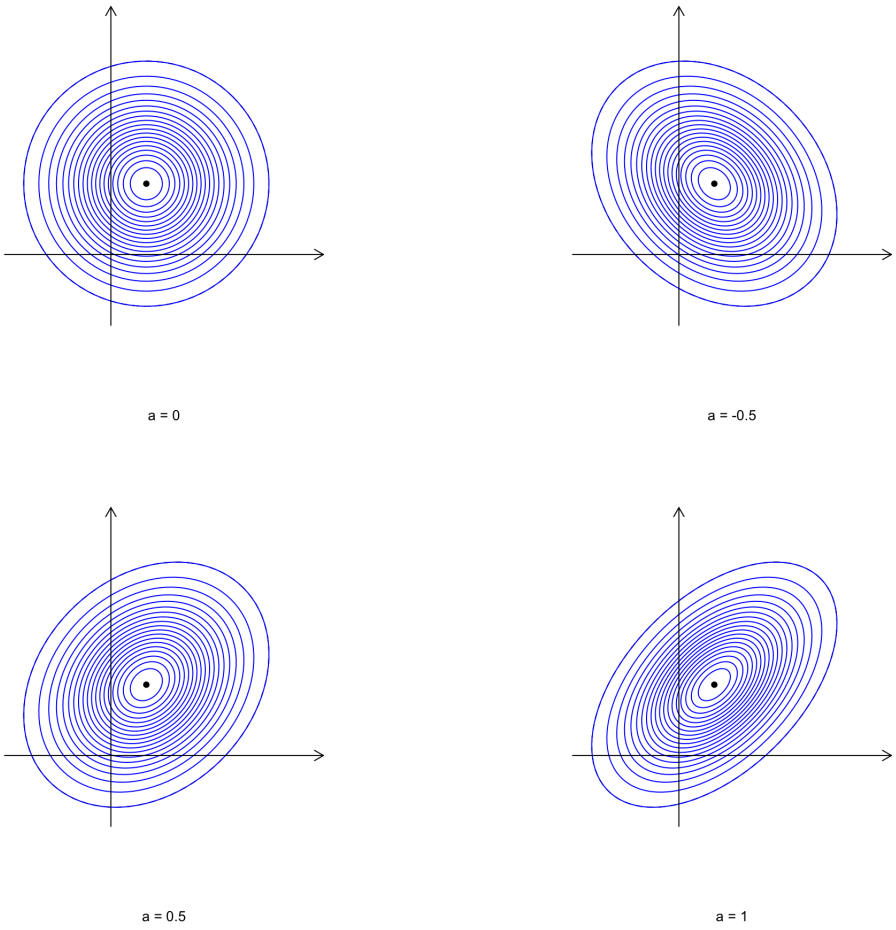

图 3.3: 第 6 题 (a) 图.

(b) [2 分] 对 $a=\dfrac{1}{2}$，确定以 $\boldsymbol{\mu}$ 为中心的 $\boldsymbol{X}$ 的区域，该区域以 0.90 的概率覆盖真实参数 $\boldsymbol{\mu}$，画出该区域的图形.

【解】 根据定理 4.7：如果 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\Sigma)$，则 $U=(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\sim\chi^2_p$. 当 $a=\dfrac{1}{2}$ 时，本题中

$$p=2\,,\quad \boldsymbol{\mu}=\begin{pmatrix}1\\2\end{pmatrix}\,,\quad \Sigma=\begin{pmatrix}2&\frac{1}{2}\\ \frac{1}{2}&2\end{pmatrix}\quad\Longrightarrow\quad \Sigma^{-1}=\begin{pmatrix}\frac{8}{15}&-\frac{2}{15}\\ -\frac{2}{15}&\frac{8}{15}\end{pmatrix}$$

所以有

$$\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)^{\mathrm{T}} \begin{pmatrix} \frac{8}{15} & -\frac{2}{15} \\ -\frac{2}{15} & \frac{8}{15} \end{pmatrix} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) \sim \chi_2^2$$

即

$$\frac{8}{15}(x_1-1)^2 - \frac{4}{15}(x_1-1)(x_2-2) + \frac{8}{15}(x_2-2)^2 \sim \chi_2^2$$

于是，以 $\boldsymbol{\mu}$ 为中心、以 0.90 的概率覆盖参数 $\boldsymbol{\mu}$ 的区域为

$$\left\{(x_1, x_2) \middle| \frac{8}{15}(x_1-1)^2 - \frac{4}{15}(x_1-1)(x_2-2) + \frac{8}{15}(x_2-2)^2 < \chi_2^2(0.10) = 4.60517\right\}$$

该区域的图形如图3.4所示，绘图的 R 代码如下：

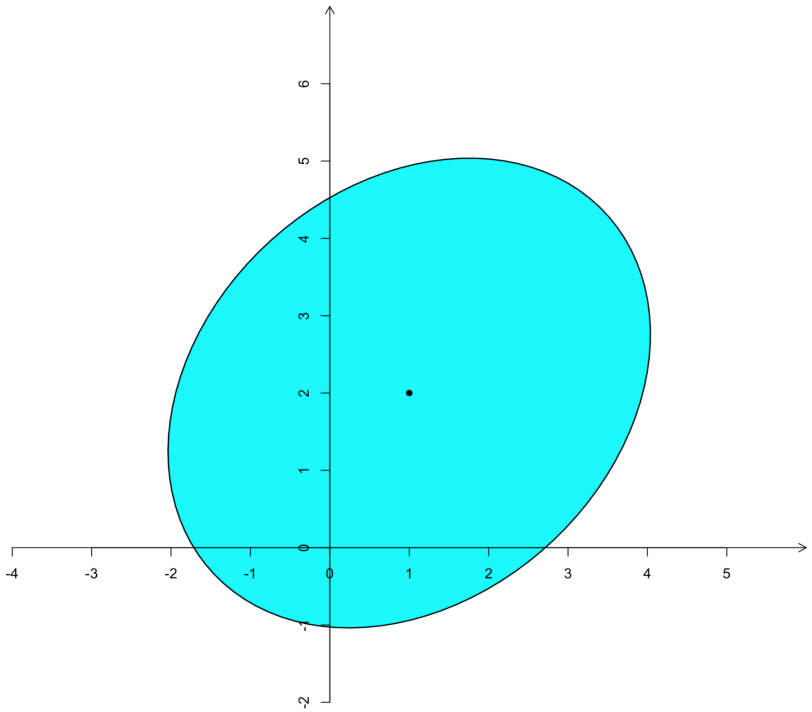

图 3.4: 第 6 题 (b) 图.

```
mu = c(1, 2)
S = matrix(c(2, 0.5, 0.5, 2), nrow = 2, byrow = TRUE)
plot(ellipse(S, center = mu, level = 0.9), type = "l", asp = 1, axes = FALSE,
     xlab = '', ylab = '', lty = 1, xlim = c(-3, 6), ylim = c(-2, 7), lwd = 2)
polygon(ellipse(S, center = mu, level = 0.9), col = "cyan")
points(mu[1], mu[2], pch = 16)
arrows(-3, 0, 6, 0, length = 0.1)
arrows(0, -2, 0, 7, length = 0.1)
axis(1, at = -4:5, pos = 0)
axis(2, at = -2:6, pos = 0)
```

7. 设有概率密度函数

$$f(x_1, x_2) = \begin{cases} \dfrac{1}{8x_2} \mathrm{e}^{-\left(\frac{x_1}{2x_2}+\frac{x_2}{4}\right)}, & x_1, x_2 > 0 \\ 0, & \text{其它} \end{cases} \tag{3.6}$$

(a) [2 分] 计算 $f_{X_2}(x_2)$.

【解】 当 $x_2 > 0$ 时，我们有

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\,\mathrm{d}x_1 = \int_0^{\infty} \frac{1}{8x_2} \mathrm{e}^{-\left(\frac{x_1}{2x_2}+\frac{x_2}{4}\right)}\,\mathrm{d}x_1$$

$$= \frac{1}{8x_2}\mathrm{e}^{-\frac{x_2}{4}} \int_0^{\infty} \mathrm{e}^{-\frac{x_1}{2x_2}}\,\mathrm{d}x_1 = \frac{1}{4}\mathrm{e}^{-\frac{x_2}{4}}$$

所以 X_2 的边际密度函数为

$$f_{X_2}(x_2) = \begin{cases} \dfrac{1}{4}\mathrm{e}^{-\frac{x_2}{4}}, & x_2 > 0 \\ 0, & \text{其它} \end{cases}$$

即 X_2 服从参数 $\theta = \dfrac{1}{4}$ 的指数分布.

(b) [2 分] 计算 $f(x_1 | x_2)$.

【解】 对于给定的 $x_2 > 0$，X_1 的条件密度为

$$f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)} = \begin{cases} \dfrac{\dfrac{1}{8x_2}\mathrm{e}^{-\left(\frac{x_1}{2x_2}+\frac{x_2}{4}\right)}}{\dfrac{1}{4}\mathrm{e}^{-\frac{x_2}{4}}}, & x_1 > 0 \\ 0, & x_1 \le 0 \end{cases}$$

$$= \begin{cases} \dfrac{1}{2x_2}\mathrm{e}^{-\frac{x_1}{2x_2}}, & x_1 > 0 \\ 0, & x_1 \le 0 \end{cases}$$

则当给定 $x_2 > 0$ 时，$(X_1 \mid X_2 = x_2)$ 服从参数 $\theta = \dfrac{1}{2x_2}$ 的指数分布.

(c) [2 分] 给出利用 X_2 的一个函数对 X_1 的最佳逼近.

【解】 在均方误差最小意义下，用 X_2 的一个函数对 X_1 的最佳逼近为 $\mathbb{E}(X_1 \mid X_2)$. 求得

$$\mathbb{E}(X_1 \mid X_2) = \int_{-\infty}^{\infty} x_1 \cdot f(x_1 \mid x_2)\,\mathrm{d}x_1 = \int_0^{\infty} x_1 \cdot \frac{1}{2x_2}\mathrm{e}^{-\frac{x_1}{2x_2}}\mathrm{d}x_1$$

$$= -\int_0^{\infty} x_1\,\mathrm{d}\left(\mathrm{e}^{-\frac{x_1}{2x_2}}\right) = \int_0^{\infty} \mathrm{e}^{-\frac{x_1}{2x_2}}\,\mathrm{d}x_1$$

$$= 2x_2$$

所以，用 X_2 的一个函数对 X_1 的最佳逼近为 $\mathbb{E}(X_1 \mid X_2) = 2X_2$.

(d) [2 分] 计算最佳逼近的误差的方差.

【解】 最佳逼近的误差为

$$U = X_1 - \mathbb{E}\left(X_1 \,|\, X_2\right) = X_1 - 2X_2$$

由重期望公式

$$\mathbb{E}\left(X_1\right) = \mathbb{E}\left[\mathbb{E}\left(X_1 \,|\, X_2\right)\right]$$

可知 $\mathbb{E}(U) = 0$，于是，最佳逼近的误差的方差为

$$\begin{aligned}\mathbb{V}\mathrm{ar}(U) &= \mathbb{V}\mathrm{ar}\left(X_1 - 2X_2\right) = \mathbb{E}\left[\left(X_1 - 2X_2\right)^2\right] \\ &= \mathbb{E}\left(X_1^2 - 4X_1X_2 + 4X_2^2\right) = \mathbb{E}\left(X_1^2\right) - 4\mathbb{E}\left(X_1X_2\right) + 4\mathbb{E}\left(X_2^2\right)\end{aligned}$$

由已知 X_2 服从参数 $\theta = \dfrac{1}{4}$ 的指数分布可得，

$$\mathbb{E}\left(X_2\right) = 4\,, \quad \mathbb{V}\mathrm{ar}\left(X_2\right) = 4^2 = 16 \implies \mathbb{E}\left(X_2^2\right) = \mathbb{V}\mathrm{ar}\left(X_2\right) + \left[\mathbb{E}\left(X_2\right)\right]^2 = 32$$

由已知给定 $x_2 > 0$ 时，$\left(X_1 \,|\, X_2 = x_2\right)$ 服从参数 $\theta = \dfrac{1}{2x_2}$ 的指数分布可得，

$$\mathbb{E}\left(X_1 \,|\, X_2\right) = 2x_2\,, \quad \mathbb{V}\mathrm{ar}\left(X_1 \,|\, X_2\right) = 4x_2^2$$

由条件期望的性质可知

$$\begin{aligned}\mathbb{E}\left(X_1\right) &= \mathbb{E}\left[\mathbb{E}\left(X_1 \,|\, X_2\right)\right] = \mathbb{E}\left(2X_2\right) = 2\mathbb{E}\left(X_2\right) = 8 \\ \mathbb{V}\mathrm{ar}\left(X_1\right) &= \mathbb{E}\left\{\mathbb{V}\mathrm{ar}\left(X_1 \,|\, X_2\right)\right\} + \mathbb{V}\mathrm{ar}\left\{\mathbb{E}\left(X_1 \,|\, X_2\right)\right\} \\ &= \mathbb{E}\left(4X_2^2\right) + \mathbb{V}\mathrm{ar}\left(2X_2\right) = 4\mathbb{E}\left(X_2^2\right) + 4\mathbb{V}\mathrm{ar}\left(X_2\right) = 192 \\ \implies \quad \mathbb{E}\left(X_1^2\right) &= \mathbb{V}\mathrm{ar}\left(X_1\right) + \left[\mathbb{E}\left(X_1\right)\right]^2 = 256\end{aligned}$$

再计算得

$$\begin{aligned}\mathbb{E}\left(X_1X_2\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2\, f\left(x_1\,,\, x_2\right)\, \mathrm{d}x_1\mathrm{d}x_2 \\ &= \int_0^{\infty}\int_0^{\infty} x_1x_2 \cdot \frac{1}{8x_2}\, \mathrm{e}^{-\left(\frac{x_1}{2x_2} + \frac{x_2}{4}\right)}\, \mathrm{d}x_1\mathrm{d}x_2 \\ &= \frac{1}{8}\int_0^{\infty} \mathrm{e}^{-\frac{x_2}{4}} \left(\int_0^{\infty} x_1 \cdot \mathrm{e}^{-\frac{x_1}{2x_2}}\, \mathrm{d}x_1\right) \mathrm{d}x_2 \\ &= \frac{1}{2}\int_0^{\infty} x_2^2\, \mathrm{e}^{-\frac{x_2}{4}}\, \mathrm{d}x_2 = 64\end{aligned}$$

故，最佳逼近的误差的方差为

$$\begin{aligned}\mathbb{V}\mathrm{ar}(U) &= \mathbb{E}\left(X_1^2\right) - 4\mathbb{E}\left(X_1X_2\right) + 4\mathbb{E}\left(X_2^2\right) \\ &= 256 - 4\times 64 + 4\times 32 = 128\end{aligned}$$

Chapter 4

第 4 周作业参考答案

1. [2 分] 证明

$$f_{\boldsymbol{Y}}(\boldsymbol{y})=\begin{cases}\dfrac{1}{2}y_1-\dfrac{1}{4}y_2\,, & 0\le y_1\le 2,\ |y_2|\le 1-|1-y_1|\ ,\\ 0\,, & \text{其它}\end{cases} \tag{4.1}$$

是一个概率密度函数.

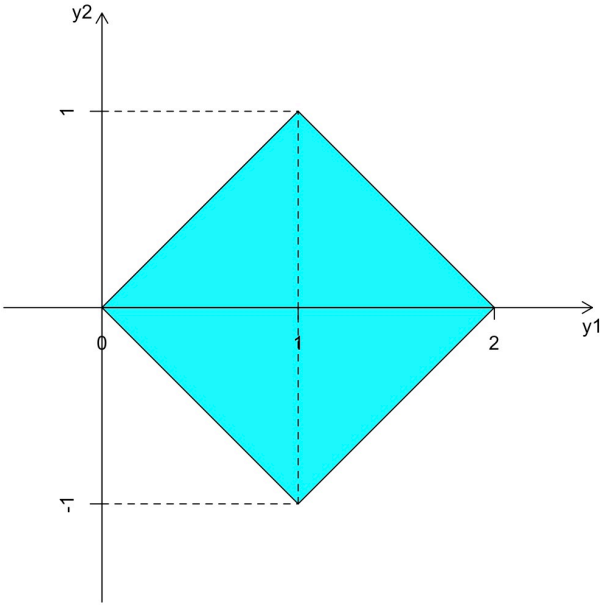

图 4.1: 第 1 题图.

【证明】 函数 $f_{\boldsymbol{Y}}(\boldsymbol{y})$ 非零值的定义域如图4.1所示，记该区域为

$$D=\{(y_1,\,y_2):\,0\le y_1\le 2,\,|y_2|\le 1-|1-y_1|\}$$

于是，当 $0\le y_1\le 1$ 时，

$$|y_2|\le 1-|1-y_1|=y_1 \quad\Longrightarrow\quad y_1-y_2\ge 0 \quad\Longrightarrow\quad f_{\boldsymbol{Y}}(\boldsymbol{y})=\frac{1}{2}y_1-\frac{1}{4}y_2\ge 0$$

当 $1 \leq y_1 \leq 2$ 时，

$$|y_2| \leq 1 - |1 - y_1| = 2 - y_1 \leq 1 \quad \Longrightarrow \quad f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{2}\, y_1 - \frac{1}{4}\, y_2 \geq 0$$

又因为

$$\begin{aligned}
\iint\limits_{-\infty}^{+\infty} f_{\boldsymbol{Y}}(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y} &= \iint\limits_{D} \left(\frac{1}{2}\, y_1 - \frac{1}{4}\, y_2\right) \mathrm{d}y_1 \mathrm{d}y_2 \\
&= \int_0^1 \mathrm{d}y_1 \left[\int_{-y_1}^{y_1} \left(\frac{1}{2}\, y_1 - \frac{1}{4}\, y_2\right) \mathrm{d}y_2\right] + \int_1^2 \mathrm{d}y_1 \left[\int_{y_1-2}^{2-y_1} \left(\frac{1}{2}\, y_1 - \frac{1}{4}\, y_2\right) \mathrm{d}y_2\right] \\
&= \frac{1}{3} + \frac{2}{3} \\
&= 1
\end{aligned}$$

故 $f_{\boldsymbol{Y}}(\boldsymbol{y})$ 是一个概率密度函数.

2. 设 $\boldsymbol{X} = (X_1,\, X_2)^{\mathrm{T}}$ 的概率密度函数为

$$f\left(x_1\,,\, x_2\right) = \begin{cases} 4\, x_1 x_2\, \mathrm{e}^{-x_1^2}\,, & x_1 > 0\,,\, 0 < x_2 < 1\,, \\ 0\,, & \text{其它.} \end{cases} \tag{4.2}$$

(a) [2 分] 计算 $\mathbb{E}(\boldsymbol{X})$ 与 $\mathbb{V}\mathrm{ar}(\boldsymbol{X})$.

【解】 记

$$f_{X_1}\left(x_1\right) = \begin{cases} 2\, x_1\, \mathrm{e}^{-x_1^2}\,, & x_1 > 0 \\ 0\,, & \text{其它}, \end{cases}$$

$$f_{X_2}\left(x_2\right) = \begin{cases} 2\, x_2\,, & 0 < x_2 < 1 \\ 0\,, & \text{其它.} \end{cases}$$

因为

$$f\left(x_1\,,\, x_2\right) = f_{X_1}\left(x_1\right) \cdot f_{X_2}\left(x_2\right)$$

所以 $f_{X_1}\left(x_1\right)$ 是 X_1 的边际概率密度函数，$f_{X_2}\left(x_2\right)$ 是 X_2 的边际概率密度函数，且 X_1 与 X_2

相互独立. 于是

$$\mathbb{E}(X_1) = \int_{-\infty}^{\infty} x_1 \cdot f_{X_1}(x_1)\,\mathrm{d}x_1 = \int_0^{\infty} 2\,x_1^2\,\mathrm{e}^{-x_1^2}\,\mathrm{d}x_1 = \frac{\sqrt{\pi}}{2}$$

$$\mathbb{E}\left(X_1^2\right) = \int_{-\infty}^{\infty} x_1^2 \cdot f_{X_1}(x_1)\,\mathrm{d}x_1 = \int_0^{\infty} 2\,x_1^3\,\mathrm{e}^{-x_1^2}\,\mathrm{d}x_1 = 1$$

$$\mathbb{V}\mathrm{ar}(X_1) = \mathbb{E}\left(X_1^2\right) - \left[\mathbb{E}(X_1)\right]^2 = 1 - \frac{\pi}{4}$$

$$\mathbb{E}(X_2) = \int_{-\infty}^{\infty} x_2 \cdot f_{X_2}(x_2)\,\mathrm{d}x_2 = \int_0^{1} 2\,x_2^2\,\mathrm{d}x_2 = \frac{2}{3}$$

$$\mathbb{E}\left(X_2^2\right) = \int_{-\infty}^{\infty} x_2^2 \cdot f_{X_2}(x_2)\,\mathrm{d}x_2 = \int_0^{1} 2\,x_2^3\,\mathrm{d}x_2 = \frac{1}{2}$$

$$\mathbb{V}\mathrm{ar}(X_2) = \mathbb{E}\left(X_2^2\right) - \left[\mathbb{E}(X_2)\right]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$$

$$\mathbb{C}\mathrm{ov}(X_1,\, X_2) = 0$$

故

$$\mathbb{E}(\boldsymbol{X}) = \begin{pmatrix} \dfrac{\sqrt{\pi}}{2} \\ \\ \dfrac{2}{3} \end{pmatrix}, \qquad \mathbb{V}\mathrm{ar}(\boldsymbol{X}) = \begin{pmatrix} 1 - \dfrac{\pi}{4} & 0 \\ \\ 0 & \dfrac{1}{18} \end{pmatrix}.$$

(b) [2 分] 计算 $\mathbb{E}(X_1 \,|X_2)$ 与 $\mathbb{E}(X_2 \,|X_1)$.

【解】 因为 X_1 与 X_2 相互独立，所以

$$\mathbb{E}(X_1 \,|X_2) = \mathbb{E}(X_1) = \frac{\sqrt{\pi}}{2}, \qquad \mathbb{E}(X_2 \,|X_1) = \mathbb{E}(X_2) = \frac{2}{3}.$$

(c) [2 分] 计算 $\mathbb{V}\mathrm{ar}(X_1 \,|X_2)$ 与 $\mathbb{V}\mathrm{ar}(X_2 \,|X_1)$.

【解】 因为 X_1 与 X_2 相互独立，所以

$$\mathbb{V}\mathrm{ar}(X_1 \,|X_2) = \mathbb{V}\mathrm{ar}(X_1) = 1 - \frac{\pi}{4}, \qquad \mathbb{V}\mathrm{ar}(X_2 \,|X_1) = \mathbb{V}\mathrm{ar}(X_2) = \frac{1}{18}.$$

3. [2 分] 设 $\boldsymbol{X} = (X_1,\, X_2)^{\mathrm{T}}$ 的概率密度函数为

$$f(x_1\,,\, x_2) = \begin{cases} \dfrac{1}{2\pi}, & 0 < x_1 < 2\pi\,,\, 0 < x_2 < 1\,, \\ \\ 0\,, & \text{其它}. \end{cases} \tag{4.3}$$

令

$$\begin{cases} U_1 = (\sin X_1)\sqrt{-2\,\ln X_2} \\ \\ U_2 = (\cos X_1)\sqrt{-2\,\ln X_2} \end{cases} \tag{4.4}$$

求 $\boldsymbol{U}=(U_1,\,U_2)^{\mathrm{T}}$ 的概率密度函数 $g\,(u_1\,,\,u_2)$.

【解】 由 (4.4) 式可得

$$\frac{U_1}{U_2}=\tan X_1 \qquad \Longrightarrow \qquad X_1=\arctan\frac{U_1}{U_2}$$

$$U_1^2+U_2^2=-2\,\ln X_2 \qquad \Longrightarrow \qquad X_2=\exp\left\{-\frac{U_1^2+U_2^2}{2}\right\}$$

变换的 Jacobian 行列式为

$$J=\begin{vmatrix}\dfrac{\partial x_1}{\partial u_1} & \dfrac{\partial x_1}{\partial u_2}\\[2ex] \dfrac{\partial x_2}{\partial u_1} & \dfrac{\partial x_2}{\partial u_2}\end{vmatrix}=\begin{vmatrix}\dfrac{u_2}{u_1^2+u_2^2} & -\dfrac{u_1}{u_1^2+u_2^2}\\[2ex] -u_1\,\mathrm{e}^{-\frac{u_1^2+u_2^2}{2}} & -u_2\,\mathrm{e}^{-\frac{u_1^2+u_2^2}{2}}\end{vmatrix}=-\mathrm{e}^{-\frac{u_1^2+u_2^2}{2}}$$

于是，$\boldsymbol{U}=(U_1,\,U_2)^{\mathrm{T}}$ 的概率密度函数 $g\,(u_1\,,\,u_2)$ 为

$$g\,(u_1\,,\,u_2)=f\left(\arctan\frac{u_1}{u_2}\,,\,\mathrm{e}^{-\frac{u_1^2+u_2^2}{2}}\right)\cdot|J|=\frac{1}{2\pi}\,\mathrm{e}^{-\frac{u_1^2+u_2^2}{2}}\,,\quad -\infty<u_1,\,u_2<+\infty\,.$$

4. [2 分] 设 $\boldsymbol{X}\sim N_p(\boldsymbol{\mu},\,\Sigma)$，其概率密度函数为

$$f(\boldsymbol{x})=|2\pi\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}\,. \tag{4.5}$$

若 $\mathcal{A}$ 为 $p\times p$ 的非奇异矩阵，$\boldsymbol{c}\in\mathbb{R}^p$ 为常数向量. 证明：

$$\boldsymbol{Y}=\mathcal{A}\,\boldsymbol{X}+\boldsymbol{c}\sim N_p\left(\mathcal{A}\,\boldsymbol{\mu}+\boldsymbol{c},\,\mathcal{A}\,\Sigma\mathcal{A}^{\mathrm{T}}\right)\,. \tag{4.6}$$

【证明】 因为 $\boldsymbol{Y}=\mathcal{A}\,\boldsymbol{X}+\boldsymbol{c}$，其中 $\mathcal{A}$ 为 $p\times p$ 的非奇异矩阵，$\boldsymbol{c}\in\mathbb{R}^p$ 为常数向量，所以

$$\boldsymbol{X}=\mathcal{A}^{-1}\,(\boldsymbol{Y}-\boldsymbol{c})=\mathcal{A}^{-1}\boldsymbol{Y}-\mathcal{A}^{-1}\boldsymbol{c}$$

该变换的 Jacobian 矩阵为

$$J=\frac{\partial\boldsymbol{X}}{\partial\boldsymbol{Y}}=\mathcal{A}^{-1}$$

于是，$\boldsymbol{Y}$ 的概率密度函数为

$$\begin{aligned}
g_{\boldsymbol{Y}}(\boldsymbol{y}) &= f_{\boldsymbol{X}}\left(\mathcal{A}^{-1}\boldsymbol{y}-\mathcal{A}^{-1}\boldsymbol{c}\right)\cdot|\det(J)|\\
&= |2\pi\Sigma|^{-1/2}\cdot\exp\left\{-\frac{1}{2}\left(\mathcal{A}^{-1}\boldsymbol{y}-\mathcal{A}^{-1}\boldsymbol{c}-\boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathcal{A}^{-1}\boldsymbol{y}-\mathcal{A}^{-1}\boldsymbol{c}-\boldsymbol{\mu}\right)\right\}\cdot\left\|\mathcal{A}^{-1}\right\|\\
&= |2\pi\Sigma|^{-1/2}\cdot\left\|\mathcal{A}\right\|^{-1}\cdot\exp\left\{-\frac{1}{2}\left(\mathcal{A}^{-1}\left(\boldsymbol{y}-\mathcal{A}\boldsymbol{\mu}-\boldsymbol{c}\right)\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathcal{A}^{-1}\left(\boldsymbol{y}-\mathcal{A}\boldsymbol{\mu}-\boldsymbol{c}\right)\right)\right\}\\
&= |2\pi\Sigma|^{-1/2}\cdot\left\|\mathcal{A}\right\|^{-1/2}\cdot\left\|\mathcal{A}^{\mathrm{T}}\right\|^{-1/2}\cdot
\end{aligned}$$

$$\exp\left\{-\frac{1}{2}\left(\boldsymbol{y}-\mathcal{A}\boldsymbol{\mu}-\boldsymbol{c}\right)^{\mathrm{T}}\left(\mathcal{A}^{-1}\right)^{\mathrm{T}}\Sigma^{-1}\mathcal{A}^{-1}\left(\boldsymbol{y}-\mathcal{A}\boldsymbol{\mu}-\boldsymbol{c}\right)\right\}$$

$$=\left|2\pi\mathcal{A}\Sigma\mathcal{A}^{\mathrm{T}}\right|^{-1/2}\cdot\exp\left\{-\frac{1}{2}\left[\boldsymbol{y}-\left(\mathcal{A}\boldsymbol{\mu}+\boldsymbol{c}\right)\right]^{\mathrm{T}}\left(\mathcal{A}\Sigma\mathcal{A}^{\mathrm{T}}\right)^{-1}\left[\boldsymbol{y}-\left(\mathcal{A}\boldsymbol{\mu}+\boldsymbol{c}\right)\right]\right\}$$

这恰好是 $N_p\left(\mathcal{A}\boldsymbol{\mu}+\boldsymbol{c},\,\mathcal{A}\Sigma\mathcal{A}^{\mathrm{T}}\right)$ 的概率密度函数，所以

$$\boldsymbol{Y}=\mathcal{A}\,\boldsymbol{X}+\boldsymbol{c}\sim N_p\left(\mathcal{A}\,\boldsymbol{\mu}+\boldsymbol{c},\,\mathcal{A}\,\Sigma\mathcal{A}^{\mathrm{T}}\right).$$

5. 考虑矩不存在的 Cauchy 分布，从而中心极限定理 (CLT) 无法应用.

(a) [2 分] 取三个不同的样本容量 n，对来自 Cauchy 分布总体的样本均值 $\overline{\boldsymbol{x}}$ 进行模拟，作直方图以及相应的核密度曲线图. **提示：** Cauchy 分布可以通过 rcauchy(n, location = 0, scale = 1) 进行模拟.

【解】 用以下的 R 代码来进行模拟. 样本容量 $n=10$ 时的直方图以及相应的核密度曲线如图 4.2 所示. 样本容量 $n=100$ 时的直方图以及相应的核密度曲线如图 4.3 所示. 样本容量 $n=300$ 时的直方图以及相应的核密度曲线如图 4.4 所示.

```
rm(list = ls(all = TRUE)) # 清除当前所有变量与对象
graphics.off() # 清除当前所有图形
library(ggplot2)
Cauchy_Hist = function(n = 10, m = 100) {
  X = array(0, dim = m)
  for (i in 1:500) X[i] = mean(rcauchy(n, location = 0, scale = 1))
  simCauchy = data.frame(X = X)
  ggplot(simCauchy, aes(X)) +
    geom_histogram(binwidth = 1.06 * sd(X) / (500^(0.2))) +
    geom_density(color = 'red')
}
Cauchy_Hist()
Cauchy_Hist(n = 100, m = 300)
Cauchy_Hist(n = 300, m = 900)
```

(b) [2 分] 当 $n\to\infty$ 时，你预期会出现什么情况?.

【解】 由图4.2、4.3、4.4 可见，当 $n\to\infty$ 时，并未呈现出明确的规律性，这是因为 Cauchy 分布的各阶矩不存在造成的，此时中心极限定理 (CLT) 的条件不满足.

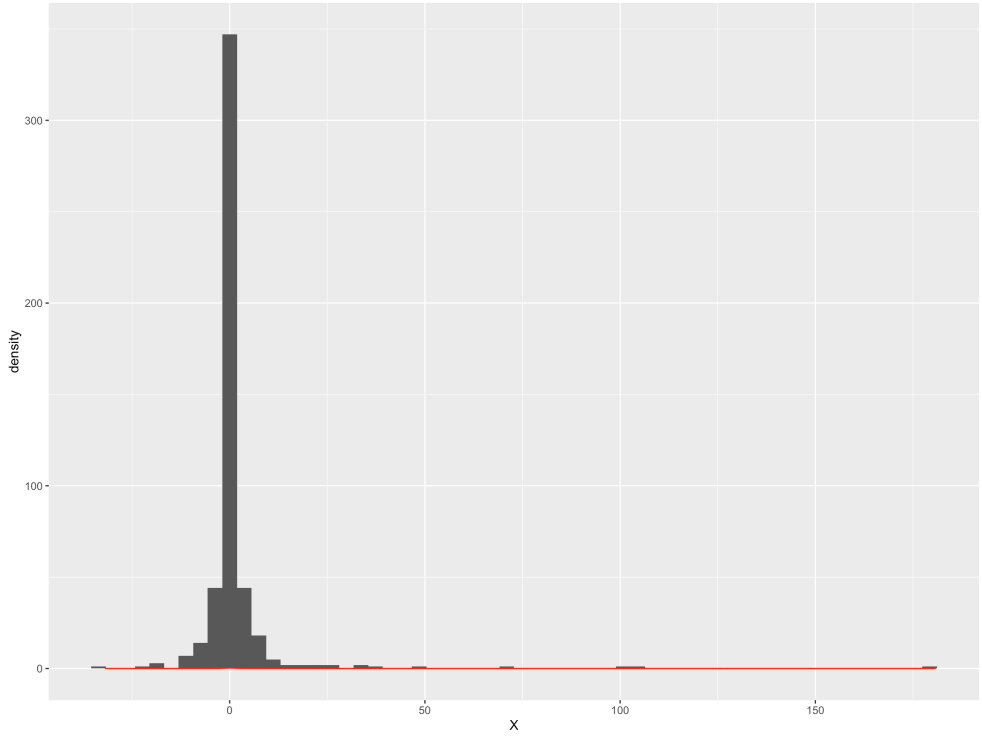

图 4.2: 样本容量 10 时的直方图与核密度曲线.

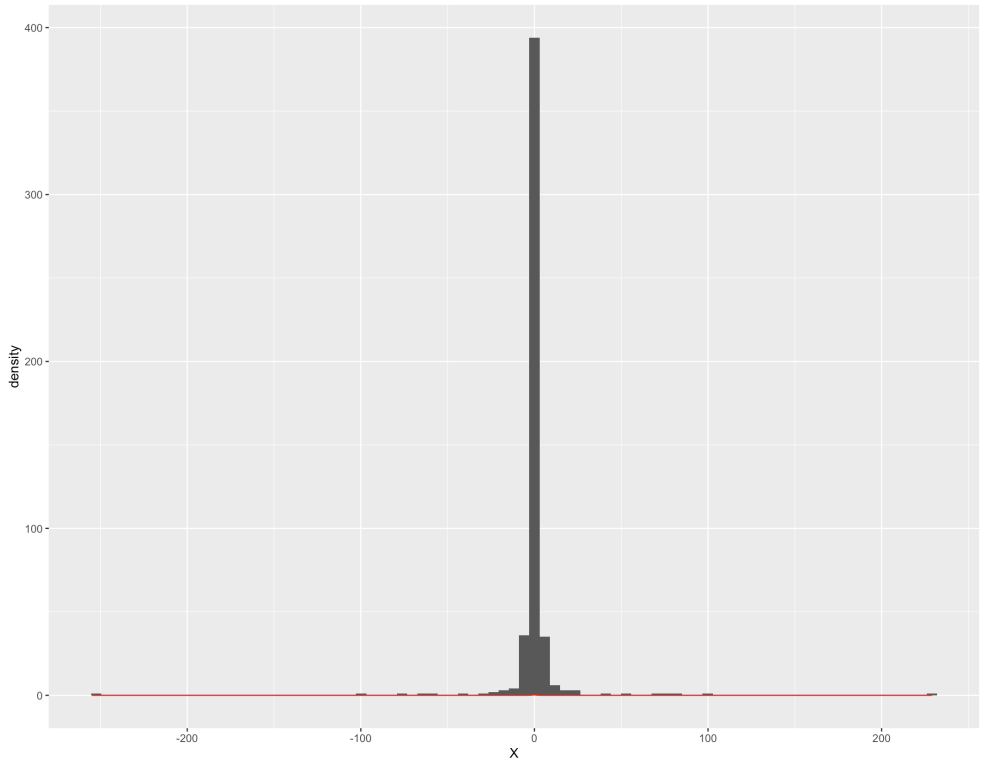

图 4.3: 样本容量 100 时的直方图与核密度曲线.

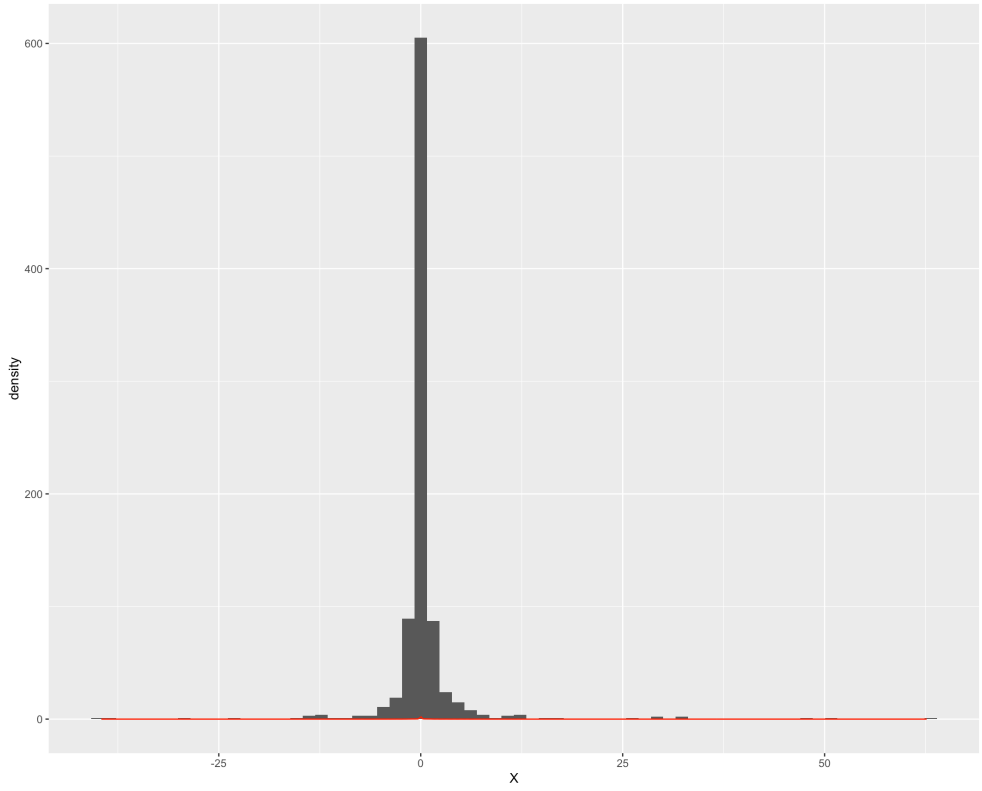

图 4.4: 样本容量 300 时的直方图与核密度曲线.

Chapter 5

第 5 周作业参考答案

1. [2 分] 假设 $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$，其中

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.1}$$

令

$$\mathcal{A} = (1, 1), \qquad \mathcal{B} = (1, -1) \tag{5.2}$$

证明 $\mathcal{A}\boldsymbol{X}$ 与 $\mathcal{B}\boldsymbol{X}$ 相互独立.

【证明】 因为

$$\mathbb{C}\mathrm{ov}(\mathcal{A}\boldsymbol{X}, \mathcal{B}\boldsymbol{X}) = \mathcal{A}\,\mathbb{C}\mathrm{ov}(\boldsymbol{X}, \boldsymbol{X})\,\mathcal{B}^{\mathrm{T}} = \mathcal{A}\,\mathbb{V}\mathrm{ar}(\boldsymbol{X})\,\mathcal{B}^{\mathrm{T}} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0$$

由教材 173 页的 Corollary 5.2 可知，$\mathcal{A}\boldsymbol{X}$ 与 $\mathcal{B}\boldsymbol{X}$ 相互独立.

2. 假设

$$\boldsymbol{X} \sim N_2\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \right), \quad (\boldsymbol{Y} \,|\boldsymbol{X}\,) \sim N_2\left(\begin{pmatrix} X_1 \\ X_1 + X_2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) \tag{5.3}$$

(a) [2 分] 确定 $Y_2 \,|\, Y_1$ 的分布.

【解】 因为

$$\begin{pmatrix} X_1 \\ X_1 + X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \mathcal{A}\boldsymbol{X}$$

所以 $\boldsymbol{Y} | \boldsymbol{X} \sim N_2(\mathcal{A}\boldsymbol{X}, \mathcal{I}_2)$，其协方差矩阵 $\mathcal{I}_2$ 与 $\boldsymbol{X}$ 无关，于是，由教材 175 页 Theorem 5.4

可知

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ Y_1 \\ Y_2 \end{pmatrix} \sim N_4(\boldsymbol{\mu},\ \Sigma)$$

其中

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \mathcal{A}\,\boldsymbol{\mu}_1 + \boldsymbol{b} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 3 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{11}\mathcal{A}^{\mathrm{T}} \\ \mathcal{A}\,\Sigma_{11} & \Omega + \mathcal{A}\,\Sigma_{11}\mathcal{A}^{\mathrm{T}} \end{pmatrix}$$

$$= \begin{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} & \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 1 & 2 & 3 \\ 1 & 2 & 1 & 3 \\ 2 & 1 & 3 & 3 \\ 3 & 3 & 3 & 7 \end{pmatrix}$$

从而，由教材 174 页 Theorem 5.2 可得

$$\boldsymbol{Y} = (\boldsymbol{0}\,,\, \mathcal{I}_2) \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 1 \\ 3 \end{pmatrix}\ ,\ \begin{pmatrix} 3 & 3 \\ 3 & 7 \end{pmatrix} \right)$$

最后，由教材 174 页 Theorem 5.3 我们有

$$(Y_2\,|Y_1\,) \sim N_1\left(3 + 3 \cdot (3)^{-1}\,(y_1 - 1)\ ,\ 7 - 3 \cdot (3)^{-1} \cdot 3\right) = N_1\,(y_1 + 2\ ,\ 4)\,.$$

(b) [2 分] 确定 $\boldsymbol{W} = \boldsymbol{X} - \boldsymbol{Y}$ 的分布.

【解】 由于

$$\boldsymbol{W} = \boldsymbol{X} - \boldsymbol{Y} = (\mathcal{I}_2\,,\,-\mathcal{I}_2)\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{pmatrix} = \begin{pmatrix}1 & 0 & -1 & 0\\ 0 & 1 & 0 & -1\end{pmatrix}\begin{pmatrix}X_1\\ X_2\\ Y_1\\ Y_2\end{pmatrix} = \begin{pmatrix}X_1 - Y_1\\ X_2 - Y_2\end{pmatrix}$$

根据教材 174 页 Theorem 5.2 我们有 $\boldsymbol{W} \sim N_2\,(\boldsymbol{\mu}^*\,,\,\Sigma^*)$，其中

$$\boldsymbol{\mu}^* = \begin{pmatrix}1 & 0 & -1 & 0\\ 0 & 1 & 0 & -1\end{pmatrix}\begin{pmatrix}1\\ 2\\ 1\\ 3\end{pmatrix} = \begin{pmatrix}0\\ -1\end{pmatrix}$$

$$\Sigma^* = \begin{pmatrix}1 & 0 & -1 & 0\\ 0 & 1 & 0 & -1\end{pmatrix}\begin{pmatrix}2 & 1 & 2 & 3\\ 1 & 2 & 1 & 3\\ 2 & 1 & 3 & 3\\ 3 & 3 & 3 & 7\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\\ -1 & 0\\ 0 & -1\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 3\end{pmatrix}$$

3. 假设

$$\begin{pmatrix}X\\ Y\\ Z\end{pmatrix} \sim N_3(\boldsymbol{\mu}\,,\,\Sigma) \tag{5.4}$$

若已知

$$Y \,|\, Z \sim N_1(-Z\,,\,1) \tag{5.5}$$

$$\mu_{Z|Y} = -\frac{1}{3} - \frac{1}{3}Y \tag{5.6}$$

$$(X \,|\, Y\,,\,Z) \sim N_1(2 + 2Y + 3Z\,,\,1) \tag{5.7}$$

(a) [2 分] 计算 $\boldsymbol{\mu}$ 和 Σ.

【解】 因为 $\begin{pmatrix}X\\ Y\\ Z\end{pmatrix} \sim N_3(\boldsymbol{\mu}\,,\,\Sigma)$，所以 $Z \sim N_1\,(\mu_z\,,\,\sigma_{zz})$. 由 (5.5) 和教材 175 页 Theorem 5.4 可知

$$\begin{pmatrix}Z\\ Y\end{pmatrix} \sim N_2\left(\begin{pmatrix}\mu_z\\ -\mu_z\end{pmatrix}\,,\,\begin{pmatrix}\sigma_{zz} & -\sigma_{zz}\\ -\sigma_{zz} & 1 + \sigma_{zz}\end{pmatrix}\right)$$

于是

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} Z \\ Y \end{pmatrix} \sim N_2 \left(\begin{pmatrix} -\mu_z \\ \mu_z \end{pmatrix}, \begin{pmatrix} 1+\sigma_{zz} & -\sigma_{zz} \\ -\sigma_{zz} & \sigma_{zz} \end{pmatrix} \right)$$

再根据教材 174 页 Theorem 5.3 可得

$$Z \,|\, Y \sim N_1 \left(\mu_z - \frac{\sigma_{zz}}{1+\sigma_{zz}} (y+\mu_z) \,,\, \sigma_{zz} - \frac{\sigma_{zz}^2}{1+\sigma_{zz}} \right)$$

结合 (5.6) 式，我们得到

$$\mathbb{E}(Z \,|\, Y) = \mu_z - \frac{\sigma_{zz}}{1+\sigma_{zz}} \mu_z - \frac{\sigma_{zz}}{1+\sigma_{zz}} Y = \mu_{Z|Y} = -\frac{1}{3} - \frac{1}{3} Y$$

从而有

$$-\frac{\sigma_{zz}}{1+\sigma_{zz}} = -\frac{1}{3}, \quad \mu_z - \frac{\sigma_{zz}}{1+\sigma_{zz}} \mu_z = -\frac{1}{3}$$

由此解得

$$\sigma_{zz} = \frac{1}{2}, \quad \mu_z = -\frac{1}{2}$$

因此有

$$\begin{pmatrix} Y \\ Z \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \right)$$

再由 (5.7) 式，其条件均值为

$$2 + 2Y + 3Z = (2\,,\,3) \begin{pmatrix} Y \\ Z \end{pmatrix} + 2$$

条件方差 1 与 Y, Z 无关，根据教材 175 页 Theorem 5.4 可得 $\begin{pmatrix} Y \\ Z \\ X \end{pmatrix} \sim N_3(\boldsymbol{\mu}^*, \Sigma^*)$. 其中

$$\boldsymbol{\mu}^* = \begin{pmatrix} \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \\ (2\,,\,3) \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} + 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{3}{2} \end{pmatrix}$$

$$\Sigma^* = \begin{pmatrix} \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} & \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ (2\,,3) \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} & 1 + (2\,,3) \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} & \frac{3}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & \frac{1}{2} & \frac{9}{2} \end{pmatrix}$$

因为

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} Y \\ Z \\ X \end{pmatrix}$$

故

$$\boldsymbol{\mu} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{3}{2} \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} & \frac{3}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & \frac{1}{2} & \frac{9}{2} \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} \frac{9}{2} & \frac{3}{2} & \frac{1}{2} \\ \frac{3}{2} & \frac{3}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

(b) [2 分] 确定 $X \,|\, Y$ 的分布.

【解】 因为

$$\begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

$$\sim N_2 \left(\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{9}{2} & \frac{3}{2} & \frac{1}{2} \\ \frac{3}{2} & \frac{3}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}^{\mathrm{T}} \right)$$

$$\sim N_2 \left(\begin{pmatrix} \frac{1}{2} \\ \frac{3}{2} \end{pmatrix}, \begin{pmatrix} \frac{3}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{9}{2} \end{pmatrix} \right)$$

由教材 174 页 Theorem 5.3 可知

$$X \,|\, Y \sim N_1\left(\frac{3}{2}+\frac{3}{2}\left(\frac{3}{2}\right)^{-1}\left(y-\frac{1}{2}\right),\, \frac{9}{2}-\frac{3}{2}\left(\frac{3}{2}\right)^{-1}\left(\frac{3}{2}\right)^{-1}\right) = N_1\left(1+y,\, 3\right)$$

(c) [2 分] 确定 $X \,|\, Y+Z$ 的分布.

【解】 因为

$$\begin{pmatrix} Y+Z \\ X \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

$$\sim N_2\left(\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix},\, \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{9}{2} & \frac{3}{2} & \frac{1}{2} \\ \frac{3}{2} & \frac{3}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}^{\mathrm{T}}\right)$$

$$\sim N_2\left(\begin{pmatrix} 0 \\ \frac{3}{2} \end{pmatrix},\, \begin{pmatrix} 1 & 2 \\ 2 & \frac{9}{2} \end{pmatrix}\right)$$

由教材 174 页 Theorem 5.3 可知

$$(X \,|\, Y+Z=t) \sim N_1\left(\frac{3}{2}+2\times(1)^{-1}(t-0),\, \frac{9}{2}-2\times(1)^{-1}\times 2\right) = N_1\left(\frac{3}{2}+2\,t,\, \frac{1}{2}\right)$$

4. 已知

$$Z \sim N_1(0\,,\,1) \tag{5.8}$$

$$Y \,|\, Z \sim N_1(1+Z\,,\,1) \tag{5.9}$$

$$(X \,|\, Y\,,\, Z) \sim N_1(1-Y\,,\,1) \tag{5.10}$$

(a) [2 分] 确定 $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ 的分布.

【解】 因为 $Z \sim N_1(0\,,\,1)$，$Y \,|\, Z \sim N_1(1+Z\,,\,1)$，条件分布的均值为 Z 的线性函数，且方差与 Z 无关，由教材 175 页 Theorem 5.4 可得

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

又因为 $(X\,|\,Y\,,\,Z)\sim N_1(1-Y\,,\,1)$，其均值

$$1-Y=\begin{pmatrix}0 & -1\end{pmatrix}\begin{pmatrix}Z\\Y\end{pmatrix}+1$$

是 $\begin{pmatrix}Z\\Y\end{pmatrix}$ 的仿射变换，而其方差与 $\begin{pmatrix}Z\\Y\end{pmatrix}$ 无关，于是，再根据教材 175 页 Theorem 5.4 可知

$$\begin{pmatrix}Z\\Y\\X\end{pmatrix}\sim N_3\left(\boldsymbol{\mu}_1\,,\;\Sigma_1\right)$$

其中

$$\boldsymbol{\mu}_1=\begin{pmatrix}0\\1\\\begin{pmatrix}0 & -1\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}+1\end{pmatrix}=\begin{pmatrix}0\\1\\0\end{pmatrix}$$

$$\Sigma_1=\begin{pmatrix}\begin{pmatrix}1 & 1\\1 & 2\end{pmatrix} & \begin{pmatrix}1 & 1\\1 & 2\end{pmatrix}\begin{pmatrix}0\\-1\end{pmatrix}\\(0\,,-1)\begin{pmatrix}1 & 1\\1 & 2\end{pmatrix} & 1+(0\,,-1)\begin{pmatrix}1 & 1\\1 & 2\end{pmatrix}\begin{pmatrix}0\\-1\end{pmatrix}\end{pmatrix}=\begin{pmatrix}1 & 1 & -1\\1 & 2 & -2\\-1 & -2 & 3\end{pmatrix}$$

故

$$\begin{pmatrix}X\\Y\\Z\end{pmatrix}=\begin{pmatrix}0 & 0 & 1\\0 & 1 & 0\\1 & 0 & 0\end{pmatrix}\begin{pmatrix}Z\\Y\\X\end{pmatrix}\sim N_3\left(\begin{pmatrix}0\\1\\0\end{pmatrix}\,,\;\begin{pmatrix}3 & -2 & -1\\-2 & 2 & 1\\-1 & 1 & 1\end{pmatrix}\right)$$

(b) [2 分] 确定 $(Y\,|\,X\,,\,Z)$ 的分布.

【解】 因为

$$\begin{pmatrix}X\\Z\\Y\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\0 & 0 & 1\\0 & 1 & 0\end{pmatrix}\begin{pmatrix}X\\Y\\Z\end{pmatrix}\sim N_3\left(\begin{pmatrix}0\\0\\1\end{pmatrix}\,,\;\begin{pmatrix}3 & -1 & -2\\-1 & 1 & 1\\-2 & 1 & 2\end{pmatrix}\right)$$

根据教材 174 页 Theorem 5.3 可得

$$Y\,|X,\ Z \sim N_1\left(1+(-2\,,\,1)\begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}^{-1}\left(\begin{pmatrix} x \\ z \end{pmatrix}-\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right)\ ,\right.$$
$$\left. 2-(-2\,,\,1)\begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}^{-1}\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right)$$
$$= N_1\left(1-\frac{x}{2}+\frac{z}{2}\ ,\ \frac{1}{2}\right)$$

(c) [2 分] 确定 $\begin{pmatrix} U \\ V \end{pmatrix}=\begin{pmatrix} 1+Z \\ 1-Y \end{pmatrix}$ 的分布.

【解】 因为

$$\begin{pmatrix} U \\ V \end{pmatrix}=\begin{pmatrix} 1+Z \\ 1-Y \end{pmatrix}=\begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}+\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

根据教材 174 页 Theorem 5.2 可得

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N_2\left(\boldsymbol{\mu}\ ,\ \Sigma\right)$$

其中

$$\boldsymbol{\mu}=\begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}+\begin{pmatrix} 1 \\ 1 \end{pmatrix}=\begin{pmatrix} 0 \\ -1 \end{pmatrix}+\begin{pmatrix} 1 \\ 1 \end{pmatrix}=\begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

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

故

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim N_2\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\ ,\ \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}\right)$$

(d) [2 分] 计算 $\mathbb{E}\,(Y\,|\,U=2)$.

【解】 因为 $U=1+Z$，所以 $U=2$ 的充分必要条件为 $Z=1$，于是我们有 $\mathbb{E}(Y\,|\,U=2)=\mathbb{E}(Y\,|\,Z=1)$. 又因为已知 $Y\,|\,Z \sim N_1(1+Z,\ 1)$，故

$$\mathbb{E}(Y\,|\,U=2)=\mathbb{E}(Y\,|\,Z=1)=1+1=2$$

5. 已知

$$\boldsymbol{X} \sim N_3\left(\begin{pmatrix}1\\2\\3\end{pmatrix},\ \begin{pmatrix}11 & -6 & 2\\ -6 & 10 & -4\\ 2 & -4 & 6\end{pmatrix}\right) \tag{5.11}$$

(a) [2 分] 利用 X_1 与 X_2 的一个线性函数，求 X_3 的最佳线性逼近.

【解】 我们要用到

> 定理 4.3：假设 $\boldsymbol{X}_1 \in \mathbb{R}^k$，$\boldsymbol{X}_2 \in \mathbb{R}^{p-k}$，$\boldsymbol{U} = \boldsymbol{X}_2 - \mathbb{E}(\boldsymbol{X}_2 \,|\, \boldsymbol{X}_1)$，则
>
> i. $\mathbb{E}(\boldsymbol{U}) = \boldsymbol{0}$.
>
> ii. 利用 $\boldsymbol{X}_1$ 的一个函数 $h(\boldsymbol{X}_1)$ 对 $\boldsymbol{X}_2$ 的最佳逼近为 $\mathbb{E}(\boldsymbol{X}_2 \,|\, \boldsymbol{X}_1)$，其中 $h: \mathbb{R}^k \longrightarrow \mathbb{R}^{p-k}$，所谓最佳是指在最小均方误差 (MSE) 意义下的最佳，其中
>
> $$\mathrm{MSE}(h) = \mathbb{E}\left\{[\boldsymbol{X}_2 - h(\boldsymbol{X}_1)]^{\mathrm{T}}\,[\boldsymbol{X}_2 - h(\boldsymbol{X}_1)]\right\}$$

在正态分布理论中 (教材 176 页)，如果

$$\begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix} \sim N_p\left(\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\ \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right)$$

由 $\boldsymbol{X}_1 \in \mathbb{R}^r$ 对 $\boldsymbol{X}_2 \in \mathbb{R}^{p-r}$ 的最佳线性逼近可以表示为

$$\boldsymbol{X}_2 = \boldsymbol{\beta}_0 + \mathcal{B}\boldsymbol{X}_1 + \boldsymbol{U}$$

其中

$$\mathcal{B} = \Sigma_{21}\Sigma_{11}^{-1}\,,\quad \boldsymbol{\beta}_0 = \boldsymbol{\mu}_2 - \mathcal{B}\boldsymbol{\mu}_1\,,\quad \boldsymbol{U} \sim N_{p-r}(\boldsymbol{0}\,,\ \Sigma_{22\cdot 1})$$

当 $p - r = 1$ 时，$\boldsymbol{X}_2$ 与 $\boldsymbol{X}_1$ 的多重相关系数的平方则为

$$\rho^2_{2\cdot 1\ldots r} = \frac{\boldsymbol{\sigma}_{21}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{12}}{\boldsymbol{\sigma}_{22}}$$

本题中，因为

$$\boldsymbol{X} = \begin{pmatrix}X_1\\ X_2\\ X_3\end{pmatrix} \sim N_3\left(\begin{pmatrix}1\\2\\3\end{pmatrix},\ \begin{pmatrix}11 & -6 & 2\\ -6 & 10 & -4\\ 2 & -4 & 6\end{pmatrix}\right)$$

则

$$\mathcal{B} = \begin{pmatrix}2 & -4\end{pmatrix}\begin{pmatrix}11 & -6\\ -6 & 10\end{pmatrix}^{-1} = \begin{pmatrix}-0.05405405 & -0.4324324\end{pmatrix}$$

$$\boldsymbol{\beta}_0 = 3 - \begin{pmatrix}-0.05405405 & -0.4324324\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix} = 3.918919$$

于是，由 X_1 与 X_2 的一个线性函数对 X_3 的最佳线性逼近为

$$X_3 = 3.918919 + \begin{pmatrix} -0.05405405 & -0.4324324 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$
$$= 3.918919 - 0.05405405X_1 - 0.4324324X_2$$

(b) [2 分] 计算 X_3 与 (X_1, X_2) 的多重相关系数.

【解】 X_3 与 (X_1, X_2) 的多重相关系数的平方为

$$\rho^2_{3\cdot 1\,2} = \frac{1}{6} \times \begin{pmatrix} 2 & -4 \end{pmatrix} \begin{pmatrix} 11 & -6 \\ -6 & 10 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ -4 \end{pmatrix} = 0.2702703$$

(c) [2 分] 令 $Z_1 = X_2 - X_3$, $Z_2 = X_2 + X_3$, 如果 $(Z_3 \,|\, Z_1, Z_2) \sim N_1(Z_1 + Z_2, 10)$, 确定 $\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}$ 的分布.

【解】 因为

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} X_2 - X_3 \\ X_2 + X_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

而

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

$$\begin{pmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 24 & 4 \\ 4 & 8 \end{pmatrix}$$

所以

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} -1 \\ 5 \end{pmatrix}, \begin{pmatrix} 24 & 4 \\ 4 & 8 \end{pmatrix} \right)$$

又因为条件分布

$$(Z_3 | \, Z_1, \, Z_2) \sim N_1(Z_1 + Z_2, \, 10) = N_1 \left(\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}, \, 10 \right)$$

其均值是 Z_1、Z_2 的线性函数，方差与 Z_1、Z_2 无关，根据教材 175 页 Theorem 5.4 可知

$$\begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} \sim N_3\left(\boldsymbol{\mu}\,,\ \Sigma\right)$$

其中

$$\boldsymbol{\mu} = \begin{pmatrix} -1 \\ 5 \\ \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 5 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \\ 4 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \begin{matrix} 24 & 4 \\ 4 & 8 \end{matrix} & \begin{pmatrix} 24 & 4 \\ 4 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 24 & 4 \\ 4 & 8 \end{pmatrix} & 10 + \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 24 & 4 \\ 4 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 24 & 4 & 28 \\ 4 & 8 & 12 \\ 28 & 12 & 40 \end{pmatrix}$$

6. 假设 $(X\,,\,Y\,,\,Z)^{\mathrm{T}}$ 服从三维正态分布，且

$$(Y \,|\, Z) \sim N_1(2Z\,,\,24) \tag{5.12}$$

$$(Z \,|\, X) \sim N_1(2X+3\,,\,14) \tag{5.13}$$

$$X \sim N_1(1\,,\,4) \tag{5.14}$$

$$\rho_{XY} = 0.5 \tag{5.15}$$

(a) [2 分] 确定 $(X\,,\,Y\,,\,Z)^{\mathrm{T}}$ 的分布.

【解】 因为 $X \sim N_1(1\,,\,4)$，$(Z \,|\, X) \sim N_1(2X+3\,,\,14)$，由教材 175 页 Theorem 5.4 可得

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim N_2\left(\begin{pmatrix} 1 \\ 2\times 1+3 \end{pmatrix}, \begin{pmatrix} 4 & 4\times 2 \\ 2\times 4 & 14+2\times 4\times 2 \end{pmatrix}\right) = N_2\left(\begin{pmatrix} 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 & 8 \\ 8 & 30 \end{pmatrix}\right)$$

由此可见

$$\mathbb{C}\mathrm{ov}(X\,,\,Z) = 8$$

又因为 $Z \sim N_1(5\,,\,30)$，$(Y \,|\, Z) \sim N_1(2Z\,,\,24)$，再由教材 175 页 Theorem 5.4 可得

$$\begin{pmatrix} Z \\ Y \end{pmatrix} \sim N_2\left(\begin{pmatrix} 5 \\ 2\times 5 \end{pmatrix}, \begin{pmatrix} 30 & 30\times 2 \\ 2\times 30 & 24+2\times 30\times 2 \end{pmatrix}\right) = N_2\left(\begin{pmatrix} 5 \\ 10 \end{pmatrix}, \begin{pmatrix} 30 & 60 \\ 60 & 144 \end{pmatrix}\right)$$

由此可见

$$\mathbb{C}\mathrm{ov}(Y\,,\,Z)=60\,,\quad Y\sim N_1(10\,,\,144)$$

再由

$$\frac{1}{2}=\rho_{XY}=\frac{\mathbb{C}\mathrm{ov}(X\,,\,Y)}{\sqrt{\mathbb{D}(X)}\,\sqrt{\mathbb{D}(Y)}}=\frac{\mathbb{C}\mathrm{ov}(X\,,\,Y)}{\sqrt{4}\,\sqrt{144}}$$

得出 $\mathbb{C}\mathrm{ov}(X\,,\,Y)=12$，故有

$$\begin{pmatrix}X\\Y\\Z\end{pmatrix}\sim N_3\left(\begin{pmatrix}1\\10\\5\end{pmatrix}\,,\,\begin{pmatrix}4&12&8\\12&144&60\\8&60&30\end{pmatrix}\right)$$

(b) [2 分] 对于给定的 Z 值，计算 X 与 Y 的偏相关系数.

【解】 由教材 174 页 Theorem 5.3 可知

$$\begin{aligned}\begin{pmatrix}X\\Y\end{pmatrix}\,\Big|\,Z&\sim N_2\left(\begin{pmatrix}1\\10\end{pmatrix}+\begin{pmatrix}8\\60\end{pmatrix}(30)^{-1}(z-5)\,,\,\begin{pmatrix}4&12\\12&144\end{pmatrix}-\begin{pmatrix}8\\60\end{pmatrix}(30)^{-1}(8\,,\,60)\right)\\&\sim N_2\left(\begin{pmatrix}\frac{7}{3}+\frac{4}{15}\,z\\2\,z\end{pmatrix}\,,\,\begin{pmatrix}\frac{28}{5}&-4\\-4&24\end{pmatrix}\right)\end{aligned}$$

于是，对于给定的 Z 值，X 与 Y 的偏相关系数为

$$\frac{-4}{\sqrt{\frac{28}{5}\times 24}}=-\sqrt{\frac{5}{42}}\approx -0.3450328$$

(c) [2 分] 你认为利用 Y 和 Z 的一个线性函数逼近 X 是否合理?

【解】 在正态分布理论中 (教材 176 页)，如果

$$\begin{pmatrix}\boldsymbol{X}_1\\\boldsymbol{X}_2\end{pmatrix}\sim N_p\left(\begin{pmatrix}\boldsymbol{\mu}_1\\\boldsymbol{\mu}_2\end{pmatrix}\,,\,\begin{pmatrix}\Sigma_{11}&\Sigma_{12}\\\Sigma_{21}&\Sigma_{22}\end{pmatrix}\right)$$

由 $\boldsymbol{X}_1\in\mathbb{R}^r$ 对 $\boldsymbol{X}_2\in\mathbb{R}^{p-r}$ 的最佳线性逼近可以表示为

$$\boldsymbol{X}_2=\boldsymbol{\beta}_0+\mathcal{B}\boldsymbol{X}_1+\boldsymbol{U}$$

其中

$$\mathcal{B}=\Sigma_{21}\Sigma_{11}^{-1}\,,\quad \boldsymbol{\beta}_0=\boldsymbol{\mu}_2-\mathcal{B}\boldsymbol{\mu}_1\,,\quad \boldsymbol{U}\sim N_{p-r}\left(\boldsymbol{0}\,,\,\Sigma_{22\cdot 1}\right)$$

当 $p-r=1$ 时，X_2 与 $\boldsymbol{X}_1$ 的多重相关系数的平方则为

$$\rho^2_{2\cdot 1\ldots r}=\frac{\boldsymbol{\sigma}_{21}\Sigma_{11}^{-1}\boldsymbol{\sigma}_{12}}{\boldsymbol{\sigma}_{22}}$$

本题中，因为

$$\begin{pmatrix}Y\\Z\\X\end{pmatrix}=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}\begin{pmatrix}X\\Y\\Z\end{pmatrix}\sim N_3\left(\begin{pmatrix}10\\5\\1\end{pmatrix},\begin{pmatrix}144&60&12\\60&30&8\\12&8&4\end{pmatrix}\right)$$

于是，X 与 Y,Z 的多重相关系数的平方为

$$\rho^2_{X\cdot YZ}=\frac{1}{4}\,(12\,,\,8)\begin{pmatrix}144&60\\60&30\end{pmatrix}^{-1}\begin{pmatrix}12\\8\end{pmatrix}=0.7$$

说明 X 与 Y,Z 的线性相关程度很强，所以用 Y 和 Z 的一个线性函数逼近 X 是合理的.

7. 设

$$\boldsymbol{X}\sim N_4\left(\begin{pmatrix}1\\2\\3\\4\end{pmatrix},\begin{pmatrix}4&1&2&4\\1&4&2&1\\2&2&16&1\\4&1&1&9\end{pmatrix}\right)\tag{5.16}$$

(a) [2 分] 给出用 $(X_1\,,\,X_4)$ 的一个函数对 X_2 的最佳线性逼近，并解释逼近的效果.

【解】 因为

$$\begin{pmatrix}X_1\\X_4\\X_2\end{pmatrix}=\begin{pmatrix}1&0&0&0\\0&0&0&1\\0&1&0&0\end{pmatrix}\begin{pmatrix}X_1\\X_2\\X_3\\X_4\end{pmatrix}=\mathcal{A}\boldsymbol{X}$$

所以

$$\begin{pmatrix}X_1\\X_4\\X_2\end{pmatrix}\sim N_3\left(\begin{pmatrix}1\\4\\2\end{pmatrix},\begin{pmatrix}4&4&1\\4&9&1\\1&1&4\end{pmatrix}\right)$$

利用条件逼近的结论，我们知道

$$X_2=\beta_0+\mathcal{B}\begin{pmatrix}X_1\\X_4\end{pmatrix}$$

其中

$$\mathcal{B} = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 4 & 4 \\ 4 & 9 \end{pmatrix}^{-1} = \begin{pmatrix} 0.25 & 0 \end{pmatrix}$$

$$\beta_0 = 2 - \begin{pmatrix} 0.25 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \end{pmatrix} = 1.75$$

于是，用 (X_1, X_4) 的一个函数对 X_2 的最佳线性逼近为

$$X_2 = 1.75 + 0.25X_1$$

X_2 与 (X_1, X_4) 的多重相关系数的平方为

$$\rho^2_{2\cdot 1\,4} = \frac{1}{4} \times \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 4 & 4 \\ 4 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 0.0625$$

(b) [2 分] 给出用 (X_1, X_3, X_4) 的一个函数对 X_2 的最佳线性逼近，与 (a) 的结果进行对比.

【解】 因为

$$\begin{pmatrix} X_1 \\ X_3 \\ X_4 \\ X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} = \mathcal{A}\boldsymbol{X}$$

所以

$$\begin{pmatrix} X_1 \\ X_3 \\ X_4 \\ X_2 \end{pmatrix} \sim N_4 \left(\begin{pmatrix} 1 \\ 3 \\ 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 2 & 4 & 1 \\ 2 & 16 & 1 & 2 \\ 4 & 1 & 9 & 1 \\ 1 & 2 & 1 & 4 \end{pmatrix} \right)$$

利用条件逼近的结论，我们知道

$$X_2 = \beta_0 + \mathcal{B} \begin{pmatrix} X_1 \\ X_3 \\ X_4 \end{pmatrix}$$

其中

$$\mathcal{B}=\begin{pmatrix}1 & 2 & 1\end{pmatrix}\begin{pmatrix}4 & 2 & 4\\ 2 & 16 & 1\\ 4 & 1 & 9\end{pmatrix}^{-1}=\begin{pmatrix}0.1790541 & 0.1013514 & 0.02027027\end{pmatrix}$$

$$\beta_0=2-\begin{pmatrix}0.1790541 & 0.1013514 & 0.02027027\end{pmatrix}\begin{pmatrix}1\\ 3\\ 4\end{pmatrix}=1.435811$$

于是，用 $(X_1\,,X_3\,,X_4)$ 的一个函数对 X_2 的最佳线性逼近为

$$X_2=1.435811+0.1790541X_1+0.1013514X_3+0.02027027X_4$$

X_2 与 $(X_1,\ X_3,\ X_4)$ 的多重相关系数的平方为

$$\rho^2_{2\cdot 1\,3\,4}=\frac{1}{4}\times\begin{pmatrix}1 & 2 & 1\end{pmatrix}\begin{pmatrix}4 & 2 & 4\\ 2 & 16 & 1\\ 4 & 1 & 9\end{pmatrix}^{-1}\begin{pmatrix}1\\ 2\\ 1\end{pmatrix}=0.1005068$$

我们看到,在添加了一个变量 X_3 之后,多重相关系数的平方由此前的 0.0625 提高到了 0.1005068，说明用 $(X_1,\ X_3,\ X_4)$ 对 X_2 的逼近，相比仅用 $(X_1,\ X_4)$ 的逼近，效果会更好一些.

Chapter 6

第 6 周作业参考答案

1. 设有二维分布总体，其概率密度函数为

$$f\left(x_1,\ x_2\right)=\frac{1}{\theta_1\theta_2}\exp\left(-\frac{x_1}{\theta_1}-\frac{x_2}{\theta_2}\right)\ ,\quad x_1\ ,\ x_2>0 \tag{6.1}$$

从中抽取一个容量为 n 的简单随机样本.

(a) [2 分] 求 $\boldsymbol{\theta}=\left(\theta_1\ ,\ \theta_2\right)^{\mathrm{T}}$ 的极大似然估计量.

【解】 似然函数为

$$\begin{aligned}L\left(\boldsymbol{\theta},\,\boldsymbol{x}\right)&=f\left(x_{i1},\,x_{i2}\right)=\prod_{i=1}^{n}\frac{1}{\theta_1\theta_2}\exp\left\{-\frac{x_{i1}}{\theta_1}-\frac{x_{i2}}{\theta_2}\right\}\\&=\theta_1^{-n}\theta_2^{-n}\exp\left\{-\sum_{i=1}^{n}\frac{x_{i1}}{\theta_1}-\sum_{i=1}^{n}\frac{x_{i2}}{\theta_2}\right\}\end{aligned}$$

对数似然函数则为

$$\ell\left(\boldsymbol{\theta},\,\boldsymbol{x}\right)=\ln L\left(\boldsymbol{\theta},\,\boldsymbol{x}\right)=-n\ln\theta_1-n\ln\theta_2-\frac{1}{\theta_1}\sum_{i=1}^{n}x_{i1}-\frac{1}{\theta_2}\sum_{i=1}^{n}x_{i2}$$

因为

$$\begin{cases}\dfrac{\partial\ell\left(\boldsymbol{\theta},\,\boldsymbol{x}\right)}{\partial\theta_1}&=&-\dfrac{n}{\theta_1}+\dfrac{1}{\theta_1^2}\sum\limits_{i=1}^{n}x_{i1}=0\\[2ex]\dfrac{\partial\ell\left(\boldsymbol{\theta},\,\boldsymbol{x}\right)}{\partial\theta_2}&=&-\dfrac{n}{\theta_2}+\dfrac{1}{\theta_2^2}\sum\limits_{i=1}^{n}x_{i2}=0\end{cases}$$

由此解得

$$\widehat{\theta}_1=\frac{1}{n}\sum_{i=1}^{n}X_{i1}=\overline{X}_{[1]}\,,\quad\widehat{\theta}_2=\frac{1}{n}\sum_{i=1}^{n}X_{i2}=\overline{X}_{[2]}$$

即得 $\boldsymbol{\theta}$ 的极大似然估计量为

$$\widehat{\boldsymbol{\theta}}=\begin{pmatrix}\widehat{\theta}_1\\\widehat{\theta}_2\end{pmatrix}=\begin{pmatrix}\overline{X}_{[1]}\\\overline{X}_{[2]}\end{pmatrix}$$

(b) [2 分] 确定其 Cramer-Rao 下界.

【解】 因为二维总体 $\boldsymbol{X}=\begin{pmatrix}X_1\\X_2\end{pmatrix}$ 的联合概率密度函数为

$$\begin{aligned}f\left(x_1,\ x_2\right)&=\frac{1}{\theta_1\theta_2}\exp\left(-\frac{x_1}{\theta_1}-\frac{x_2}{\theta_2}\right),\qquad x_1\ ,\ x_2>0\\&=\left\{\frac{1}{\theta_1}\exp\left(-\frac{x_1}{\theta_1}\right)\cdot\mathrm{I}_{\{x_1\in(0,\infty)\}}\right\}\cdot\left\{\frac{1}{\theta_2}\exp\left(-\frac{x_2}{\theta_2}\right)\cdot\mathrm{I}_{\{x_2\in(0,\infty)\}}\right\}\end{aligned}$$

因此，X_1 与 X_2 相互独立，且均服从指数分布，指数分布的参数分别为 θ_1 与 θ_2. 于是，我们有

$$\mathbb{E}\left(X_1\right)=\theta_1\,,\ \mathbb{V}\mathrm{ar}\left(X_1\right)=\theta_1^2\,,\quad \mathbb{E}\left(X_2\right)=\theta_2\,,\ \mathbb{V}\mathrm{ar}\left(X_2\right)=\theta_2^2\,,\quad \mathbb{C}\mathrm{ov}\left(X_1\ ,\ X_2\right)=0$$

由于似然函数

$$\begin{aligned}L\left(\mathcal{X};\ \boldsymbol{\theta}\right)=\prod_{i=1}^{n}f\left(x_{i1},\ x_{i2}\right)&=\left\{\frac{1}{\theta_1^n}\exp\left(-\frac{1}{\theta_1}\sum_{i=1}^{n}x_{i1}\right)\cdot\mathrm{I}_{\{x_{11},\,x_{21},\,\ldots,\,x_{n1}\in(0,\infty)\}}\right\}\\&\quad\cdot\left\{\frac{1}{\theta_2^n}\exp\left(-\frac{1}{\theta_2}\sum_{i=1}^{n}x_{i2}\right)\cdot\mathrm{I}_{\{x_{12},\,x_{22},\,\ldots,\,x_{n2}\in(0,\infty)\}}\right\}\end{aligned}$$

对数似然函数

$$\begin{aligned}\ell\left(\mathcal{X};\ \boldsymbol{\theta}\right)=\ln L\left(\mathcal{X};\ \boldsymbol{\theta}\right)&=\left(-n\,\ln\theta_1-\frac{1}{\theta_1}\sum_{i=1}^{n}x_{i1}\right)\cdot\mathrm{I}_{\{x_{11},\,x_{21},\,\ldots,\,x_{n1}\in(0,\infty)\}}\\&\quad+\left(-n\,\ln\theta_2-\frac{1}{\theta_2}\sum_{i=1}^{n}x_{i2}\right)\cdot\mathrm{I}_{\{x_{12},\,x_{22},\,\ldots,\,x_{n2}\in(0,\infty)\}}\end{aligned}$$

评分函数

$$\boldsymbol{s}\left(\mathcal{X};\ \boldsymbol{\theta}\right)=\frac{\partial\,\ell\left(\mathcal{X};\ \boldsymbol{\theta}\right)}{\partial\,\boldsymbol{\theta}}=\begin{pmatrix}-\dfrac{n}{\theta_1}+\dfrac{1}{\theta_1^2}\sum\limits_{i=1}^{n}X_{i1}\\-\dfrac{n}{\theta_2}+\dfrac{1}{\theta_2^2}\sum\limits_{i=1}^{n}X_{i2}\end{pmatrix}=-n\begin{pmatrix}\dfrac{1}{\theta_1}\\\dfrac{1}{\theta_2}\end{pmatrix}+\sum_{i=1}^{n}\begin{pmatrix}\dfrac{1}{\theta_1^2}X_{i1}\\\dfrac{1}{\theta_2^2}X_{i2}\end{pmatrix}$$

于是，Fisher 信息矩阵

$$\mathcal{F}_n=\mathbb{V}\mathrm{ar}\left(\boldsymbol{s}\left(\mathcal{X};\ \boldsymbol{\theta}\right)\right)=\mathbb{V}\mathrm{ar}\left\{-n\begin{pmatrix}\dfrac{1}{\theta_1}\\\dfrac{1}{\theta_2}\end{pmatrix}+\sum_{i=1}^{n}\begin{pmatrix}\dfrac{1}{\theta_1^2}X_{i1}\\\dfrac{1}{\theta_2^2}X_{i2}\end{pmatrix}\right\}=\mathbb{V}\mathrm{ar}\left\{\sum_{i=1}^{n}\begin{pmatrix}\dfrac{1}{\theta_1^2}X_{i1}\\\dfrac{1}{\theta_2^2}X_{i2}\end{pmatrix}\right\}$$

因为 $\begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$，所以

$$\mathcal{F}_n = n\,\mathbb{V}\mathrm{ar}\left\{ \begin{pmatrix} \frac{1}{\theta_1^2} X_1 \\ \frac{1}{\theta_2^2} X_2 \end{pmatrix} \right\}$$

由于

$$\mathbb{V}\mathrm{ar}\left\{ \begin{pmatrix} \frac{1}{\theta_1^2} X_1 \\ \frac{1}{\theta_2^2} X_2 \end{pmatrix} \right\} = \mathbb{V}\mathrm{ar}\left\{ \begin{pmatrix} \frac{1}{\theta_1^2} & 0 \\ 0 & \frac{1}{\theta_2^2} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \right\} = \begin{pmatrix} \frac{1}{\theta_1^2} & 0 \\ 0 & \frac{1}{\theta_2^2} \end{pmatrix} \cdot \mathbb{V}\mathrm{ar} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\theta_1^2} & 0 \\ 0 & \frac{1}{\theta_2^2} \end{pmatrix}^{\mathrm{T}}$$

$$= \begin{pmatrix} \frac{1}{\theta_1^2} & 0 \\ 0 & \frac{1}{\theta_2^2} \end{pmatrix} \begin{pmatrix} \theta_1^2 & 0 \\ 0 & \theta_2^2 \end{pmatrix} \begin{pmatrix} \frac{1}{\theta_1^2} & 0 \\ 0 & \frac{1}{\theta_2^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\theta_1^2} & 0 \\ 0 & \frac{1}{\theta_2^2} \end{pmatrix}$$

因此

$$\mathcal{F}_n = \mathbb{V}\mathrm{ar}\left(\boldsymbol{s}\left(\mathcal{X};\ \boldsymbol{\theta}\right)\right) = n \begin{pmatrix} \frac{1}{\theta_1^2} & 0 \\ 0 & \frac{1}{\theta_2^2} \end{pmatrix}$$

所以，Cramer-Rao 下界为

$$\mathcal{F}_n^{-1} = \frac{1}{n} \begin{pmatrix} \theta_1^2 & 0 \\ 0 & \theta_2^2 \end{pmatrix}$$

(c) [2 分] 能否找到 $\boldsymbol{\theta}$ 的一个最小方差无偏估计量?

【解】 我们已经知道 $\boldsymbol{\theta}$ 的 MLE 为 $\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \overline{X}_{[1]} \\ \overline{X}_{[2]} \end{pmatrix}$，其中

$$\overline{X}_{[1]} = \frac{1}{n}\sum_{i=1}^{n} X_{i1}\ , \quad \overline{X}_{[2]} = \frac{1}{n}\sum_{i=1}^{n} X_{i2}$$

因为 X_1 与 X_2 相互独立，均服从指数分布表，参数分别为 θ_1 与 θ_2，所以

$$\mathbb{E}\left(\widehat{\boldsymbol{\theta}}\right) = \begin{pmatrix} \mathbb{E}\left(\overline{X}_{[1]}\right) \\ \mathbb{E}\left(\overline{X}_{[2]}\right) \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \boldsymbol{\theta}$$

说明 $\widehat{\boldsymbol{\theta}}$ 是 $\boldsymbol{\theta}$ 的一个无偏估计量. $\widehat{\boldsymbol{\theta}}$ 的协方差矩阵为

$$\mathbb{V}\mathrm{ar}\left(\widehat{\boldsymbol{\theta}}\right)=\mathbb{V}\mathrm{ar}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}X_{i1}\\ \frac{1}{n}\sum_{i=1}^{n}X_{i2}\end{pmatrix}=\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}X_{i1}\\ X_{i2}\end{pmatrix}\right]$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}\begin{pmatrix}X_{i1}\\ X_{i2}\end{pmatrix}=\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}\begin{pmatrix}\theta_1^2 & 0\\ 0 & \theta_2^2\end{pmatrix}=\frac{1}{n}\begin{pmatrix}\theta_1^2 & 0\\ 0 & \theta_2^2\end{pmatrix}=\mathcal{F}_n^{-1}$$

故，$\boldsymbol{\theta}$ 的 MLE 就是 $\boldsymbol{\theta}$ 的一个最小方差无偏估计量.

2. 考虑总体 $N_p\left(\boldsymbol{\mu},\,\Sigma_0\right)$，其中 Σ_0 已知，设 $\{\boldsymbol{x}_i\}_{i=1}^n$ 是取自该总体的一个简单随机样本.

(a) [2 分] 计算 $\boldsymbol{\mu}$ 的 Cramer-Rao 下界.

【解】 似然函数

$$L\left(\mathcal{X};\;\boldsymbol{\mu}\right)=\left|2\pi\Sigma_0\right|^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma_0^{-1}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)\right\}$$

于是，对数似然函数

$$\ell\left(\mathcal{X};\;\boldsymbol{\mu}\right)=\ln L\left(\mathcal{X};\;\boldsymbol{\mu}\right)=-\frac{n}{2}\ln\left(2\pi\Sigma_0\right)-\frac{1}{2}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma_0^{-1}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)$$

评分函数则为

$$\boldsymbol{s}\left(\mathcal{X};\;\boldsymbol{\mu}\right)=\frac{\partial}{\partial\boldsymbol{\mu}}\ell\left(\mathcal{X};\;\boldsymbol{\mu}\right)=\sum_{i=1}^{n}\Sigma_0^{-1}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)=\Sigma_0^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i-\boldsymbol{\mu}\right)$$

当 $n=1$ 时，我们有

$$\begin{aligned}\mathcal{F}_1&=\mathbb{V}\mathrm{ar}(\boldsymbol{s})=\mathbb{V}\mathrm{ar}\left[\Sigma_0^{-1}\left(\boldsymbol{x}_1-\boldsymbol{\mu}\right)\right]=\Sigma_0^{-1}\;\mathbb{V}\mathrm{ar}\left(\boldsymbol{x}_1-\boldsymbol{\mu}\right)\;\left(\Sigma_0^{-1}\right)^{\mathrm{T}}\\&=\Sigma_0^{-1}\;\mathbb{V}\mathrm{ar}\left(\boldsymbol{x}_1\right)\;\Sigma_0^{-1}=\Sigma_0^{-1}\;\Sigma_0\;\Sigma_0^{-1}\\&=\Sigma_0^{-1}\end{aligned}$$

在独立性假设下，可得 Fisher 信息矩阵如下

$$\mathcal{F}_n=n\mathcal{F}_1=n\Sigma_0^{-1}$$

故，$\boldsymbol{\mu}$ 的 Cramer-Rao 下界为

$$\mathcal{F}_n^{-1}=\left(n\Sigma_0^{-1}\right)^{-1}=\frac{1}{n}\,\Sigma_0$$

(b) [2 分] 能否给出 $\boldsymbol{\mu}$ 的一个最小方差无偏估计量?

【解】 取样本均值向量

$$\overline{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i$$

因为

$$\mathbb{E}\left(\overline{\boldsymbol{X}}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\boldsymbol{x}_i\right) = \boldsymbol{\mu}$$

所以 $\overline{\boldsymbol{X}}$ 是 $\boldsymbol{\mu}$ 的一个无偏估计量. 根据正态分布的理论，我们知道 $\overline{\boldsymbol{X}}$ 的协方差矩阵为

$$\mathbb{V}\mathrm{ar}\left(\overline{\boldsymbol{X}}\right) = \frac{1}{n}\,\Sigma_0$$

恰好等于 $\boldsymbol{\mu}$ 的 Cramer-Rao 下界，所以，$\overline{\boldsymbol{X}}$ 就是 $\boldsymbol{\mu}$ 的一个最小方差无偏估计量.

3. 假设 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$，其中 Σ 未知，但我们已知 $\Sigma = \mathrm{diag}\left(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{pp}\right)$，如果 $\{\boldsymbol{x}_i\}_{i=1}^{n}$ 是取自该总体的容量为 n 的一个简单随机样本

(a) [2 分] 求 $\boldsymbol{\mu}$ 的极大似然估计.

【解】 似然函数

$$L\left(\mathcal{X}; \boldsymbol{\mu}, \Sigma\right) = \left|2\,\pi\,\Sigma\right|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\right\}$$

于是有对数似然函数

$$\begin{aligned}
\ell\left(\mathcal{X}; \boldsymbol{\mu}, \Sigma\right) &= \ln L\left(\mathcal{X}; \boldsymbol{\mu}, \Sigma\right) \\
&= -\frac{n}{2}\ln\left|2\,\pi\,\Sigma\right| - \frac{1}{2}\sum_{i=1}^{n}\left(\boldsymbol{x}_i - \boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\boldsymbol{x} - \boldsymbol{\mu}\right) \\
&\quad \downarrow \quad \text{见教材 191-192 页的推导过程，亦可参见课堂讲义相关内容} \\
&= -\frac{n}{2}\ln\left|2\,\pi\,\Sigma\right| - \frac{n}{2}\,\mathrm{tr}\left(\Sigma^{-1}\,\mathcal{S}\right) - \frac{n}{2}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right) \qquad (6.2)
\end{aligned}$$

由于 Σ 是正态总体的协方差矩阵，所以它是非负定矩阵，从而 Σ^{-1} 也是非负定矩阵，即有

$$\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right) \geq 0$$

因此，为使对数似然函数 $\ell\left(\mathcal{X}; \boldsymbol{\mu}, \Sigma\right)$ 最大，则 (6.2) 式中的

$$\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\overline{\boldsymbol{x}} - \boldsymbol{\mu}\right) = 0$$

由此即得 $\boldsymbol{\mu}$ 的极大似然估计量为

$$\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i$$

(b) [2 分] 求 Σ 的极大似然估计.

【解】 将 $\boldsymbol{\mu}$ 的极大似然估计 $\widehat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$ 代入 (6.2) 式，此时对数似然函数为

$$\ell\left(\mathcal{X}\,;\,\boldsymbol{\mu}\,,\,\Sigma\right) = -\frac{n}{2}\,\ln\left|2\,\pi\,\Sigma\right| - \frac{n}{2}\,\mathrm{tr}\left(\Sigma^{-1}\,\mathcal{S}\right)$$

其中

$$\mathcal{S} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)^{\mathrm{T}} \triangleq \left(s_{ij}\right)_{p\times p}\,, \quad s_{ij} = \frac{1}{n}\sum_{k=1}^{n}\left(x_{ki} - \overline{x}_i\right)\left(x_{kj} - \overline{x}_j\right)$$

因为

$$\Sigma = \mathrm{diag}\left(\sigma_{11},\,\sigma_{22},\,\ldots,\,\sigma_{pp}\right) = \begin{pmatrix} \sigma_{11} & & & \\ & \sigma_{22} & & \\ & & \ddots & \\ & & & \sigma_{pp} \end{pmatrix}$$

所以 Σ 的逆矩阵

$$\Sigma^{-1} = \begin{pmatrix} \dfrac{1}{\sigma_{11}} & & & \\ & \dfrac{1}{\sigma_{22}} & & \\ & & \ddots & \\ & & & \dfrac{1}{\sigma_{pp}} \end{pmatrix} = \mathrm{diag}\left(\sigma_{11}^{-1}\,,\,\sigma_{22}^{-1}\,,\,\ldots\,,\,\sigma_{pp}^{-1}\right)$$

于是，对数似然函数可以表示为

$$\begin{aligned}
\ell\left(\mathcal{X}\,;\,\boldsymbol{\mu}\,,\,\Sigma\right) &= -\frac{n}{2}\left[p\ln(2\,\pi) + \sum_{k=1}^{p}\ln\left(\sigma_{kk}\right)\right] - \frac{1}{2}\,\mathrm{tr}\left\{\Sigma^{-1}\sum_{i=1}^{n}\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)\,\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)^{\mathrm{T}}\right\} \\
&= -\frac{np}{2}\,\ln(2\,\pi) - \frac{n}{2}\sum_{k=1}^{p}\ln\left(\sigma_{kk}\right) - \frac{1}{2}\,\mathrm{tr}\left\{\sum_{i=1}^{n}\left[\Sigma^{-1}\,\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)^{\mathrm{T}}\right]\right\} \\
&= -\frac{np}{2}\,\ln(2\,\pi) - \frac{n}{2}\sum_{k=1}^{p}\ln\left(\sigma_{kk}\right) - \frac{1}{2}\sum_{i=1}^{n}\,\mathrm{tr}\left\{\Sigma^{-1}\,\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)\left(\boldsymbol{x}_i - \overline{\boldsymbol{x}}\right)^{\mathrm{T}}\right\} \\
&= -\frac{np}{2}\,\ln(2\,\pi) - \frac{n}{2}\sum_{k=1}^{p}\ln\left(\sigma_{kk}\right) - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{p}\frac{\left(x_{ik} - \overline{x}_k\right)^2}{\sigma_{kk}}
\end{aligned}$$

为求 σ_{jj} 的极大似然估计，对数似然函数关于 σ_{jj} 求导并令其为零，得似然方程组如下

$$\frac{\partial\,\ell}{\partial\,\sigma_{jj}} = -\frac{n}{2}\cdot\frac{1}{\sigma_{jj}} + \frac{1}{2\,\sigma_{jj}^2}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2 = 0\,, \quad j = 1,\,2,\,\ldots,\,p$$

由此解得 σ_{jj} 的极大似然估计为

$$\widehat{\sigma}_{jj}=\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij}-\overline{x}_j\right)^2=s_{jj}\,,\quad j=1,\,2,\,\ldots,\,p$$

故，Σ 的极大似然估计为

$$\widehat{\Sigma}=\operatorname{diag}\left(s_{11},\,s_{22},\,\ldots,\,s_{pp}\right)=\begin{pmatrix}s_{11}&&&\\&s_{22}&&\\&&\ddots&\\&&&s_{pp}\end{pmatrix}$$

4. [2 分] 证明定理 6.1.

定理 6.1: 设 $\boldsymbol{s}=\boldsymbol{s}\left(\mathcal{X};\,\boldsymbol{\theta}\right)$ 是评分函数，如果 $\widehat{\boldsymbol{\theta}}=\boldsymbol{t}=\boldsymbol{t}\left(\mathcal{X};\,\boldsymbol{\theta}\right)$ 是 $\mathcal{X}$ 与 $\boldsymbol{\theta}$ 的任一函数，则在正则条件下有

$$\mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}}\right)=\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left(\boldsymbol{t}^{\mathrm{T}}\right)-\mathbb{E}\left(\frac{\partial\boldsymbol{t}^{\mathrm{T}}}{\partial\boldsymbol{\theta}}\right).\tag{6.3}$$

提示：从

$$\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left(\boldsymbol{t}^{\mathrm{T}}\right)=\frac{\partial}{\partial\boldsymbol{\theta}}\int\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\cdot L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\mathrm{d}\mathcal{X}$$

出发，注意到

$$\boldsymbol{s}\left(\mathcal{X};\,\boldsymbol{\theta}\right)=\frac{1}{L\left(\mathcal{X};\,\boldsymbol{\theta}\right)}\cdot\frac{\partial}{\partial\boldsymbol{\theta}}L\left(\mathcal{X};\,\boldsymbol{\theta}\right).$$

【证】 因为 $\boldsymbol{t}=\boldsymbol{t}\left(\mathcal{X};\,\boldsymbol{\theta}\right)$ 是 $\mathcal{X}$ 的函数，似然函数 $L\left(\mathcal{X};\,\boldsymbol{\theta}\right)$ 是 $\mathcal{X}$ 的联合分布密度，由随机变量函数的数学期望的计算公式，我们有

$$\mathbb{E}\left(\boldsymbol{t}^{\mathrm{T}}\right)=\int\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\cdot L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\mathrm{d}\mathcal{X}$$

再利用正则性条件，则有

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\theta}}\mathbb{E}\left(\boldsymbol{t}^{\mathrm{T}}\right)&=\frac{\partial}{\partial\boldsymbol{\theta}}\int\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\cdot L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\mathrm{d}\mathcal{X}=\int\frac{\partial}{\partial\boldsymbol{\theta}}\left[\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\cdot L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\right]\mathrm{d}\mathcal{X}\\&=\int\left\{\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\right]\cdot L\left(\mathcal{X};\,\boldsymbol{\theta}\right)+\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\cdot\left[\frac{\partial}{\partial\boldsymbol{\theta}}L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\right]\right\}\mathrm{d}\mathcal{X}\\&=\int\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\right]\cdot L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\mathrm{d}\mathcal{X}+\int\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\cdot\left[\frac{1}{L\left(\mathcal{X};\,\boldsymbol{\theta}\right)}\cdot\frac{\partial}{\partial\boldsymbol{\theta}}L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\right]\cdot L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\mathrm{d}\mathcal{X}\\&=\mathbb{E}\left(\frac{\partial\boldsymbol{t}^{\mathrm{T}}}{\partial\boldsymbol{\theta}}\right)+\int\boldsymbol{t}^{\mathrm{T}}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\cdot\boldsymbol{s}\left(\mathcal{X};\,\boldsymbol{\theta}\right)\cdot L\left(\mathcal{X};\,\boldsymbol{\theta}\right)\mathrm{d}\mathcal{X}\\&=\mathbb{E}\left(\frac{\partial\boldsymbol{t}^{\mathrm{T}}}{\partial\boldsymbol{\theta}}\right)+\mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}}\right)\end{aligned}$$

移项后即得所证结果.

5. 设 $\boldsymbol{s}(\mathcal{X};\ \boldsymbol{\theta})$ 是评分函数，$\widehat{\boldsymbol{\theta}} = \boldsymbol{t} = \boldsymbol{t}(\mathcal{X})$ 是 $\boldsymbol{\theta}$ 的任意一个无偏估计量，即 $\mathbb{E}(\boldsymbol{t}) = \boldsymbol{\theta}$.

(a) [2 分] 证明

$$\mathrm{E}\left[\boldsymbol{s}(\mathcal{X};\ \boldsymbol{\theta})\right] = \mathbf{0}\,. \tag{6.4}$$

【证】 利用评分函数的定义，我们有

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{s}(\mathcal{X};\ \boldsymbol{\theta})\right] &= \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\mathcal{X};\ \boldsymbol{\theta})\right] \\
&= \mathbb{E}\left[\frac{1}{L(\mathcal{X};\ \boldsymbol{\theta})} \cdot \frac{\partial}{\partial \boldsymbol{\theta}} L(\mathcal{X};\ \boldsymbol{\theta})\right] \\
&= \int \left[\frac{1}{L(\mathcal{X};\ \boldsymbol{\theta})} \cdot \frac{\partial}{\partial \boldsymbol{\theta}} L(\mathcal{X};\ \boldsymbol{\theta})\right] \cdot L(\mathcal{X};\ \boldsymbol{\theta})\,\mathrm{d}\mathcal{X} \\
&= \int \frac{\partial}{\partial \boldsymbol{\theta}} L(\mathcal{X};\ \boldsymbol{\theta})\,\mathrm{d}\mathcal{X} \quad \longleftarrow \quad \text{正则性条件} \\
&= \frac{\partial}{\partial \boldsymbol{\theta}} \int L(\mathcal{X};\ \boldsymbol{\theta})\,\mathrm{d}\mathcal{X} \\
&= \frac{\partial}{\partial \boldsymbol{\theta}} 1 \\
&= \mathbf{0}
\end{aligned}$$

(b) [2 分] 证明

$$\mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}}\right) = \mathbb{C}\mathrm{ov}\left(\boldsymbol{s},\ \boldsymbol{t}\right) = \mathcal{I}_k\ . \tag{6.5}$$

【证】 利用协方差矩阵的定义，我们有

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left(\boldsymbol{s},\ \boldsymbol{t}\right) &= \mathbb{E}\left\{\left[\boldsymbol{s} - \mathbb{E}(\boldsymbol{s})\right]\left[\boldsymbol{t} - \mathbb{E}(\boldsymbol{t})\right]^{\mathrm{T}}\right\} \\
&= \mathbb{E}\left[\boldsymbol{s}(\boldsymbol{t} - \boldsymbol{\theta})^{\mathrm{T}}\right] = \mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}} - \boldsymbol{s}\boldsymbol{\theta}^{\mathrm{T}}\right) \\
&= \mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}}\right) - \mathbb{E}\left(\boldsymbol{s}\boldsymbol{\theta}^{\mathrm{T}}\right) = \mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}}\right) - \mathbb{E}\left(\boldsymbol{s}\right)\boldsymbol{\theta}^{\mathrm{T}} \\
&= \mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}}\right)
\end{aligned}$$

再根据已证明的**定理** 6.1，则有

$$\mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}}\right) = \frac{\partial}{\partial \boldsymbol{\theta}}\mathbb{E}\left(\boldsymbol{t}^{\mathrm{T}}\right) - \mathbb{E}\left(\frac{\partial \boldsymbol{t}^{\mathrm{T}}}{\partial \boldsymbol{\theta}}\right) = \frac{\partial}{\partial \boldsymbol{\theta}}\left(\boldsymbol{\theta}^{\mathrm{T}}\right) - \mathbb{E}\left(\frac{\partial \boldsymbol{t}^{\mathrm{T}}}{\partial \boldsymbol{\theta}}\right) = \mathcal{I}_k - \mathbb{E}\left(\frac{\partial \boldsymbol{t}^{\mathrm{T}}}{\partial \boldsymbol{\theta}}\right)$$

又因为 $\boldsymbol{t} = \boldsymbol{t}(\mathcal{X})$ 是 $\boldsymbol{\theta}$ 的无偏估计量，它是一个统计量，其中不含未知参数，所以

$$\frac{\partial \boldsymbol{t}^{\mathrm{T}}}{\partial \boldsymbol{\theta}} = \mathbf{0} \quad \Longrightarrow \quad \mathbb{E}\left(\frac{\partial \boldsymbol{t}^{\mathrm{T}}}{\partial \boldsymbol{\theta}}\right) = \mathbf{0}$$

故有

$$\mathbb{E}\left(\boldsymbol{s}\boldsymbol{t}^{\mathrm{T}}\right)=\mathbb{C}\mathrm{ov}\left(\boldsymbol{s},\ \boldsymbol{t}\right)=\mathcal{I}_k$$

6. 设 $\boldsymbol{X}\sim N_2\left(\boldsymbol{\mu},\ \Sigma\right)$，其中已知

$$\Sigma=\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix} \tag{6.6}$$

我们从该总体抽取了容量 $n=6$ 的一个简单随机样本，计算得

$$\overline{\boldsymbol{x}}=\begin{pmatrix}1\\ \frac{1}{2}\end{pmatrix} \tag{6.7}$$

(a) [2 分] 求解下述假设检验问题.

$$H_0:\ \boldsymbol{\mu}=\begin{pmatrix}2\\ \frac{2}{3}\end{pmatrix}\ \longleftrightarrow\ H_1:\ \boldsymbol{\mu}\neq\begin{pmatrix}2\\ \frac{2}{3}\end{pmatrix} \tag{6.8}$$

【解】 这是我们讨论的检验问题 1，检验统计量为

$$-2\log\lambda=n\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}\Sigma^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}_0\right)\ \overset{H_0\text{ 真}}{\sim}\ \chi_p^2$$

已知

$$\boldsymbol{\mu}_0=\begin{pmatrix}2\\ \frac{2}{3}\end{pmatrix},\quad \Sigma=\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix},\quad \overline{\boldsymbol{x}}=\begin{pmatrix}1\\ \frac{1}{2}\end{pmatrix},\quad n=6,\quad p=2$$

计算得检验统计量的取值为

$$-2\log\lambda=6\times\left[\begin{pmatrix}1\\ \frac{1}{2}\end{pmatrix}-\begin{pmatrix}2\\ \frac{2}{3}\end{pmatrix}\right]^{\mathrm{T}}\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix}^{-1}\left[\begin{pmatrix}1\\ \frac{1}{2}\end{pmatrix}-\begin{pmatrix}2\\ \frac{2}{3}\end{pmatrix}\right]=4.778$$

显著水平 $\alpha=0.05$ 时，χ_2^2 分布的上侧 α 分位数 $\chi_2^2(0.05)=5.991465$，因为

$$-2\log\lambda=4.778<5.991465=\chi_2^2(0.05)$$

所以，此时我们不能拒绝 $H_0:\ \boldsymbol{\mu}=\begin{pmatrix}2\\ \frac{2}{3}\end{pmatrix}$. 上述计算可使用下述 R 代码实现:

```
n = 6 # 样本容量
p = 2 # 正态总体的维数
xbar = matrix(c(1, 1/2), nrow = 2, byrow = TRUE) # 样本均值向量
mu_0 = matrix(c(2, 2/3), nrow = 2, byrow = TRUE) # 零假设时的参数取值
```

```
Sigma = matrix(c(2, -1, -1, 2), nrow = 2, byrow = TRUE) # 已知的总体协方差矩阵
n * t(xbar - mu_0) %*% solve(Sigma) %*% (xbar - mu_0) # 计算检验统计量的值
alpha = 0.05 # 假设检验的显著水平
qchisq(alpha, p, lower.tail = FALSE) # 假设检验的临界值
```

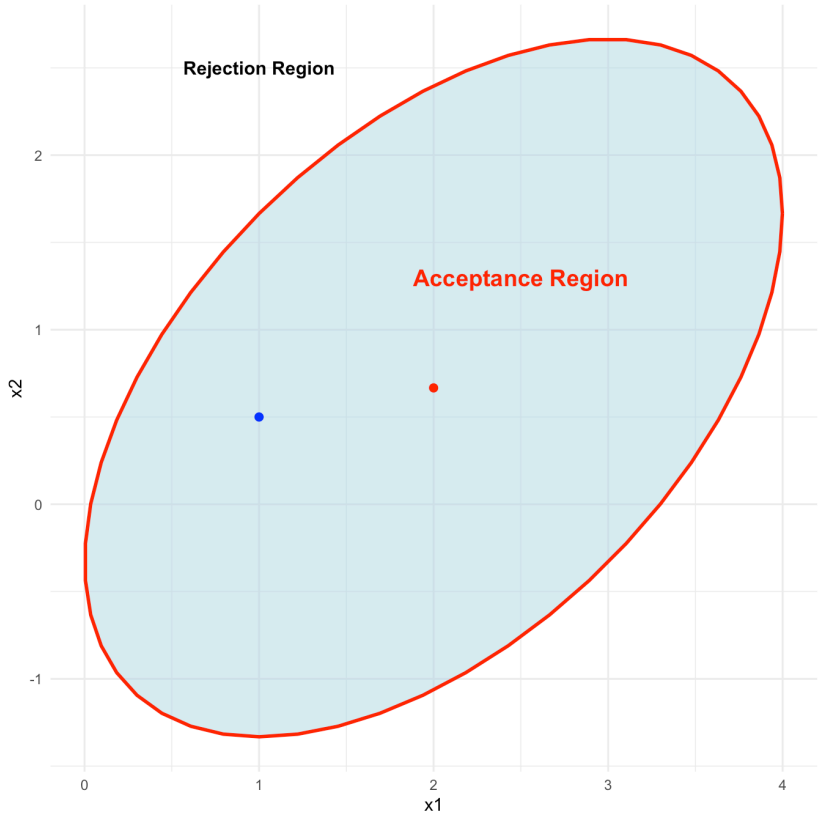

图 6.1: 假设检验问题的接受域与拒绝域 (协方差矩阵已知).

(b) [2 分] 作拒绝域的可视化图形.

【解】 该检验问题的拒绝域为

$$\left\{\boldsymbol{x}\in\mathbb{R}^2 \;\middle|\; -2\log\lambda(\boldsymbol{x}) = n\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}\Sigma^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right) > 5.991465 = \chi_2^2(0.05)\right\}$$

或等价地表示为

$$\left\{\boldsymbol{x}\in\mathbb{R}^2 \;\middle|\; \left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}\Sigma^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right) > \frac{\chi_2^2(0.05)}{n} \triangleq r^2\right\}$$

在 $\mathbb{R}^2$ 中，$\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}\Sigma^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right) = r^2$ 是以 $\boldsymbol{\mu}_0$ 为中心、以 r 为半径、以 Σ^{-1} 为度量矩阵的一个椭圆. 于是，该检验问题拒绝域的形状 (如图6.1所示) 可以用下述 R 代码实现:

```
library(car)
library(ggplot2)
```

```r
n = 6
p = 2
alpha = 0.05
mu = c(2, 2/3) # 均值向量
S = matrix(c(2, -1, -1, 2), nrow = 2, byrow = TRUE) # 协方差矩阵
M = solve(S) # 度量矩阵
r = sqrt(qchisq(alpha, p, lower.tail = FALSE)) # 半径

# 提取椭圆点
ell = car::ellipse(mu, M, radius = r, draw = FALSE)
ell_df = as.data.frame(ell)
colnames(ell_df) = c("x1", "x2")

# 绘图
Fig_1 = ggplot() +
  geom_polygon(data = ell_df, aes(x1, x2), fill = "lightblue", alpha = 0.5) +
  geom_path(data = ell_df, aes(x1, x2), color = "red", linewidth = 1) +
  geom_point(aes(x = 2, y = 2/3), size = 2, color = "red") +
  geom_point(aes(x = 1, y = 1/2), size = 2, color = "blue") +
  annotate("text",
           x = 2.5, y = 1.3,
           label = "Acceptance Region",
           color = "red", fontface = 2, size = 5) +
  annotate("text",
           x = 1, y = 2.5,
           label = "Rejection Region",
           color = "black", fontface = 2, size = 4) +
  coord_equal() +
  labs(title="") +
  theme_minimal()
Fig_1
```

图 6.1 中的阴影部分是 H_0 的接受域，$\mathbb{R}^2$ 中除去阴影部分之外的区域即是 H_0 的拒绝域.

7. 设 $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$，其中 Σ 未知. 从中抽取了容量 $n = 6$ 的一个样本，计算得样本均值和样本方差如下：

$$\overline{\boldsymbol{x}} = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}, \quad \mathcal{S} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \tag{6.9}$$

(a) [2 分] 求解下述检验问题.

$$H_0:\ \boldsymbol{\mu}=\begin{pmatrix}2\\ \\ \dfrac{2}{3}\end{pmatrix}\ \longleftrightarrow\ H_1:\ \boldsymbol{\mu}\neq\begin{pmatrix}2\\ \\ \dfrac{2}{3}\end{pmatrix} \tag{6.10}$$

【解】 这是我们讨论的**检验问题 2**，检验统计量为

$$\frac{n-p}{p}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}\mathcal{S}^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}_0\right)\ \overset{H_0\ \text{true}}{\sim}\ F_{p,\ n-p}$$

因为

$$\boldsymbol{\mu}_0=\begin{pmatrix}2\\ \\ \dfrac{2}{3}\end{pmatrix},\quad \mathcal{S}=\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix},\quad \overline{\boldsymbol{x}}=\begin{pmatrix}1\\ \\ \dfrac{1}{2}\end{pmatrix},\quad n=6\,,\quad p=2$$

所以我们有

$$\begin{aligned}\frac{n-p}{p}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}\mathcal{S}^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}_0\right)&=\frac{6-2}{2}\left(\begin{pmatrix}1\\ \\ \dfrac{1}{2}\end{pmatrix}-\begin{pmatrix}2\\ \\ \dfrac{2}{3}\end{pmatrix}\right)^{\mathrm{T}}\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix}^{-1}\left(\begin{pmatrix}1\\ \\ \dfrac{1}{2}\end{pmatrix}-\begin{pmatrix}2\\ \\ \dfrac{2}{3}\end{pmatrix}\right)\\ &=1.5926\end{aligned}$$

取检验的显著水平 $\alpha=0.05$，则上侧分位数 $F_{2,\ 4}(0.05)=6.944$. 由于

$$\frac{n-p}{p}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}_0\right)^{\mathrm{T}}\mathcal{S}^{-1}\left(\overline{\boldsymbol{x}}-\boldsymbol{\mu}_0\right)=1.5926<6.944=F_{2,\ 4}(0.05)$$

所以，显著水平 $\alpha=0.05$ 时，我们不能拒绝 $H_0:\ \boldsymbol{\mu}=\begin{pmatrix}2\\ \\ \dfrac{2}{3}\end{pmatrix}$. 上述计算的 R 代码如下:

```
n = 6 # 样本容量
p = 2 # 正态总体的维数
xbar = matrix(c(1, 1/2), nrow = 2, byrow = TRUE) # 样本均值向量
mu_0 = matrix(c(2, 2/3), nrow = 2, byrow = TRUE) # 零假设时的参数取值
S = matrix(c(2, -1, -1, 2), nrow = 2, byrow = TRUE) # 样本协方差矩阵
F_test = ((n - p) / p) * t(xbar - mu_0) %*% solve(S) %*% (xbar - mu_0)
F_test  # 检验统计量的值
alpha = 0.05 # 假设检验的显著水平
qf(alpha, p, n - p, lower.tail = FALSE) # 假设检验的临界值
pf(F_test, p, n - p, lower.tail = FALSE) # 假设检验的p值
```

(b) [2 分] 作拒绝域的可视化图形.

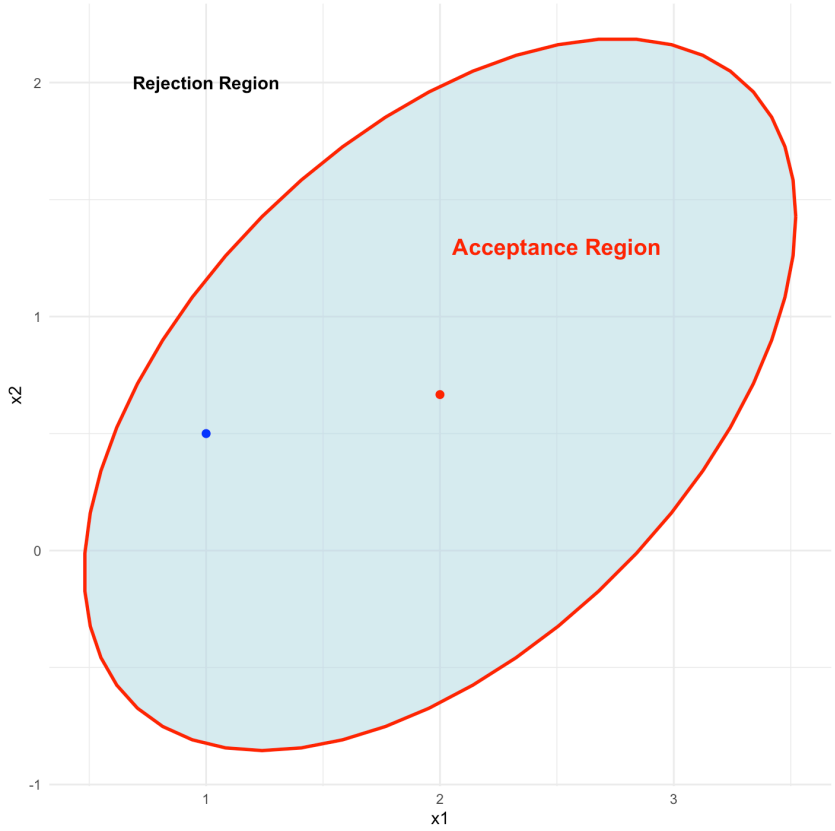

图 6.2: 假设检验问题的接受域与拒绝域 (协方差矩阵未知).

【解】 该检验问题的拒绝域为

$$\left\{ \boldsymbol{x} \in \mathbb{R}^2 \;\middle|\; \frac{n-p}{p} \left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)^{\mathrm{T}} S^{-1} \left(\boldsymbol{x}-\boldsymbol{\mu}_0\right) > 6.944272 = F_{2,4}(0.05) \right\}$$

或等价地表示为

$$\left\{ \boldsymbol{x} \in \mathbb{R}^2 \;\middle|\; \left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)^{\mathrm{T}} S^{-1} \left(\boldsymbol{x}-\boldsymbol{\mu}_0\right) > F_{2,4}(0.05) \cdot \frac{p}{n-p} \triangleq r^2 \right\}$$

在 $\mathbb{R}^2$ 中，$\left(\boldsymbol{x}-\boldsymbol{\mu}_0\right)^{\mathrm{T}} S^{-1} \left(\boldsymbol{x}-\boldsymbol{\mu}_0\right) = r^2$ 是以 $\boldsymbol{\mu}_0$ 为中心、以 r 为半径、以 S^{-1} 为度量矩阵的一个椭圆. 于是，该检验问题接受域、拒绝域的形状 (如图6.2所示) 可以用下述 R 代码实现:

```
library(car)
library(ggplot2)

n = 6
p = 2
alpha = 0.05
mu = c(2, 2/3) # 均值向量
S = matrix(c(2, -1, -1, 2), nrow = 2, byrow = TRUE) # 协方差矩阵
```

```
M = solve(S) # 度量矩阵
r = sqrt(qf(alpha, p, n - p, lower.tail = FALSE) * p / (n-p)) # 半径

# 提取椭圆点
ell = car::ellipse(mu, M, radius = r, draw = FALSE)
ell_df = as.data.frame(ell)
colnames(ell_df) = c("x1", "x2")

# 绘图
Fig_2 = ggplot() +
  geom_polygon(data = ell_df, aes(x1, x2), fill = "lightblue", alpha = 0.5) +
  geom_path(data = ell_df, aes(x1, x2), color = "red", linewidth = 1) +
  geom_point(aes(x = 2, y = 2/3), size = 2, color = "red") +
  geom_point(aes(x = 1, y = 1/2), size = 2, color = "blue") +
  annotate("text",
           x = 2.5, y = 1.3,
           label = "Acceptance Region",
           color = "red", fontface = 2, size = 5) +
  annotate("text",
           x = 1, y = 2,
           label = "Rejection Region",
           color = "black", fontface = 2, size = 4) +
  coord_equal() +
  labs(title="") +
  theme_minimal()
Fig_2
```

图 6.2 中的阴影部分是 H_0 的接受域，$\mathbb{R}^2$ 中除去阴影部分之外的区域即是 H_0 的拒绝域.

Chapter 7

第 7 周作业参考答案

1. 从二元正态分布总体模拟抽样一个简单随机样本，其中

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix} \tag{7.1}$$

检验假设 H_0： $2\mu_1 - \mu_2 = 0.2$.

(a) [2 分] 首先，假设 Σ 已知.

【解】 由于

$$2\mu_1 - \mu_2 = 0.2 \iff \begin{pmatrix} 2 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = 0.2 \iff \mathcal{A}\boldsymbol{\mu} = \boldsymbol{a}$$

这是我们课堂中讨论的**检验问题 5**，检验统计量为

$$n\left(\mathcal{A}\overline{\boldsymbol{x}} - \boldsymbol{a}\right)^{\mathrm{T}} \left(\mathcal{A}\Sigma\mathcal{A}^{\mathrm{T}}\right)^{-1} \left(\mathcal{A}\overline{\boldsymbol{x}} - \boldsymbol{a}\right) \overset{H_0\text{ true}}{\sim} \chi_q^2$$

对于给定的显著水平 α，当上述检验统计量的值大于 $\chi_q^2(\alpha)$ 时，我们拒绝 H_0.

我们从该总体当中模拟抽样，得到容量 $n = 10$ 的样本，并计算检验统计量以及检验的临界值，相应的 R 代码如下：

```
rm(list = ls(all = TRUE))
graphics.off()
library(MASS)
mu.0 = matrix(c(1, 2), nrow = 2, byrow = FALSE)
Sigma.0 = matrix(c(1, 0.5, 0.5, 2), nrow = 2, byrow = FALSE)
n = 10
a = 0.2
x = mvrnorm(n, mu.0, Sigma.0) # 模拟抽样
```

```
x.bar = as.matrix(colMeans(x)) # 计算样本均值向量
A = matrix(c(2, -1), nrow = 1, byrow = TRUE) # 矩阵A
test = n * t(A %*% x.bar - a) %*% solve(A %*% Sigma.0 %*% t(A)) %*%
    (A %*% x.bar - a)  # 计算经验统计量的值
test
qchisq(0.05, 1, lower.tail = FALSE) # 假设检验的临界值
```

计算得到

$$n\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\,\Sigma\,\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)=0.01605$$

取检验的显著水平 $\alpha=0.05$，检验的临界值为

$$\chi_q^2(\alpha)=\chi_1^2(0.05)=3.841459$$

因为检验统计量的值 $0.01605<3.841459=\chi_1^2(0.05)$，所以，在显著水平 $\alpha=0.05$ 时，我们接受假设 H_0：$2\mu_1-\mu_2=0.2$.

(b) [2 分] 其次，假设 Σ 未知.

【解】 这是我们课堂中讨论的**检验问题 6**，检验统计量为

$$(n-1)\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\,\mathcal{S}\,\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)\overset{H_0\ \text{true}}{\sim}T^2_{q,\ n-1}=\frac{(n-1)\,q}{n-q}\,F_{q,\ n-q}$$

利用上述模拟数据，我们计算得检验统计量的值为

$$(n-1)\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\mathcal{S}\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)=0.0119398$$

显著水平 $\alpha=0.05$ 时，检验的临界值为

$$F_{q,\ n-q}(\alpha)=F_{1,\ 9}(0.05)=5.117355$$

因为检验统计量的值 $0.0119398<5.117355=F_{1,\ 9}(0.05)$，所以当显著水平 $\alpha=0.05$ 时，我们接受 H_0：$2\mu_1-\mu_2=0.2$. 上述计算可用以下 R 代码实现：

```
S = cov(x) * (n - 1) / n # 样本协方差矩阵
test.2 = (n-1) * t(A %*% x.bar - a) %*% solve(A %*% S %*% t(A)) %*%
      (A %*% x.bar - a) # 计算经验统计量的值
test.2
qf(0.05, 1, n-1, lower.tail = FALSE) # 假设检验的临界值
```

(c) [2 分] 比较上述结果.

【解】当样本容量 $n=10$，检验的显著水平 $\alpha=0.05$ 时，两种检验方法均接受了零假设 H_0：$2\mu_1-\mu_2=0.2$. 当样本容量 n 变大时，样本协方差矩阵 $\mathcal{S}$ 会收敛于总体的协方差矩阵 Σ，两种检验

方法的结论亦会趋于一致.

2. 对上课用到的美国公司数据集.

(a) [2 分] 使用 X_1 至 X_6 全部六个变量的观测数据，检验能源行业的均值向量与制造业的均值向量是否相同.

【解】 这是我们课堂中讨论的检验问题 8，检验统计量为

$$\frac{n_1 n_2 \left(n_1+n_2-p-1\right)}{p\left(n_1+n_2\right)^2}\left\{\left(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2\right)-\boldsymbol{\delta}\right\}^{\mathrm{T}} \mathcal{S}^{-1}\left\{\left(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2\right)-\boldsymbol{\delta}\right\} \overset{H_0 \text{ true}}{\sim} F_{p,\ n_1+n_2-p-1}$$

其中

$$\mathcal{S}=\frac{1}{n_1+n_2}\left(n_1 \mathcal{S}_1+n_2 \mathcal{S}_2\right)$$

拒绝域为

$$\frac{n_1 n_2 \left(n_1+n_2-p-1\right)}{p\left(n_1+n_2\right)^2}\left\{\left(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2\right)-\boldsymbol{\delta}\right\}^{\mathrm{T}} \mathcal{S}^{-1}\left\{\left(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2\right)-\boldsymbol{\delta}\right\} \geq F_{p,\ n_1+n_2-p-1}(\alpha)$$

对于美国公司数据集当中的能源行业与制造业两个部分的数据，我们可计算得

$$\frac{n_1 n_2 \left(n_1+n_2-p-1\right)}{p\left(n_1+n_2\right)^2}\left\{\left(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2\right)-\boldsymbol{\delta}\right\}^{\mathrm{T}} \mathcal{S}^{-1}\left\{\left(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2\right)-\boldsymbol{\delta}\right\}=2.152601$$

检验的临界值为

$$F_{p,\ n_1+n_2-p-1}(\alpha)=F_{6,\ 18}(0.05)=2.661305$$

由于检验统计量的值 $2.152601<2.661305=F_{6,\ 18}(0.05)$，当检验的显著水平 $\alpha=0.05$ 时，我们接受零假设 H_0：$\boldsymbol{\mu}_1-\boldsymbol{\mu}_2=\mathbf{0}$. 用到的 R 代码如下：

```
rm(list = ls(all = TRUE))
graphics.off()
# energy sector data
x = rbind(c(13621, 4848, 4572, 485, 898.9, 23.4),
          c(1117, 1038, 478, 59.7, 91.7, 3.8),
          c(1633, 701, 679, 74.3, 135.9, 2.8),
          c(5651, 1254, 2002, 310.7, 407.9, 6.2),
          c(5835, 4053, 1601, -93.8, 173.8, 10.8),
          c(3494, 1653, 1442, 160.9, 320.3, 6.4),
          c(1654, 451, 779, 84.8, 130.4, 1.6),
          c(1679, 1354, 687, 93.8, 154.6, 4.6),
          c(1257, 355, 181, 167.5, 304, 0.6),
          c(1743, 597, 717, 121.6, 172.4, 3.5),
          c(1440, 1617, 639, 81.7, 126.4, 3.5),
```

```
          c(14045, 15636, 2754, 418, 1462, 27.3),
          c(3010, 749, 1120, 146.3, 209.2, 3.4),
          c(3086, 1739, 1507, 202.7, 335.2, 4.9),
          c(1995, 2662, 341, 34.7, 100.7, 2.3)
          )
x = as.data.frame(x)
bar.x = as.matrix(colMeans(x)) # 样本均值向量
bar.x
# manufacturing sector data
y = rbind(c(1093, 1679, 1070, 100.9, 164.5, 20.8),
          c(1128, 1516, 430, -47, 26.7, 13.2),
          c(1804, 2564, 483, 70.5, 164.9, 26.6),
          c(4662, 4781, 2988, 28.7, 371.5, 66.2),
          c(6307, 8199, 598, -771.5, -524.3, 57.5),
          c(2366, 3305, 1117, 131.2, 256.5, 25.2),
          c(4084, 4346, 3023, 302.7, 521.7, 37.5),
          c(10348, 5721, 1915, 223.6, 322.5, 49.5),
          c(752, 2149, 101, 11.1, 15.2, 2.6),
          c(10528, 14992, 5377, 312.7, 710.7, 184.8)
          )
y = as.data.frame(y)
bar.y = as.matrix(colMeans(y)) # 样本均值向量
bar.y
p = 6 # 数据维数
n1 = nrow(x) # 能源行业的样本容量
n2 = nrow(y) # 制造业的样本容量
S1 = (n1 - 1) * cov(x) / n1 # 能源行业的样本协方差矩阵
S2 = (n2 - 1) * cov(y) / n2 # 制造业的样本协方差矩阵
S = (n1 * S1 + n2 * S2) / (n1 + n2) # 合并的样本协方差矩阵
test.1 = (n1 * n2 * (n1 + n2 - p - 1) / (p * (n1 + n2)^2)) *
      t(bar.x - bar.y) %*% solve(S) %*% (bar.x - bar.y)
test.1 # 检验统计量的值
qf(0.05, p, n1 + n2 - p - 1, lower.tail = FALSE) # 检验的临界值
```

(b) [2 分] 计算均值差的联合置信区间.

【解】 置信度为 $1-\alpha$ 时，均值差 $\boldsymbol{\delta}=\boldsymbol{\mu}_1-\boldsymbol{\mu}_2$ 的联合置信区间为

$$\delta_j=\overline{x}_{1j}-\overline{x}_{2j}\pm\sqrt{\frac{p\left(n_1+n_2\right)^2}{n_1n_2\left(n_1+n_2-p-1\right)}\,F_{p,\ n_1+n_2-p-1}(\alpha)\ s_{jj}}\ ,\quad j=1,\ 2,\ \ldots,\ p$$

对于美国公司数据集当中的能源行业与制造业两个部分的数据，计算可得

$$\begin{aligned}
-7639.3970 &\le \delta_1=\mu_{1\,1}-\mu_{1\,2}\le 7192.9970\\
-9613.2203 &\le \delta_2=\mu_{2\,1}-\mu_{2\,2}\le 4923.7537\\
-2923.8431 &\le \delta_3=\mu_{3\,1}-\mu_{3\,2}\le 2103.3097\\
-295.4852 &\le \delta_4=\mu_{4\,1}-\mu_{4\,2}\le 535.9586\\
-527.1320 &\le \delta_5=\mu_{5\,1}-\mu_{5\,2}\le 790.9387\\
-102.3026 &\le \delta_6=\mu_{6\,1}-\mu_{6\,2}\le 19.5359
\end{aligned}$$

上述计算用到的 R 代码如下：

```
a = array(0, dim = p)
for (i in 1:p) {
    a[i] = sqrt(((p * (n1 + n2)^2) / (n1 * n2 *(n1 + n2 - p - 1))) *
        qf(0.05, p, n1 + n2 - p - 1, lower.tail = FALSE) * S[i, i]) }
a = as.matrix(a)
delta.lower = bar.x - bar.y - a
delta.upper = bar.x - bar.y + a
CI = cbind(delta.lower, delta.upper)
colnames(CI) = c("Lower", "Upper")
CI
```

3. 设 $\boldsymbol{X}\sim N_2\left(\boldsymbol{\mu},\,\Sigma\right)$，其中 Σ 已知

$$\Sigma=\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix}\tag{7.2}$$

从中抽取了容量 $n=6$ 的一个简单随机样本，计算得

$$\overline{\boldsymbol{x}}=\begin{pmatrix}1\\ \dfrac{1}{2}\end{pmatrix}\tag{7.3}$$

(a) [2 分] 求解假设检验问题 $H_0:\ \mu_1+\mu_2=\dfrac{7}{2}\ \longleftrightarrow\ H_1:\ \mu_1+\mu_2\neq\dfrac{7}{2}$.

【解】 这是我们课堂中讨论的**检验问题 5**，检验统计量为

$$n\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\,\Sigma\,\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)\quad\overset{H_0\ \text{true}}{\sim}\quad\chi^2_q$$

对于给定的显著水平 α，当上述检验统计量的值很大时我们拒绝零假设 H_0. 因为

$$\mu_1+\mu_2=\frac{7}{2} \iff \begin{pmatrix}1 & 1\end{pmatrix}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=\frac{7}{2} \iff \mathcal{A}\boldsymbol{\mu}=\boldsymbol{a}$$

所以我们有

$$\mathcal{A}=\begin{pmatrix}1 & 1\end{pmatrix},\quad \boldsymbol{a}=\frac{7}{2},\quad n=6,\quad q=1,\quad \overline{\boldsymbol{x}}=\begin{pmatrix}1\\ \frac{1}{2}\end{pmatrix},\quad \Sigma=\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix}$$

计算得检验统计量的取值为

$$n\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\,\Sigma\,\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)=12$$

取检验的显著水平 $\alpha=0.05$，检验的临界值为 $\chi_1^2(0.05)=3.841459$. 由于检验统计量的值 $12>3.841459=\chi_1^2(0.05)$，我们拒绝零假设 $H_0:\ \mu_1+\mu_2=\dfrac{7}{2}$. 计算用到的 R 代码如下：

```
rm(list = ls(all = TRUE))
graphics.off()
n = 6
q = 1
a = 7/2
A = matrix(c(1, 1), nrow = 1, byrow = TRUE)
x.bar = matrix(c(1, 1/2), nrow = 2, byrow = TRUE)
Sigma = matrix(c(2, -1, -1, 2), nrow = 2, byrow = TRUE)
n * t(A %*% x.bar - a) %*% solve(A %*% Sigma %*% t(A)) %*% (A %*% x.bar - a)
qchisq(0.05, q, lower.tail = FALSE)
```

(b) [2 分] 作拒绝域的可视化图形.

【解】 该检验问题的拒绝域为

$$\left\{\boldsymbol{x}\in\mathbb{R}^2 \;\middle|\; n\left(\mathcal{A}\boldsymbol{x}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\,\Sigma\,\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\boldsymbol{x}-\boldsymbol{a}\right)>\chi_q^2(\alpha)\right\}$$

具体可表示为

$$\left\{\boldsymbol{x}\in\mathbb{R}^2 \;\middle|\; 6\times\left(x_1+x_2-\frac{7}{2}\right)\times\frac{1}{2}\times\left(x_1+x_2-\frac{7}{2}\right)>3.841459\right\}$$

也就是

$$\left\{\boldsymbol{x}\in\mathbb{R}^2 \;\middle|\; \left(x_1+x_2-\frac{7}{2}\right)^2>\frac{3.841459}{3}=1.280486\right\}$$

亦即

$$\left\{\boldsymbol{x}\in\mathbb{R}^2 \mid x_1+x_2>4.631586 \text{ 或 } x_1+x_2<2.368414\right\}$$

拒绝域的形状如图 7.1 所示，绘图用到的 R 代码如下：

```
plot(x = c(-4, 6), y = c(-4, 6), asp = 1, type = "n", axes = FALSE,
     xlab = "", ylab = "")
x = c(-3.7, 6, 6, -1.4, -3.7)
y = c(6, -3.7, -1.4, 6, 6)
polygon(x, y, col = "grey90", border = "grey90")
arrows(-4, 0, 6, 0, length = 0.15, code = 2)
arrows(0, -4, 0, 6, length = 0.15, code = 2)
lines(c(-1.4, 6), c(6, -1.4), lwd = 2, col = "red")
lines(c(-3.7, 6), c(6, -3.7), lwd = 2, col = "red")
text(1, 2.5, labels = "Acceptance Region", cex = 1.5, col = "red", srt = -45)
text(3.5, 4, labels = "Rejection Region", cex = 1.5, col = "blue")
text(-2, -1, labels = "Rejection Region", cex = 1.5, col = "blue")
points(1, 1/2, pch = 16, cex = 1.5, col = "black")
```

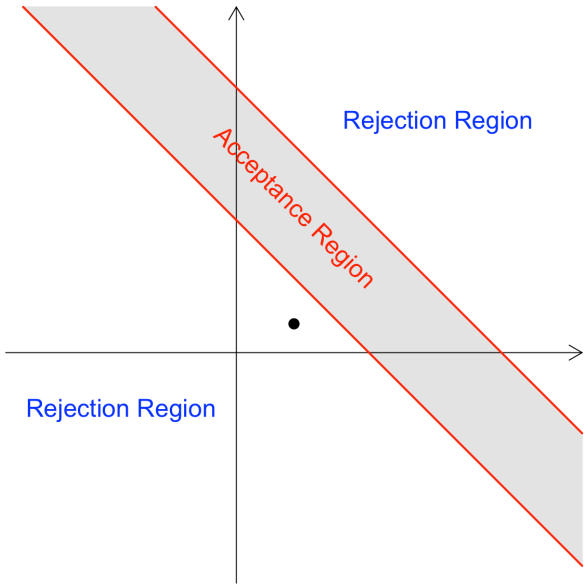

图 7.1: 假设检验问题 (线性组合) 的接受域与拒绝域 (协方差矩阵已知).

4. 设 $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \Sigma)$，其中 Σ 未知. 从中抽取了容量 $n = 6$ 的一个简单随机样本，计算得

$$\overline{\boldsymbol{x}} = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}, \quad \mathcal{S} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \tag{7.4}$$

(a) [2 分] 求解假设检验问题 $H_0:\ \mu_1-\mu_2=\dfrac{1}{2} \longleftrightarrow H_1:\ \mu_1-\mu_2\neq\dfrac{1}{2}$.

【解】 这是我们课堂上讨论的**检验问题 6**，检验统计量为

$$\frac{n-q}{q}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\mathcal{S}\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right) \overset{H_0\text{ true}}{\sim} F_{q,\ n-q}$$

给定检验的显著水平 α，当该统计量的值大于 $F_{q,\ n-q}(\alpha)$ 时我们拒绝零假设 H_0. 由于

$$\mu_1-\mu_2=\frac{1}{2} \iff \begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}=\frac{1}{2} \iff \mathcal{A}\boldsymbol{\mu}=\boldsymbol{a}$$

于是，

$$\mathcal{A}=\begin{pmatrix}1 & -1\end{pmatrix},\quad \boldsymbol{a}=\frac{1}{2},\quad n=6,\quad q=1,\quad \overline{\boldsymbol{x}}=\begin{pmatrix}1\\ \dfrac{1}{2}\end{pmatrix},\quad \mathcal{S}=\begin{pmatrix}2 & -1\\ -1 & 2\end{pmatrix}$$

从而检验统计量的值

$$\frac{n-q}{q}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\mathcal{S}\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)=0$$

检验的临界值 $F_{1,\ 5}(0.05)=6.607891$，因为检验统计量的值 $0<6.607891=F_{1,\ 5}(0.05)$，所以，当检验的显著水平 $\alpha=0.05$ 时，我们接受零假设 $H_0:\ \mu_1-\mu_2=\dfrac{1}{2}$. 计算的 R 代码如下：

```
rm(list = ls(all = TRUE))
graphics.off()
n = 6
q = 1
a = 1/2
A = matrix(c(1, -1), nrow = 1, byrow = TRUE)
x.bar = matrix(c(1, 1/2), nrow = 2, byrow = TRUE)
S = matrix(c(2, -1, -1, 2), nrow = 2, byrow = TRUE)
test = ((n-q)/q) * t(A %*% x.bar - a) %*%
    solve(A %*% S %*% t(A)) %*% (A %*% x.bar - a)
test
qf(0.05, q, n-q, lower.tail = FALSE)
```

(b) [2 分] 作拒绝域的可视化图形.

【解】 该检验问题的拒绝域为

$$\left\{\boldsymbol{x}\in\mathbb{R}^2 \;\middle|\; \frac{n-q}{q}\left(\mathcal{A}\boldsymbol{x}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\mathcal{S}\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\boldsymbol{x}-\boldsymbol{a}\right)>F_{q,\,n-q}(\alpha)\right\}$$

具体可表示为

$$\left\{ \boldsymbol{x} \in \mathbb{R}^2 \;\middle|\; \frac{6-1}{1} \left(x_1 - x_2 - \frac{1}{2} \right)^{\mathrm{T}} \times \frac{1}{6} \times \left(x_1 - x_2 - \frac{1}{2} \right) > 6.607891 \right\}$$

也就是

$$\left\{ \boldsymbol{x} \in \mathbb{R}^2 \;\middle|\; \left(x_1 - x_2 - \frac{1}{2} \right)^2 > 7.929469 \right\}$$

亦即

$$\left\{ \boldsymbol{x} \in \mathbb{R}^2 \mid x_1 - x_2 > 3.315931 \text{ 或 } x_1 - x_2 < -2.315931 \right\}$$

拒绝域的形状如图 7.2 所示，绘图用到的 R 代码如下：

```
rm(list = ls(all = TRUE))
graphics.off()
plot(x = c(-4, 5), y = c(-4, 5), asp = 1, type = "n",
     axes = FALSE, xlab = "", ylab = "")
x = c(-4, -0.7, 5, 5, 2.7, -4, -4)
y = c(-4, -4, 1.7, 5, 5, -1.7, -4)
polygon(x, y, col = "grey90", border = "grey90")
arrows(-4, 0, 5, 0, length = 0.15, code = 2)
arrows(0, -4, 0, 5, length = 0.15, code = 2)
lines(c(-0.7, 5), c(-4, 1.7), lwd = 2, col = "red")
lines(c(-4, 2.7), c(-1.7, 5), lwd = 2, col = "red")
text(2, 2, labels = "Acceptance Region", cex = 1.5, col = "red", srt = 45)
text(-3, 2, labels = "Rejection Region", cex = 1.5, col = "blue")
text(3, -2.5, labels = "Rejection Region", cex = 1.5, col = "blue")
points(1, 1/2, pch = 16, cex = 1.5, col = "black")
```

5. 已知 $\boldsymbol{X} \sim N_3(\boldsymbol{\mu}, \Sigma)$. 从中抽取了容量 $n = 10$ 的一个简单随机样本，算得

$$\overline{\boldsymbol{x}} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \mathcal{S} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix} \tag{7.5}$$

(a) [2 分] 已知 $\mathcal{S}$ 的特征值为整数，给出 $\boldsymbol{\mu}$ 的置信度为 95% 一个置信域.

【提示】为计算特征值，可以利用下式：

$$|\mathcal{S}| = \prod_{j=1}^{3} \lambda_j \,, \quad \operatorname{tr}(\mathcal{S}) = \sum_{j=1}^{3} \lambda_j \tag{7.6}$$

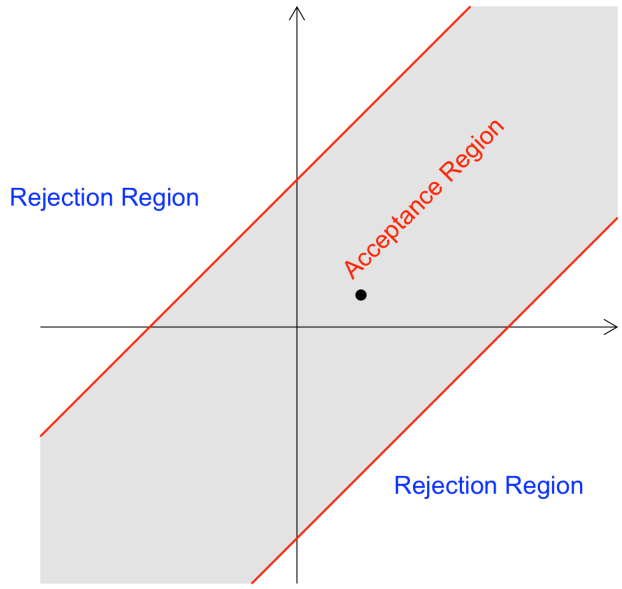

图 7.2: 假设检验问题 (线性组合) 的接受域与拒绝域 (协方差矩阵未知).

【解】 $\boldsymbol{\mu}$ 的置信域为

$$\mathrm{CR} = \left\{ \boldsymbol{\mu} \in \mathbb{R}^3 \;\middle|\; (\boldsymbol{\mu} - \overline{\boldsymbol{x}})^{\mathrm{T}} \mathcal{S}^{-1} (\boldsymbol{\mu} - \overline{\boldsymbol{x}}) \leq \frac{p}{n-p} F_{p,\, n-p}(\alpha) \right\}$$

我们看到

$$n = 10\,, \quad p = 3\,, \quad \alpha = 0.05\,, \quad \overline{\boldsymbol{x}} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}\,, \quad \mathcal{S} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 1 & 4 \end{pmatrix}\,, \quad F_{3,\,7}(0.05) = 4.3468$$

记 $\mathcal{S}$ 的特征为 $\lambda_1, \lambda_2, \lambda_3$. 因为 $\mathcal{S}$ 的特征值为整数，而且

$$|\mathcal{S}| = \lambda_1 \lambda_2 \lambda_3 = 18\,, \quad \mathrm{tr}(\mathcal{S}) = \lambda_1 + \lambda_2 + \lambda_3 = 10$$

由此可得

$$\lambda_1 = 6\,, \quad \lambda_2 = 3\,, \quad \lambda_3 = 1$$

于是，$\mathcal{S}^{-1}$ 的特征值为

$$\xi_1 = \lambda_1^{-1} = \frac{1}{6}\,, \quad \xi_2 = \lambda_2^{-1} = \frac{1}{3}\,, \quad \xi_3 = \lambda_3^{-1} = 1$$

如果我们令 $\boldsymbol{y} \triangleq \boldsymbol{\mu} - \overline{\boldsymbol{x}}$，则有

$$(\boldsymbol{\mu} - \overline{\boldsymbol{x}})^{\mathrm{T}} \mathcal{S}^{-1} (\boldsymbol{\mu} - \overline{\boldsymbol{x}}) = \boldsymbol{y}^{\mathrm{T}} \mathcal{S}^{-1} \boldsymbol{y}$$

二次型 $\boldsymbol{y}^{\mathrm{T}} \mathcal{S}^{-1} \boldsymbol{y}$ 可以通过一个正交变换化为标准形，即有

$$\boldsymbol{y}^{\mathrm{T}} \mathcal{S}^{-1} \boldsymbol{y} = \xi_1 z_1^2 + \xi_2 z_2^2 + \xi_3 z_3^2 = \frac{1}{6} z_1^2 + \frac{1}{3} z_2^2 + z_3^2$$

于是，$\boldsymbol{\mu}$ 的置信域为

$$\mathrm{CR} = \left\{ \boldsymbol{\mu} \in \mathbb{R}^3 \;\middle|\; \frac{1}{6} z_1^2 + \frac{1}{3} z_2^2 + z_3^2 \leq \frac{3}{10-3} \times 4.3468 = 1.8629 \right\}$$

等价地

$$\mathrm{CR} = \left\{ \boldsymbol{\mu} \in \mathbb{R}^3 \;\middle|\; z_1^2 + 2 z_2^2 + 6 z_3^2 \leq 11.1775 \right\}$$

它是 $\mathbb{R}^3$ 中的一个椭球.

(b) [2 分] 计算 μ_1，μ_2，以及 μ_3 的联合置信区间.

【解】 μ_1, μ_2, 以及 μ_3 的联合置信区间为

$$\overline{x}_j - \sqrt{\frac{p}{n-p} F_{p,\, n-p}(\alpha)\; s_{jj}} \leq \mu_j \leq \overline{x}_j + \sqrt{\frac{p}{n-p} F_{p,\, n-p}(\alpha)\; s_{jj}}\,, \quad j = 1,\ 2,\ 3$$

由于

$$\begin{aligned} & p = 3\,,\ n = 10 \\ & \overline{x}_1 = 1\,,\ \overline{x}_2 = 0\,,\ \overline{x}_3 = 2 \\ & s_{11} = 3\,,\ s_{22} = 3\,,\ s_{33} = 4 \\ & \alpha = 0.05\,,\ F_{3,\,7}(0.05) = 4.3468 \end{aligned}$$

因此可得置信度为 95% 的联合置信区间如下

$$\begin{aligned} -1.3641 \leq & \mu_1 \leq 3.3641 \\ -2.3641 \leq & \mu_2 \leq 2.3641 \\ -0.7298 \leq & \mu_3 \leq 4.7298 \end{aligned}$$

(c) [2 分] 可否认为 μ_1 等于 μ_2 与 μ_3 的平均?

【解】 欲检验的假设为

$$H_0:\ \mu_1 = \frac{\mu_2 + \mu_3}{2} \quad \longleftrightarrow \quad H_1:\ \mu_1 \neq \frac{\mu_2 + \mu_3}{2}$$

或者，等价地 $H_0:\ \mathcal{A}\boldsymbol{\mu}=\boldsymbol{a} \longleftrightarrow H_1:\ \mathcal{A}\boldsymbol{\mu}\neq\boldsymbol{a}$，其中

$$\mathcal{A}=\begin{pmatrix}1 & -\frac{1}{2} & -\frac{1}{2}\end{pmatrix}\ ,\quad \boldsymbol{\mu}=\begin{pmatrix}\mu_1\\ \mu_2\\ \mu_3\end{pmatrix}\ ,\quad \boldsymbol{a}=\boldsymbol{0}$$

这是我们在课堂讨论的**检验问题 6**，检验统计量为

$$\frac{n-q}{q}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\mathcal{S}\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)\quad\overset{H_0\ \text{true}}{\sim}\quad F_{q,\ n-q}$$

对于给定的显著水平 α，当该统计量的值大于 $F_{q,\ n-q}(\alpha)$ 时，我们拒绝零假设 H_0 . 因为 $q=1$, $n=10$, $F_{1,9}(0.05)=5.1174$，于是

$$\frac{n-q}{q}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)^{\mathrm{T}}\left(\mathcal{A}\mathcal{S}\mathcal{A}^{\mathrm{T}}\right)^{-1}\left(\mathcal{A}\overline{\boldsymbol{x}}-\boldsymbol{a}\right)=0<5.1174=F_{1,\ 9}(0.05)$$

故，当显著水平 $\alpha=0.05$ 时，我们不能拒绝 H_0，即接受 μ_1 等于 μ_2 与 μ_3 的平均.

6. 对取自两个二元正态分布总体、容量均为 10 的两个独立样本，计算得

$$\overline{\boldsymbol{x}}_1=\begin{pmatrix}3\\ 1\end{pmatrix}\ ,\quad \mathcal{S}_1=\begin{pmatrix}4 & -1\\ -1 & 2\end{pmatrix}\ ,\quad \overline{\boldsymbol{x}}_2=\begin{pmatrix}1\\ 1\end{pmatrix}\ ,\quad \mathcal{S}_2=\begin{pmatrix}2 & -2\\ -2 & 4\end{pmatrix}\tag{7.7}$$

求解以下假设检验问题:

(a) [2 分] $H_0:\ \boldsymbol{\mu}_1=\boldsymbol{\mu}_2 \longleftrightarrow H_1:\ \boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2$.

【解】 这是两个正态总体均值向量的比较问题，由于缺少两个协方差矩阵是否相等的信息，我们需要先对其进行检验，从而有以下检验问题:

$$H_0:\ \Sigma_1=\Sigma_2 \longleftrightarrow H_1:\ \Sigma_1\neq\Sigma_2$$

这是我们在课堂上讨论的**检验问题 9 (协方差矩阵的比较)**，检验统计量为

$$-2\log\lambda=n\log|\mathcal{S}|-\sum_{h=1}^{k}n_h\log|\mathcal{S}_h|\quad\overset{H_0\ \text{true}}{\sim}\quad\chi^2_m$$

其中

$$m=\frac{1}{2}(k-1)p(p+1)\ ,\quad \mathcal{S}=\frac{n_1\mathcal{S}_1+n_2\mathcal{S}_2+\cdots+n_k\mathcal{S}_k}{n_1+n_2+\cdots+n_k}$$

因为

$$n_1=n_2=10\ ,\ k=p=2\ ,\ \mathcal{S}_1=\begin{pmatrix}4 & -1\\ -1 & 2\end{pmatrix}\ ,\ \mathcal{S}_2=\begin{pmatrix}2 & -2\\ -2 & 4\end{pmatrix}$$

所以有

$$m = 3\,, \qquad \mathcal{S} = \begin{pmatrix} 3 & -1.5 \\ -1.5 & 3 \end{pmatrix}\,, \quad n = n_1 + n_2 = 20$$

于是

$$-2\log\lambda = 4.8688 < 7.8147 = \chi^2_3(0.05)$$

显著水平 $\alpha = 0.05$ 时，我们接受零假设 $H_0:\ \Sigma_1 = \Sigma_2$. 从而，两个均值向量的比较就是我们在课堂上讨论的**检验问题 8**. 欲检验的假设为

$$H_0:\ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \ \leftrightarrow\ H_1:\ \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

检验统计量为

$$T = \frac{n_1 n_2\,(n_1 + n_2 - p - 1)}{p\,(n_1 + n_2)^2}\,(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^{\mathrm{T}}\,\mathcal{S}^{-1}\,(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) \quad \overset{H_0 \text{ true}}{\sim} \quad F_{p,\ n_1+n_2-p-1}$$

给定显著水平 α，当检验统计量的值 $T > F_{p,\ n_1+n_2-p-1}(\alpha)$ 时我们拒绝 H_0. 由于

$$\overline{\boldsymbol{x}}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}\,,\ \overline{\boldsymbol{x}}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

则

$$T = 3.7778 > 3.5915 = F_{2,17}(0.05)$$

显著水平 $\alpha = 0.05$ 时，我们拒绝零假设 $H_0:\ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.

(b) [2 分] $H_0:\ \mu_{11} = \mu_{21} \ \longleftrightarrow\ H_1:\ \mu_{11} \neq \mu_{21}$.

【解】 这是一维情形两个正态总体均值的比较问题，检验统计量为

$$T = \frac{|\overline{x}_{11} - \overline{x}_{21}|}{s_w\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

其中

$$s_w^2 = \frac{1}{n_1 + n_2 - 2}\left[(n_1 - 1)\,s_1^2 + (n_2 - 1)\,s_2^2\right] = \frac{1}{10 + 10 - 2}\,(9 \times 4 + 9 \times 2) = 3$$

因为

$$T = \frac{|3 - 1|}{\sqrt{3} \times \sqrt{\dfrac{1}{10} + \dfrac{1}{10}}} = 2.582 > 2.101 = t_{18}(0.025)$$

显著水平 $\alpha = 0.05$ 时，我们拒绝零假设 $H_0:\ \mu_{11} = \mu_{21}$.

(c) [2 分] $H_0:\ \mu_{12} = \mu_{22} \ \longleftrightarrow\ H_1:\ \mu_{12} \neq \mu_{22}$.

【解】 这仍然是一维情形两个正态总体均值的比较问题，检验统计量为

$$T=\frac{|\overline{x}_{12}-\overline{x}_{22}|}{s_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$$

其中

$$s_w^2=\frac{1}{n_1+n_2-2}\left[(n_1-1)\,s_1^2+(n_2-1)\,s_2^2\right]=\frac{1}{10+10-2}\,(9\times 2+9\times 4)=3$$

因为

$$T=\frac{|1-1|}{\sqrt{3}\times\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}}=0<2.101=t_{18}(0.025)$$

显著水平 $\alpha=0.05$ 时，我们接受零假设 $H_0:\ \mu_{12}=\mu_{22}$.

(d) [2 分] 比较上述结果并作简要分析.

【解】 在二维情形中，我们拒绝了零假设 $H_0:\ \boldsymbol{\mu}_1=\boldsymbol{\mu}_2$. 而在一维情形中，我们拒绝了零假设 $H_0:\ \mu_{11}=\mu_{21}$，而接受了零假设 $H_0:\ \mu_{12}=\mu_{22}$. 这说明，我们拒绝两个均值向量相等是因为它们的第一个分量的均值有着显著差异.

7. [2 分] 对于课堂中讨论过的美国公司数据集，利用 $X_1\sim X_6$ 的全部六个变量的观测数据，检验能源行业和制造业的协方差矩阵是否相等.

【解】 虽然本题目的样本容量不够大，关于协方差矩阵的比较课堂中我们仅讨论过检验问题 9，以下我们用该方法求解，欲检验的假设为

$$H_0:\ \Sigma_1=\Sigma_2\longleftrightarrow H_1:\ \text{no constraints}$$

若记 $\mathcal{S}_1$ 是第一个总体 (能源行业) 的样本协方差矩阵，$\mathcal{S}_2$ 是第二个总体 (制造业) 的样本协方差矩阵，而

$$\mathcal{S}=\frac{n_1\,\mathcal{S}_1+n_2\,\mathcal{S}_2}{n_1+n_2}$$

则检验统计量

$$-2\log\lambda=(n_1+n_2)\log|\mathcal{S}|-n_1\,\log|\mathcal{S}_1|-n_2\,\log|\mathcal{S}_2|$$

近似服从自由度为

$$\frac{1}{2}\times(2-1)\times p\times(p+1)=\frac{p(p+1)}{2}$$

的 χ^2 分布. 已知 $p=6$，所以检验统计量 $-2\log\lambda$ 近似服从 χ^2_{21} 分布. 计算得

$$-2\,\log\lambda=175.122$$

查得临界值

$$\chi^2_{21}(0.05)=32.67057$$

因为检验统计量的值 175.122 > 32.67057，所以当检验的显著水平 $\alpha = 0.05$ 时，我们拒绝零假设 $H_0: \Sigma_1 = \Sigma_2$，认为两个总体的协方差矩阵不同. 计算的 R 代码如下:

```
test.chi = (n1 + n2) * log(det(S)) - n1 * log(det(sx.var)) -
        n2 * log(det(sy.var))
test.chi # 检验统计量的值
qchisq(0.05, 21, lower.tail = FALSE) # 检验的临界值
pchisq(test.chi, 21, lower.tail = FALSE) # 检验的 p 值
```

```
> summary(X_lm)

Call:
lm(formula = Diagonal ~ Length + Left + Right + Bottom + Top,
    data = X)

Residuals:
     Min       1Q   Median       3Q      Max
-1.16606 -0.28914 -0.01916  0.31843  1.21257

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 47.34541   34.93498   1.355  0.17859
Length       0.31930    0.14831   2.153  0.03388 *
Left        -0.50683    0.24829  -2.041  0.04403 *
Right        0.63375    0.20208   3.136  0.00229 **
Bottom       0.33250    0.05963   5.576 2.35e-07 ***
Top          0.31793    0.10391   3.060  0.00289 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4714 on 94 degrees of freedom
Multiple R-squared:  0.322,     Adjusted R-squared:  0.2859
F-statistic: 8.927 on 5 and 94 DF,  p-value: 5.757e-07
```

图 7.3: 伪钞数据集的线性模型.

8. 对于瑞士银行钞票数据集 (mclust 包中的 banknote 数据集) 当中的伪钞数据，我们想知道钞票对角线的长度 X_6 是否可以由 $X_1 \sim X_5$ 的一个线性模型来预测.

   (a) [2 分] 拟合线性模型，给出拟合结果.

   【解】 利用《回归分析》中已掌握的方法，我们拟合对角线长度 X_6 关于 $X_1 \sim X_5$ 的一个线性模型，相应的 R 代码如下:

```
rm(list = ls(all = TRUE))
library(mclust)
X = subset(banknote, Status == "counterfeit") # 伪钞数据子集
X_lm = lm(Diagonal ~ Length + Left + Right+ Bottom + Top, data = X) # 模型拟合
```

```
summary(X_lm) # 拟合结果
```

结果如图 7.3 所示.

(b) [2 分] 检验回归系数是否显著不等于零 (取显著水平 $\alpha = 0.05$).

【解】 欲检验的零假设为

$$H_0 : \beta_{\text{Length}} = \beta_{\text{Left}} = \beta_{\text{Right}} = \beta_{\text{Bottom}} = \beta_{\text{Top}} = 0$$

因为检验的 p 值 $= 5.757 \times 10^{-7} < 0.05$，所以当检验的显著水平 $\alpha = 0.05$ 时我们拒绝零假设，即认为回归系数显著不等于零.

Chapter 8

第 8 周作业参考答案

1. 利用数据矩阵的因子分解方法，简要分析瑞士银行钞票数据集 (mclust 包中的 banknote 数据集).

   (a) [2 分] 利用 R 中的 scale() 函数对数据进行标准化，将标准化之后的数据集记为 $\mathcal{X}$.

   【解】 利用以下的 R 代码读入瑞士银行钞票数据集 banknote，提取我们要分析的数据，并作标准化处理，代码还验证了经标准化之后，各个变量的样本均值为 0、样本标准差为 1.

```
> library(mclust)
> data(banknote) # 读入数据
> x = banknote[, 2:7] # 提取要分析的数据子集
> head(x) # 数据子集的前6行
  Length Left Right Bottom  Top Diagonal
1    215  131   131    9.0  9.7      141
2    215  130   130    8.1  9.5      142
3    215  130   130    8.7  9.6      142
4    215  130   130    7.5 10.4      142
5    215  130   130   10.4  7.7      142
6    216  131   130    9.0 10.1      141
> X = scale(x, center = TRUE, scale = TRUE) # 标准化
> head(X) # 标准化数据集的前6行
     Length Left Right Bottom   Top Diagonal
[1,]  -0.25  2.4  2.83  -0.29 -1.18     0.45
[2,]  -0.79 -1.2 -0.63  -0.91 -1.43     1.06
[3,]  -0.25 -1.2 -0.63  -0.50 -1.31     1.49
[4,]  -0.25 -1.2 -0.88  -1.33 -0.31     1.32
[5,]   0.28 -1.4 -0.63   0.68 -3.67     1.14
[6,]   2.14  1.9  1.35  -0.29 -0.69     0.80
> round(apply(X, 2, mean), digits = 3) # 标准化后各变量的均值为0
  Length     Left    Right   Bottom      Top Diagonal
       0        0        0        0        0        0
> apply(X, 2, sd) # 标准化后各变量的标准差为1
  Length     Left    Right   Bottom      Top Diagonal
```

```
       1        1        1        1        1        1
```

(b) [2 分] 为处理更为规范，我们将 $\mathcal{X}$ 的所有元素除以 $\sqrt{n-1}$，其中 n 为样本容量，得到的数据矩阵记为 $\mathcal{Y}$，现在 $\mathcal{Y}$ 即是我们要分析的数据矩阵.

【解】 这是为了计算样本相关矩阵作准备，相应的代码如下：

```
> n = dim(X)[1] # 数据集的行数
> n
[1] 200
> Y = as.matrix(X / sqrt(n-1)) # X 的所有元素除以 sqrt{n-1}
> head(Y) # 前6行数据
     Length   Left  Right Bottom    Top Diagonal
[1,] -0.018  0.172  0.201 -0.020 -0.084    0.032
[2,] -0.056 -0.083 -0.045 -0.065 -0.102    0.075
[3,] -0.018 -0.083 -0.045 -0.035 -0.093    0.106
[4,] -0.018 -0.083 -0.063 -0.094 -0.022    0.093
[5,]  0.020 -0.102 -0.045  0.048 -0.260    0.081
[6,]  0.151  0.133  0.095 -0.020 -0.049    0.056
```

(c) [2 分] 求矩阵 $\mathcal{R} = \mathcal{Y}^{\mathrm{T}}\mathcal{Y}$ 的特征值及其对应的单位特征向量. 矩阵 $\mathcal{R} = \mathcal{Y}^{\mathrm{T}}\mathcal{Y}$ 是原始数据的相关矩阵.

【解】 相关矩阵可以通过上述计算过程求得，也可以利用 R 中的 cor() 函数直接由原始数据求得. 代码如下：

```
> R = t(Y) %*% Y # 相关矩阵
> round(R, digits = 3)
          Length  Left Right Bottom    Top Diagonal
Length     1.000  0.23  0.15  -0.19 -0.061     0.19
Left       0.231  1.00  0.74   0.41  0.362    -0.50
Right      0.152  0.74  1.00   0.49  0.401    -0.52
Bottom    -0.190  0.41  0.49   1.00  0.142    -0.62
Top       -0.061  0.36  0.40   0.14  1.000    -0.59
Diagonal   0.194 -0.50 -0.52  -0.62 -0.594     1.00
> # 相关矩阵亦可用函数 cor() 计算
> R = cor(x)
> round(R, digits = 3)
          Length  Left Right Bottom    Top Diagonal
Length     1.000  0.23  0.15  -0.19 -0.061     0.19
Left       0.231  1.00  0.74   0.41  0.362    -0.50
Right      0.152  0.74  1.00   0.49  0.401    -0.52
Bottom    -0.190  0.41  0.49   1.00  0.142    -0.62
Top       -0.061  0.36  0.40   0.14  1.000    -0.59
Diagonal   0.194 -0.50 -0.52  -0.62 -0.594     1.00
```

得相关矩阵

$$\mathcal{R} = \begin{pmatrix} 1.000 & 0.231 & 0.152 & -0.190 & -0.061 & 0.194 \\ 0.231 & 1.000 & 0.743 & 0.414 & 0.362 & -0.503 \\ 0.152 & 0.743 & 1.000 & 0.487 & 0.401 & -0.516 \\ -0.190 & 0.414 & 0.487 & 1.000 & 0.142 & -0.623 \\ -0.061 & 0.362 & 0.401 & 0.142 & 1.000 & -0.594 \\ 0.194 & -0.503 & -0.516 & -0.623 & -0.594 & 1.000 \end{pmatrix}$$

对相关矩阵 $\mathcal{R}$ 作谱分解，可以利用 R 软件中的函数 eigen() 实现，代码如下：

```
> R_SD = eigen(R) # 相关矩阵R的谱分解
> Lambda = R_SD$values # 特征值
> round(Lambda, digits = 3)
[1] 2.95 1.28 0.87 0.45 0.27 0.19
> Gamma = R_SD$vectors # 特征向量
> round(Gamma, digits = 3)
       [,1]   [,2]   [,3]  [,4]   [,5]   [,6]
[1,] -0.007  0.815 -0.018  0.57 -0.059 -0.031
[2,]  0.468  0.342  0.103 -0.40  0.639  0.298
[3,]  0.487  0.252  0.123 -0.43 -0.614 -0.349
[4,]  0.407 -0.266  0.584  0.40 -0.215  0.462
[5,]  0.368 -0.091 -0.788  0.11 -0.220  0.419
[6,] -0.493  0.274  0.114 -0.39 -0.340  0.632
```

得特征值与特征向量如下：

$$\boldsymbol{\Lambda} = \begin{pmatrix} 2.946 & & & & & \\ & 1.278 & & & & \\ & & 0.869 & & & \\ & & & 0.450 & & \\ & & & & 0.269 & \\ & & & & & 0.189 \end{pmatrix}$$

$$\boldsymbol{\Gamma} = \begin{pmatrix} -0.007 & 0.815 & -0.018 & 0.575 & -0.059 & -0.031 \\ 0.468 & 0.342 & 0.103 & -0.395 & 0.639 & 0.298 \\ 0.487 & 0.252 & 0.123 & -0.430 & -0.614 & -0.349 \\ 0.407 & -0.266 & 0.584 & 0.404 & -0.215 & 0.462 \\ 0.368 & -0.091 & -0.788 & 0.110 & -0.220 & 0.419 \\ -0.493 & 0.274 & 0.114 & -0.392 & -0.340 & 0.632 \end{pmatrix}$$

(d) [2 分] 因子变量是数据矩阵 $\mathcal{Y}$ 的六个变量 $Y_1 \sim Y_6$ 的线性组合，写出前两个因子变量 Z_1 和 Z_2 的表达式.

【解】 前两个因子变量 Z_1 和 Z_2 是变量 $Y_1 \sim Y_6$ 的标准线性组合，各变量前的权重是对应特征向量的值，因此有

$$Z_1 = -0.007\,Y_1 + 0.468\,Y_2 + 0.487\,Y_3 + 0.407\,Y_4 + 0.368\,Y_5 - 0.493\,Y_6$$
$$Z_2 = 0.815\,Y_1 + 0.342\,Y_2 + 0.252\,Y_3 - 0.266\,Y_4 - 0.091\,Y_5 + 0.274\,Y_6$$

(e) [2 分] 计算前两个因子变量 Z_1 和 Z_2 对应的特征值之和占所有特征值之和的比例.

【解】 计算可得前两个因子变量 Z_1 和 Z_2 对应的特征值之和占所有特征值之和的比例为 70.39%，代码如下：

```
> Rate = cumsum(Lambda) / sum(Lambda) # 累计贡献率
> round(Rate, digits = 4)
[1] 0.49 0.70 0.85 0.92 0.97 1.00
> Rate[2] # 前两个主方向的累计贡献率
[1] 0.7
```

(f) [2 分] 计算观测数据 (行) 在前两个因子变量 Z_1 和 Z_2 上的坐标值.

【解】 观测数据 (行) 在前两个因子变量 Z_1 和 Z_2 上的坐标值为 $\boldsymbol{z}_1 = \mathcal{Y}\boldsymbol{\gamma}_1$，$\boldsymbol{z}_2 = \mathcal{Y}\boldsymbol{\gamma}_2$. 代码如下：

```
> options(digits = 3)
> Z_1 = Y %*% Gamma[, 1] # 在第一个因子变量上的坐标
> Z_2 = Y %*% Gamma[, 2] # 在第二个因子变量上的坐标
> head(data.frame(z1 = Z_1, z2 = Z_2)) # 前6个观测值在第一、二个因子变量上的坐标
       z1       z2
1  0.1236  0.11673
2 -0.1608 -0.03810
3 -0.1610 -0.00761
4 -0.1615 -0.00620
5 -0.1861  0.00277
6  0.0536  0.21841
```

(g) [2 分] 作观测数据在前两个因子变量 Z_1 和 Z_2 上的散点图，将真钞与伪钞的数据点分别用不同的颜色表示，你能看到什么现象.

【解】 观测数据在前两个因子变量 Z_1 和 Z_2 上的散点图如图 8.1 所示，从中我们可以发现，前两个因子变量 Z_1 和 Z_2 较好地捕捉到了数据当中的信息，在二维平面上将真钞 (genuine) 与伪钞 (conterfeit) 清晰地呈现在有明显边界的不同区域当中. 绘制图 8.1 的代码如下：

```
Z = data.frame(z1 = Z_1, z2 = Z_2, Status = banknote$Status) # 绘图所需数据集
Fig_1 = ggplot(data = Z, aes(x = z1, y = z2, col = Status)) +
  geom_point(size = 2) +
```

```
  geom_hline(yintercept = 0,
             colour = "grey50",
             linewidth = 0.5, linetype = 2) +
  geom_vline(xintercept = 0,
             colour = "grey50",
             linewidth = 0.5, linetype = 2)
Fig_1
```

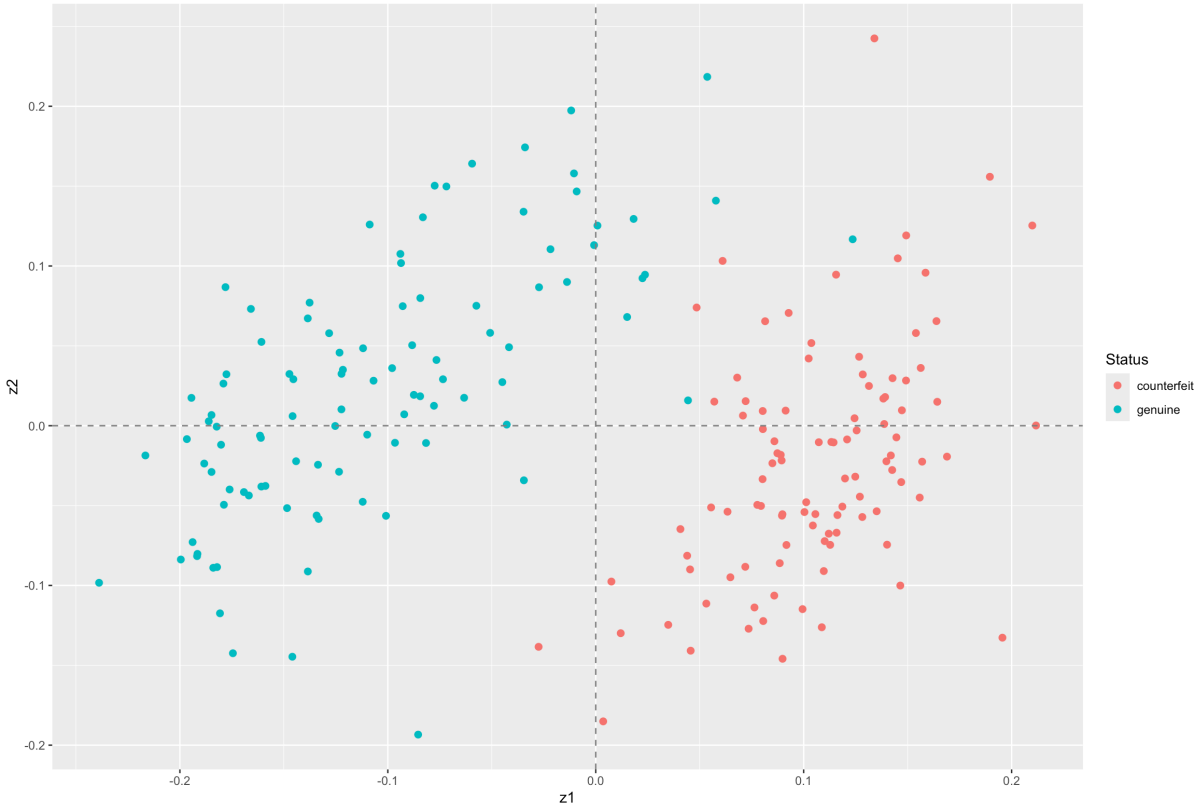

图 8.1: 钞票数据在前两个因子变量 Z_1, Z_2 上的散点图.

(h) [2 分] 计算变量数据 (列) 在前两个因子变量 W_1 和 W_2 上的坐标值.

【解】 变量数据 (列) 在前两个因子变量 W_1 和 W_2 上的坐标为 $\boldsymbol{w}_1 = \sqrt{\lambda_1}\boldsymbol{\gamma}_1$, $\boldsymbol{w}_2 = \sqrt{\lambda_2}\boldsymbol{\gamma}_2$. 代码如下:

```
> W_1 = sqrt(Lambda[1]) * Gamma[, 1] # 在第1个因子变量上的坐标
> W_2 = sqrt(Lambda[2]) * Gamma[, 2] # 在第2个因子变量上的坐标
> data.frame(w1 = W_1, w2 = W_2)
      w1     w2
1 -0.012  0.922
2  0.803  0.387
3  0.835  0.285
4  0.698 -0.301
5  0.631 -0.103
6 -0.847  0.310
```

(i) [2 分] 作变量点在前两个因子变量 W_1 和 W_2 上的散点图，你能看到什么现象.

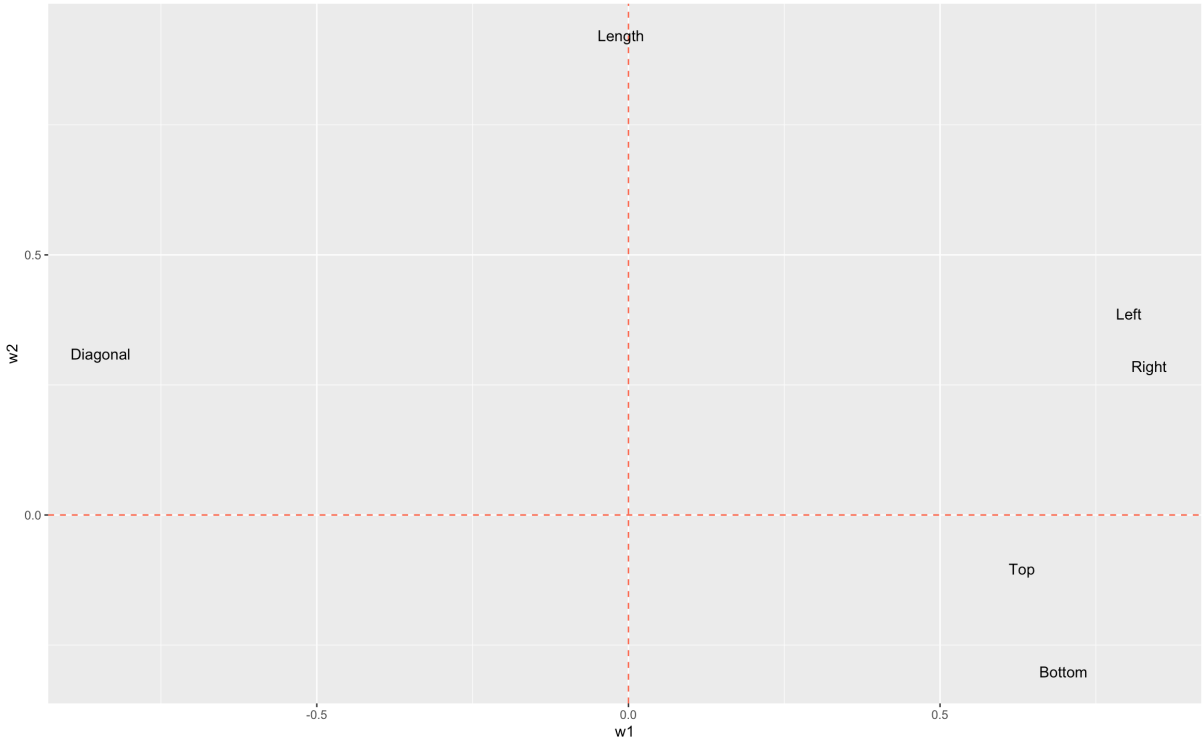

图 8.2: 变量点在前两个因子变量 W_1, W_2 上的散点图.

【解】 变量点在前两个因子变量 W_1 和 W_2 上的散点图如图 8.2 所示，从中我们可以发现，第 1 个主方向 W_1 主要受变量 X_6 (Diagonal)、X_2 (Left) 以及 X_3 (Right) 的影响，它反映钞票对角线长度与左、右两边框宽度之和的差异. 第 2 个主方向 W_2 主要受变量 X_1 (Length) 的影响，它主要反映了钞票的长度. 绘制图 8.2 的代码如下:

```
W = data.frame(w1 = W_1, w2 = W_2)
Name = c("Length", "Left", "Right", "Bottom", "Top", "Diagonal")
Fig_2 = ggplot(data = W, aes(x = w1, y = w2)) +
  geom_hline(yintercept = 0,
             colour = "tomato",
             linewidth = 0.5, linetype = 2) +
  geom_vline(xintercept = 0,
             colour = "tomato",
             linewidth = 0.5, linetype = 2) +
  geom_text(label = Name)
Fig_2
```

Chapter 9

第 9 周作业参考答案

1. 美国犯罪数据集 (uscrime.csv) 的主成分分析.

   该数据集由 11 个变量的 50 个观测值组成，它提供了 1985 年美国 50 个州报告的犯罪数量及其它一些信息，我们利用其中 7 个变量 $(X_3, \ldots, X_9)$ 的数据来作主成分分析. 数据集当中各个变量的定义如表 9.1 所示.

表 9.1: 美国犯罪数据集的变量含义.

X_1	land area	占地面积
X_2	popu 1985	1985 年的人口数量
X_3	murder	凶杀
X_4	rape	强奸
X_5	robbery	抢劫
X_6	assault	人身袭击
X_7	burglary	入室盗窃
X_8	larceny	偷盗
X_9	autotheft	汽车盗窃
X_{10}	region	美国各州所处地区
X_{11}	division	美国各州所属分部

   (a) [2 分] 读入原始数据，提取拟分析的数据集 $\mathscr{X}$.

   【解】 读入原始数据并提取拟分析的数据集，相应的 R 代码为:

```
library(data.table)
setwd("~/Desktop/2026_Multivariate Statistical Analysis/R_Code/Data")
x = fread("uscrime.csv", header = TRUE) # 读入原始数据集
```

```
X = x[, 4:10] # 提取拟分析的数据子集
head(X) # 数据子集的前6行
```

```
> head(X) # 数据子集的前6行
   murder  rape robbery assault burglary larceny autotheft
    <num> <num>   <num>   <int>    <int>   <int>     <int>
1:    1.5   7.0    12.6      62      562    1055       146
2:    2.0   6.0    12.1      36      566     929       172
3:    1.3  10.3     7.6      55      731     969       124
4:    3.5  12.0    99.5      88     1134    1531       878
5:    3.2   3.6    78.3     120     1019    2186       859
6:    3.5   9.1    70.4      87     1084    1751       484
```

图 9.1: 拟分析的数据子集的前 6 行.

(b) [2 分] 由于数据集 $\mathscr{X}$ 中各变量的数据尺度有较大差异，我们先对其作标准化处理，将标准化后的数据集记为 $\mathscr{Y}$.

【解】 对数据作标准化处理的 R 代码为:

```
Y = scale(X, center = TRUE, scale = TRUE) # 对数据集 X 作标准化
round(head(Y), digits = 3) # 标准化后数据集的前 6 行
```

```
> round(head(Y), digits = 3) # 标准化后数据集的前 6 行
     murder   rape robbery assault burglary larceny autotheft
[1,] -1.392 -1.173  -0.975  -1.077   -1.021  -1.252    -1.111
[2,] -1.262 -1.309  -0.980  -1.458   -1.010  -1.429    -0.981
[3,] -1.444 -0.723  -1.030  -1.180   -0.553  -1.373    -1.222
[4,] -0.873 -0.492  -0.022  -0.696    0.563  -0.581     2.556
[5,] -0.951 -1.635  -0.255  -0.226    0.244   0.341     2.461
[6,] -0.873 -0.887  -0.341  -0.710    0.424  -0.271     0.582
```

图 9.2: 标准化数据子集的前 6 行.

(c) [2 分] 计算数据集 $\mathscr{Y}$ 的样本协方差矩阵 $\mathscr{S}$，并与数据集 $\mathscr{X}$ 的相关矩阵 $\mathscr{R}$ 进行比较.

【解】 相应的 R 代码如下:

```
S = cov(Y) # 标准化数据集 Y 的协方差矩阵
round(S, digits = 3)
R = cor(X) # 原始数据集 X 的相关矩阵
round(R, digits = 3)
```

如图9.3所示，我们可以看到，标准化数据集 $\mathscr{Y}$ 的样本协方差矩阵 $\mathscr{S}$ 就是初始数据集 $\mathscr{X}$ 的相关矩阵 $\mathscr{R}$.

(d) [2 分] 对相关矩阵 $\mathscr{R}$ 作谱分解 $\mathscr{R} = \Gamma \Lambda \Gamma^{\mathrm{T}}$，给出谱分解的结果并作验证运算.

【解】 相应的 R 代码如下:

```
> Jordan_R = eigen(R) # R 的谱分解
> lambda = Jordan_R$values # R 的特征值
```

```
> S = cov(Y) # 标准化数据集 Y 的协方差矩阵
> round(S, digits = 3)
          murder  rape robbery assault burglary larceny autotheft
murder     1.000 0.520   0.341   0.813    0.277   0.065     0.110
rape       0.520 1.000   0.551   0.696    0.680   0.601     0.441
robbery    0.341 0.551   1.000   0.563    0.622   0.436     0.617
assault    0.813 0.696   0.563   1.000    0.521   0.317     0.330
burglary   0.277 0.680   0.622   0.521    1.000   0.801     0.700
larceny    0.065 0.601   0.436   0.317    0.801   1.000     0.555
autotheft  0.110 0.441   0.617   0.330    0.700   0.555     1.000
> R = cor(X) # 原始数据集 X 的相关矩阵
> round(R, digits = 3)
          murder  rape robbery assault burglary larceny autotheft
murder     1.000 0.520   0.341   0.813    0.277   0.065     0.110
rape       0.520 1.000   0.551   0.696    0.680   0.601     0.441
robbery    0.341 0.551   1.000   0.563    0.622   0.436     0.617
assault    0.813 0.696   0.563   1.000    0.521   0.317     0.330
burglary   0.277 0.680   0.622   0.521    1.000   0.801     0.700
larceny    0.065 0.601   0.436   0.317    0.801   1.000     0.555
autotheft  0.110 0.441   0.617   0.330    0.700   0.555     1.000
```

图 9.3: 标准化数据集的样本协方差矩阵与初始数据集的相关矩阵.

```
> round(lambda, digits = 3)
[1] 4.077 1.432 0.631 0.340 0.248 0.140 0.132
> Lambda = diag(lambda) # 特征值构成的对角矩阵 Lambda
> round(Lambda, digits = 3)
      [,1]  [,2]  [,3] [,4]  [,5] [,6]  [,7]
[1,] 4.077 0.000 0.000 0.00 0.000 0.00 0.000
[2,] 0.000 1.432 0.000 0.00 0.000 0.00 0.000
[3,] 0.000 0.000 0.631 0.00 0.000 0.00 0.000
[4,] 0.000 0.000 0.000 0.34 0.000 0.00 0.000
[5,] 0.000 0.000 0.000 0.00 0.248 0.00 0.000
[6,] 0.000 0.000 0.000 0.00 0.000 0.14 0.000
[7,] 0.000 0.000 0.000 0.00 0.000 0.00 0.132
> Gamma = as.matrix(Jordan_R$vectors) # R 的特征向量构成的矩阵 Gamma
> round(Gamma, digits = 3)
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]
[1,] -0.276  0.644 -0.010 -0.329  0.203  0.100  0.591
[2,] -0.421  0.116 -0.360  0.296 -0.759 -0.065  0.107
[3,] -0.387 -0.046  0.604  0.645  0.190  0.069  0.161
[4,] -0.388  0.456  0.011 -0.067  0.136  0.100 -0.780
[5,] -0.436 -0.257 -0.155 -0.144  0.292 -0.783  0.027
[6,] -0.360 -0.401 -0.508  0.048  0.360  0.561  0.069
[7,] -0.354 -0.366  0.472 -0.601 -0.337  0.208 -0.012
> # 验证谱分解的结果: R = Gamma * Lambda * t(Gamma)
> round(Gamma %*% Lambda %*% t(Gamma), digits = 3)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
```

```
[1,] 1.000 0.520 0.341 0.813 0.277 0.065 0.110
[2,] 0.520 1.000 0.551 0.696 0.680 0.601 0.441
[3,] 0.341 0.551 1.000 0.563 0.622 0.436 0.617
[4,] 0.813 0.696 0.563 1.000 0.521 0.317 0.330
[5,] 0.277 0.680 0.622 0.521 1.000 0.801 0.700
[6,] 0.065 0.601 0.436 0.317 0.801 1.000 0.555
[7,] 0.110 0.441 0.617 0.330 0.700 0.555 1.000
> round(R, digits = 3) # 相关矩阵 R
          murder  rape robbery assault burglary larceny autotheft
murder     1.000 0.520   0.341   0.813    0.277   0.065     0.110
rape       0.520 1.000   0.551   0.696    0.680   0.601     0.441
robbery    0.341 0.551   1.000   0.563    0.622   0.436     0.617
assault    0.813 0.696   0.563   1.000    0.521   0.317     0.330
burglary   0.277 0.680   0.622   0.521    1.000   0.801     0.700
larceny    0.065 0.601   0.436   0.317    0.801   1.000     0.555
autotheft  0.110 0.441   0.617   0.330    0.700   0.555     1.000
```

谱分解：$\mathscr{R} = \varGamma \varLambda \varGamma^{\mathrm{T}}$，其中

$$\varLambda = \begin{pmatrix} 4.077 & & & & & & \\ & 1.432 & & & & & \\ & & 0.631 & & & & \\ & & & 0.340 & & & \\ & & & & 0.248 & & \\ & & & & & 0.140 & \\ & & & & & & 0.132 \end{pmatrix}$$

$$\varGamma = \begin{pmatrix} -0.276 & 0.644 & -0.010 & -0.329 & 0.203 & 0.100 & 0.591 \\ -0.421 & 0.116 & -0.360 & 0.296 & -0.759 & -0.065 & 0.107 \\ -0.387 & -0.046 & 0.604 & 0.645 & 0.190 & 0.069 & 0.161 \\ -0.388 & 0.456 & 0.011 & -0.067 & 0.136 & 0.100 & -0.780 \\ -0.436 & -0.257 & -0.155 & -0.144 & 0.292 & -0.783 & 0.027 \\ -0.360 & -0.401 & -0.508 & 0.048 & 0.360 & 0.561 & 0.069 \\ -0.354 & -0.366 & 0.472 & -0.601 & -0.337 & 0.208 & -0.012 \end{pmatrix}$$

如图9.4所示，我们验证了谱分解的结果.

(e) [2 分] 画碎石图，问各个主成分的贡献率是多少？前 2 个主成分的累积贡献率是多少，前 3 个主成分的累积贡献率又是多少.

【解】 碎石图如图 9.5 所示，各 NPC 的贡献率以及累积贡献率如表 9.2 所示.

```
> round(Gamma %*% Lambda %*% t(Gamma), digits = 3)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
[1,] 1.000 0.520 0.341 0.813 0.277 0.065 0.110
[2,] 0.520 1.000 0.551 0.696 0.680 0.601 0.441
[3,] 0.341 0.551 1.000 0.563 0.622 0.436 0.617
[4,] 0.813 0.696 0.563 1.000 0.521 0.317 0.330
[5,] 0.277 0.680 0.622 0.521 1.000 0.801 0.700
[6,] 0.065 0.601 0.436 0.317 0.801 1.000 0.555
[7,] 0.110 0.441 0.617 0.330 0.700 0.555 1.000
> round(R, digits = 3) # 相关矩阵 R
          murder  rape robbery assault burglary larceny autotheft
murder     1.000 0.520   0.341   0.813    0.277   0.065     0.110
rape       0.520 1.000   0.551   0.696    0.680   0.601     0.441
robbery    0.341 0.551   1.000   0.563    0.622   0.436     0.617
assault    0.813 0.696   0.563   1.000    0.521   0.317     0.330
burglary   0.277 0.680   0.622   0.521    1.000   0.801     0.700
larceny    0.065 0.601   0.436   0.317    0.801   1.000     0.555
autotheft  0.110 0.441   0.617   0.330    0.700   0.555     1.000
```

图 9.4: 相关矩阵谱分解的验证.

表 9.2: 各 NPC 的贡献率与累积贡献率.

特征值	4.077	1.432	0.631	0.340	0.248	0.140	0.132
贡献率	0.582	0.205	0.090	0.049	0.035	0.020	0.019
累积贡献率	0.582	0.787	0.877	0.926	0.961	0.981	1.000

我们看到，前 2 个主成分的累积贡献率为 78.7%，前 3 个主成分的累积贡献率达到了 87.7%. 作碎石图与计算贡献率的代码如下：

```
> plot(lambda, axes = FALSE, pch = 16, type = "b", xlab = '', ylab = '',
+      xlim = c(0, 8), ylim = c(0, 5), main = 'Scree Plot')
> arrows(0, 0, 0, 5, length = 0.1)
> arrows(0, 0, 8, 0, length = 0.1)
> axis(1, at = 1:7, pos = 0)
> axis(2, at = 1:4, pos = 0)

> round(lambda / sum(lambda), digits = 3) # 贡献率
[1] 0.582 0.205 0.090 0.049 0.035 0.020 0.019
> round(cumsum(lambda) / sum(lambda), digits = 3) # 贡献率与累积贡献率
[1] 0.582 0.787 0.877 0.926 0.961 0.981 1.000
```

(f) [2 分] 计算各变量可以用前 2 个主成分解释的比例.

【解】 计算变量 X_i 与前两个 NPC (Z_1, Z_2) 的相关系数 $r_{X_i Z_j} = \sqrt{\ell_j}\, g_{ij}$. 变量 X_i 可由前两个 NPC 解释的比例即为 $r^2_{X_i Z_1} + r^2_{X_i Z_2}$. 结果见表 9.3，其中平方和列即为各个变量可以用前两个 NPC 解释的比例. 计算的 R 代码如下：

```
> # 前两大特征值对应的特征向量
```

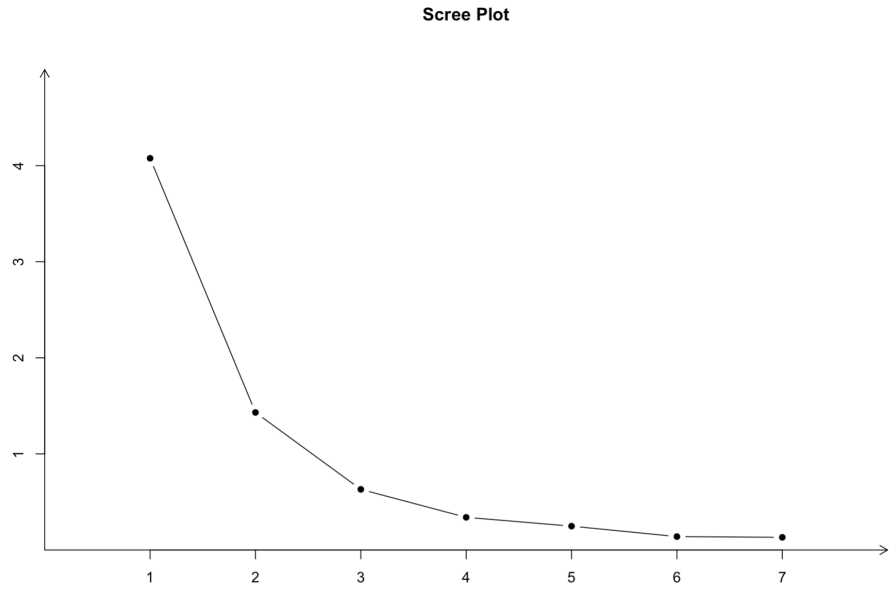

图 9.5: 碎石图.

```
> a = Gamma[, 1:2]
> round(a, digits = 3)
       [,1]   [,2]
[1,] -0.276  0.644
[2,] -0.421  0.116
[3,] -0.387 -0.046
[4,] -0.388  0.456
[5,] -0.436 -0.257
[6,] -0.360 -0.401
[7,] -0.354 -0.366
> # 前两大特征值开方的矩阵
> b = sqrt(matrix(lambda[1:2], nrow(a), ncol(a), byrow = TRUE))
> round(b, digits = 3)
      [,1]  [,2]
[1,] 2.019 1.197
[2,] 2.019 1.197
[3,] 2.019 1.197
[4,] 2.019 1.197
[5,] 2.019 1.197
[6,] 2.019 1.197
[7,] 2.019 1.197
> # 各变量与前两个NPC的相关系数
> w = a * b
> round(w, digits = 3)
       [,1]   [,2]
[1,] -0.557  0.771
```

```
[2,] -0.851  0.139
[3,] -0.782 -0.055
[4,] -0.784  0.546
[5,] -0.881 -0.308
[6,] -0.728 -0.480
[7,] -0.714 -0.438
> # 各变量与前两个NPC相关系数的平方和
> c = as.matrix(apply(w^2, 1, sum))
> round(c, digits = 4)
       [,1]
[1,] 0.9054
[2,] 0.7432
[3,] 0.6150
[4,] 0.9120
[5,] 0.8709
[6,] 0.7595
[7,] 0.7023
> # 将上述计算结果放在一个数据框中
> NPC = data.frame(NPC1 = w[, 1], NPC2 = w[, 2], SS = c[, 1])
> rownames(NPC) = colnames(x)[3:9]
> round(NPC, digits = 3)
            NPC1   NPC2    SS
popu 1985 -0.557  0.771 0.905
murder    -0.851  0.139 0.743
rape      -0.782 -0.055 0.615
robbery   -0.784  0.546 0.912
assault   -0.881 -0.308 0.871
burglary  -0.728 -0.480 0.759
larceny   -0.714 -0.438 0.702
```

(g) [2 分] 作变量在前 2 个主成分平面上的散点图，对结果进行解释.

【解】 变量在前两个 NPC 平面上的投影如图 9.6 所示，从水平方向 (第一主成分) 看，结合变量与第一主成分的相关系数的取值，我们除变量 murder 外，其余变量与第一主成分的相关系数的绝对值均大于 0.7、符号相同，且 rape (强奸)、burglary (入室盗窃) 与第一主成分相关系数的绝对值都超过了 0.85，因此，第一主成分可以看作是一个“综合犯罪变量”，反映 rape (强奸)、burglary (入室盗窃) 的程度要略大一些. 第一主成分取值越小，说明该州的综合犯罪因素越强，第一主成分取值越大，则该州的综合犯罪因素就小一些，安全性相对较好.

从图 9.6 纵向 (第二主成分) 看，结合变量与第二主成分的相关系数的取值，我们发现第二主成分主要受变量 murder (谋杀) 和 assault (人身侵犯)、larceny (偷盗) 和 autotheft (汽车偷盗) 的影响，且 murder (谋杀) 和 assault (人身侵犯) 与第二主成分的相关系数符号为正，larceny (偷盗) 和 autotheft (汽车偷盗) 与第二主成分的相关系数符号为负，因此，第二主成分可以看作是一个“人身攻击——盗窃变量”，主要反映两者的差异. 第二主成分取值越大，说明该州的人

表 9.3: 各变量可以用前 2 个主成分解释的比例.

变量	NPC1	NPC2	平方和
murder	−0.557	0.771	0.905
rape	−0.851	0.139	0.743
robbery	−0.782	−0.055	0.615
assault	−0.784	0.546	0.912
burglary	−0.881	−0.308	0.871
larceny	−0.728	−0.480	0.759
autotheft	−0.714	−0.438	0.702

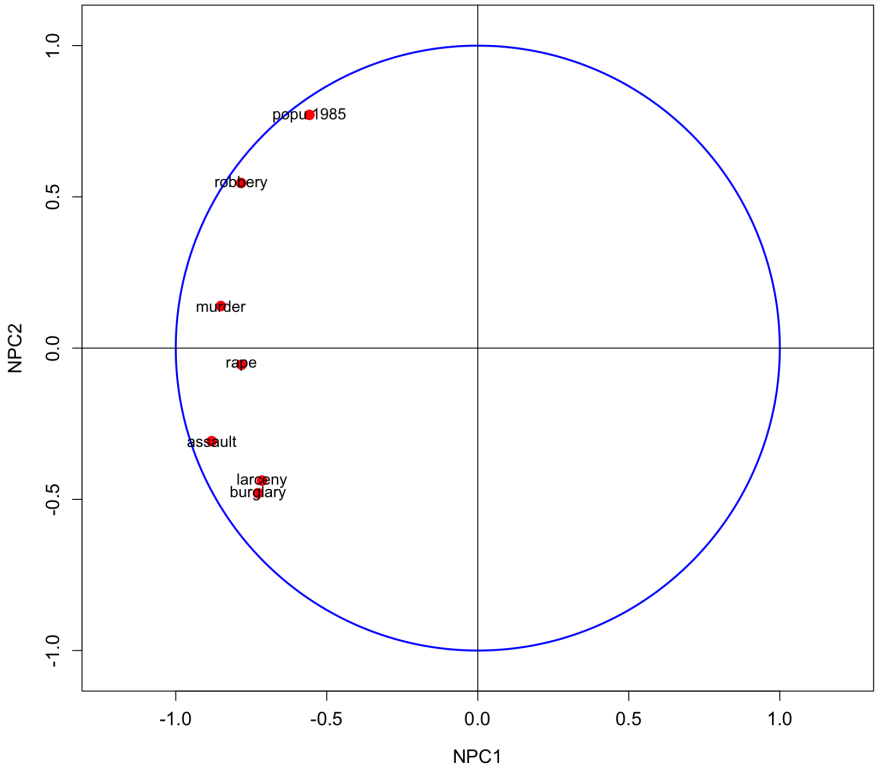

图 9.6: 变量在前两个 NPC 平面上的散点图.

身攻击现象越严重、但偷盗案件较少。第二主成分取值越小，表明该州的偷盗现象较多，但是人身攻击类案件较低.

从图 9.6 中各变量的相对位置我们还可以发现，变量 larceny (偷盗) 和 autotheft (汽车偷盗) 夹角很小，说明它们之间呈强的正相关. 变量 rape (强奸) 与 robbery (抢劫) 的夹角也很小，说明它们之间也呈现较强的正相关. 变量 murder (谋杀) 与变量 larceny (偷盗) 和 autotheft (汽车偷盗) 的夹角几乎垂直，说明这两组变量之间几乎不相关.

作变量在前两个 NPC 平面上投影的散点图代码如下：

```
# 变量在前 2 个主成分平面上的散点图
ucircle = cbind(cos((0:360)/180 * pi), sin((0:360)/180 * pi))
plot(ucircle, type = "l", lty = "solid", col = "blue",
     xlim = c(-1.05, 1.05), ylim = c(-1.05, 1.05),
     xlab = "NPC1", ylab = "NPC2",
     cex.lab = 1.2, cex.axis = 1.2, cex.main = 1.8,
     lwd = 2, asp = 1)
abline(h = 0, v = 0)
label = rownames(NPC)
points(NPC[, 1:2], pch = 16, col = 'pink', cex = 2)
text(NPC[, 1:2], label)
```

(h) [2 分] 作每个州的观测数据在前 2 个主成分平面上的散点图，能否看出美国四个地区存在不同？各州所在地区由变量 X_{10} 提供.

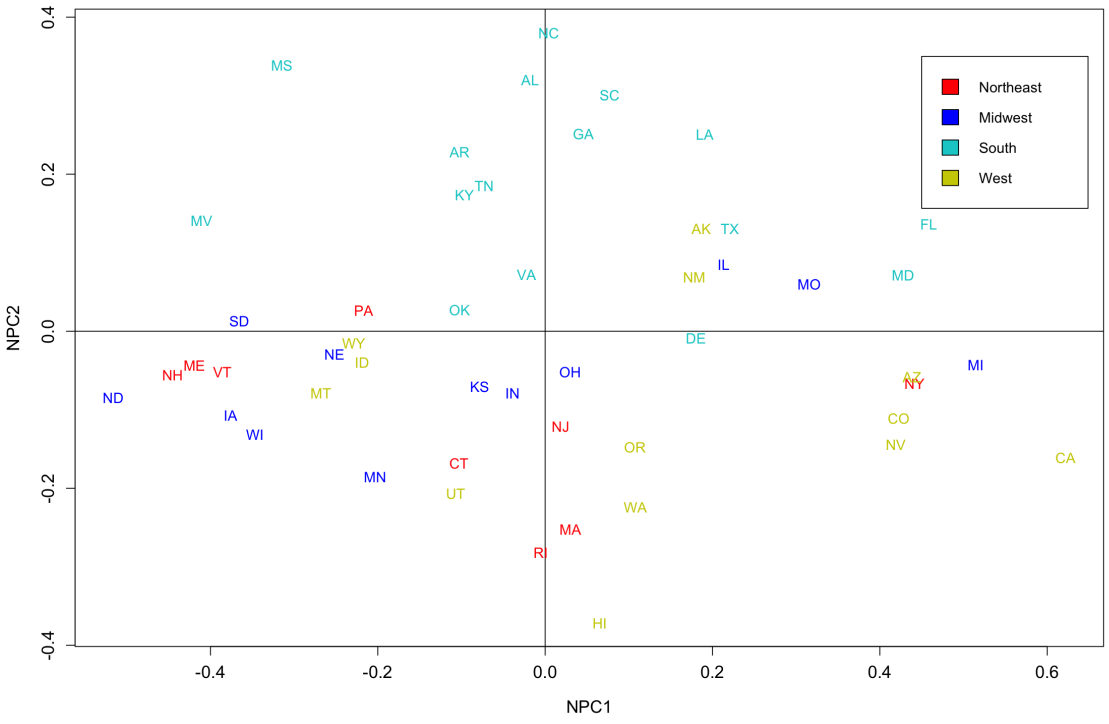

图 9.7: 观测数据 (50 个州) 在前两个 NPC 平面上的散点图.

【解】 观测数据 (50 个州) 在前两个 NPC 平面上的投影如图 9.7 所示. 从横向 (第一主成分) 看，Northeast 地区和 Midwest 地区多数州的值较小 (偏左)，而 South 与 West 地区则大小均有且分布较为均匀，说明总体来看，Northeast 和 Midwest 地区的犯罪问题要更严重一些. 从纵向 (第二主成分) 看，Northeast 地区和 Midwest 地区的值接近于零，说明这些州人身侵犯和谋杀的犯罪率几乎差不多；而 South 州的取值均大于零，说明这些州的人身攻击犯罪远高于盗窃犯罪；West 地区的取值大部分均小于零，说明这些州的盗窃犯罪多于人身攻击的犯罪. 作图的代

码如下：

```
# 各州的观测数据在前 2 个主成分平面上的散点图
X = scale(X)
n = dim(X)[1]
e = eigen(X %*% t(X) / n)
e1 = e$values
e2 = e$vectors
a = e2[, 1:2]
w = a * sqrt(matrix(e1[1:2], nrow(a), ncol(a), byrow = TRUE))
plot(w,  type = 'n', xlab = 'NPC1', ylab = 'NPC2',
     cex.lab = 1.2, cex.axis = 1.2, cex.main = 1.8, lwd = 2)
abline(h = 0, v = 0)
label2 = as.vector(x[,1])
col_reg = x$region
for (i in 1:n){
  if (col_reg[i] == "Northeast") col_reg[i] = "red"
  if (col_reg[i] == "Midwest") col_reg[i] = "blue"
  if (col_reg[i] == "South") col_reg[i] = "cyan3"
  if (col_reg[i] == "West") col_reg[i] = "yellow3"
}
text(w, label2$V1, col = col_reg)
region = c("Northeast", "Midwest", "South", "West")
legend(x = 0.45, y = 0.35, legend = region,
       fill = c("red", "blue", "cyan3", "yellow3"))
```

(i) [2 分] 是否有必要考虑第 3 个主成分?

【解】 我们看到，前两个主成分的累积贡献率为 78.7%，基本能够解释美国各州的犯罪情况，可以不必再考虑第三个主成分. 若认为 78.7% 的累积贡献率还不够高，则可以考虑第三个主成分，并作相应的分析. 考虑第三主成分的计算与分析从略.

2. 乳腺癌数据集 (Breast Cancer Wisconsin Data.csv) 的主成分分析.

数据集 (Breast Cancer Wisconsin Data.csv) 来自 Wisconsin 大学附属医院，由 William H. Wolberg 博士提供. 数据集由 11 个变量的 699 个观测值组成，数据集当中各个变量的定义见表 9.4. 我们用其中的 9 个变量 $(X_2, \ldots, X_{10})$ 的数据来作主成分分析.

(a) [2 分] 读入数据集“Breast Cancer Wisconsin Data.csv”，根据上述定义对每一个变量进行命名 (建议用英文单词或字母缩写). 检查所有变量的类型，最后一个变量 (X_{11}) 是分类变量，将它的属性转变为因子. 其余变量均为数值型，若读入的数据集中有变量非数值型，将它转变为数值型.

【解】 执行上述任务的 R 代码如下，调整后数据集的前 6 行以及数据集的结构如图 9.8 所示.

```
setwd("~/Desktop/2026_Multivariate Statistical Analysis/R_Code/Data/Breast Cancer-Wisconsin Data")
```

表 9.4: 乳腺癌数据集的变量含义.

变量	Definition	中文含义
X_1	Sample code number	样本代码编号
X_2	Clump Thickness	肿块厚度
X_3	Uniformity of Cell Size	细胞大小的一致性
X_4	Uniformity of Cell Shape	细胞形状的一致性
X_5	Marginal Adhesion	边缘黏附 (用于描述细胞边缘与周围组织的黏附程度)
X_6	Single Epithelial Cell Size	单个上皮细胞大小
X_7	Bare Nuclei	裸露的细胞核
X_8	Bland Chromatin	良性染色质
X_9	Normal Nucleoli	正常核仁
X_{10}	Mitoses	有丝分裂
X_{11}	Class	分类 (2 表示良性，4 表示恶性)

```
x = fread("Breast Cancer Wisconsin Data.csv", header = FALSE)
dim(x)
colnames(x) = c("CN", "Thick", "Size", "Shape", "MA", "SECS", "BN", "Bland",
                "Normal", "Mitosis", "Class")
x$BN = as.integer(x$BN)
x$Class = as.factor(x$Class)
x = as.data.frame(x)
head(x)
str(x)
```

(b) [2 分] 数据集当中有 16 个数据含有单一缺失值，这些缺失值在原数据集中用“?”来表示. 找到含有缺失数据的观测值，将它们从数据集当中剔除. 从剔除缺失数据的数据集当中提取变量 $(X_2, \ldots, X_{10})$ 的数据子集，它就是我们要作主成分分析的对象.

【解】 利用 summary() 函数可以发现，缺失的 16 个数据均在变量 X_7 中 (如图9.9所示)，可以使用 na.omit() 函数将它们剔除，然后再提取拟作分析的数据集，完成上述任务的 R 代码如下:

```
n = dim(x)[1] # 样本容量
n
p = dim(x)[2] # 变量个数
p
summary(x) # 变量 BN (X_7) 含有 16 个 NA 数据
x.clean = na.omit(x) # 剔除 NA 后的数据集
```

```
> head(x)
       CN Thick Size Shape MA SECS BN Bland Normal Mitosis Class
1 1000025     5    1     1  1    2  1     3      1       1     2
2 1002945     5    4     4  5    7 10     3      2       1     2
3 1015425     3    1     1  1    2  2     3      1       1     2
4 1016277     6    8     8  1    3  4     3      7       1     2
5 1017023     4    1     1  3    2  1     3      1       1     2
6 1017122     8   10    10  8    7 10     9      7       1     4
> str(x)
'data.frame':   699 obs. of  11 variables:
 $ CN     : int  1000025 1002945 1015425 1016277 1017023 1017122 1018099 1018561 1033078 1033078 ...
 $ Thick  : int  5 5 3 6 4 8 1 2 2 4 ...
 $ Size   : int  1 4 1 8 1 10 1 1 1 2 ...
 $ Shape  : int  1 4 1 8 1 10 1 2 1 1 ...
 $ MA     : int  1 5 1 1 3 8 1 1 1 1 ...
 $ SECS   : int  2 7 2 3 2 7 2 2 2 2 ...
 $ BN     : int  1 10 2 4 1 10 10 1 1 1 ...
 $ Bland  : int  3 3 3 3 3 9 3 3 1 2 ...
 $ Normal : int  1 2 1 7 1 7 1 1 1 1 ...
 $ Mitosis: int  1 1 1 1 1 1 1 1 5 1 ...
 $ Class  : Factor w/ 2 levels "2","4": 1 1 1 1 1 2 1 1 1 1 ...
```

图 9.8: 调整后的数据集.

```
x.new = x.clean[, 2:10] # 提取拟分析的数据子集
head(x.new)
```

```
> summary(x) # 变量 BN (X_7) 含有 16 个 NA 数据
       CN                Thick            Size            Shape            MA
 Min.   :   61634   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
 1st Qu.:  870688   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.: 1.000
 Median : 1171710   Median : 4.000   Median : 1.000   Median : 1.000   Median : 1.000
 Mean   : 1071704   Mean   : 4.418   Mean   : 3.134   Mean   : 3.207   Mean   : 2.807
 3rd Qu.: 1238298   3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.: 5.000   3rd Qu.: 4.000
 Max.   :13454352   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000

      SECS              BN              Bland            Normal          Mitosis       Class
 Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000   2:458
 1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 1.000   4:241
 Median : 2.000   Median : 1.000   Median : 3.000   Median : 1.000   Median : 1.000
 Mean   : 3.216   Mean   : 3.545   Mean   : 3.438   Mean   : 2.867   Mean   : 1.589
 3rd Qu.: 4.000   3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.: 4.000   3rd Qu.: 1.000
 Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000   Max.   :10.000
                  NA's   :16
```

图 9.9: 使用 summary() 函数查看缺失数据.

(c) [2 分] 将得到的数据集进行标准化，计算相关矩阵并给出结果.

【解】 对数据集进行标准化、计算相关矩阵的 R 代码如下，相关矩阵的结果如图9.10.

```
x.standard = scale(x.new, center = TRUE, scale = TRUE) # 标准化
round(head(x.standard), digits = 3) # 标准化数据集的前 6 行
x.cor = cor(x.standard) # 计算相关矩阵
round(x.cor, digits = 3) # 相关矩阵
```

(d) [2 分] 作相关矩阵的谱分解，给出谱分解的结果并作验证运算.

【解】 相关矩阵的谱分解及结果验证的 R 代码如下:

```
> round(x.cor, digits = 3) # 相关矩阵
        Thick  Size Shape    MA  SECS    BN Bland Normal Mitosis
Thick   1.000 0.642 0.653 0.488 0.524 0.593 0.554  0.534   0.351
Size    0.642 1.000 0.907 0.707 0.754 0.692 0.756  0.719   0.461
Shape   0.653 0.907 1.000 0.686 0.722 0.714 0.735  0.718   0.441
MA      0.488 0.707 0.686 1.000 0.595 0.671 0.669  0.603   0.419
SECS    0.524 0.754 0.722 0.595 1.000 0.586 0.618  0.629   0.481
BN      0.593 0.692 0.714 0.671 0.586 1.000 0.681  0.584   0.339
Bland   0.554 0.756 0.735 0.669 0.618 0.681 1.000  0.666   0.346
Normal  0.534 0.719 0.718 0.603 0.629 0.584 0.666  1.000   0.434
Mitosis 0.351 0.461 0.441 0.419 0.481 0.339 0.346  0.434   1.000
```

图 9.10: 相关矩阵.

```
> x.cor.Jordan = eigen(x.cor) # 相关矩阵的谱分解
> lambda = x.cor.Jordan$values # 特征值
> round(lambda, digits = 3)
[1] 5.899 0.776 0.539 0.460 0.380 0.302 0.294 0.261 0.088
> Lambda = diag(lambda) # 特征值构成的对角矩阵 Lambda
> round(Lambda, digits = 3)
      [,1]  [,2]  [,3] [,4] [,5]  [,6]  [,7]  [,8]  [,9]
 [1,] 5.899 0.000 0.000 0.00 0.00 0.000 0.000 0.000 0.000
 [2,] 0.000 0.776 0.000 0.00 0.00 0.000 0.000 0.000 0.000
 [3,] 0.000 0.000 0.539 0.00 0.00 0.000 0.000 0.000 0.000
 [4,] 0.000 0.000 0.000 0.46 0.00 0.000 0.000 0.000 0.000
 [5,] 0.000 0.000 0.000 0.00 0.38 0.000 0.000 0.000 0.000
 [6,] 0.000 0.000 0.000 0.00 0.00 0.302 0.000 0.000 0.000
 [7,] 0.000 0.000 0.000 0.00 0.00 0.000 0.294 0.000 0.000
 [8,] 0.000 0.000 0.000 0.00 0.00 0.000 0.000 0.261 0.000
 [9,] 0.000 0.000 0.000 0.00 0.00 0.000 0.000 0.000 0.088
> Gamma = as.matrix(x.cor.Jordan$vectors) # 特征向量构成的矩阵 Gamma
> round(Gamma, digits = 3)
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]   [,8]   [,9]
 [1,] -0.302 -0.141  0.866 -0.108 -0.080  0.243 -0.009  0.248  0.003
 [2,] -0.381 -0.047 -0.020  0.204  0.146  0.139 -0.205 -0.436  0.733
 [3,] -0.378 -0.082  0.034  0.176  0.108  0.075 -0.127 -0.583 -0.667
 [4,] -0.333 -0.052 -0.413 -0.493  0.020  0.655  0.124  0.163 -0.046
 [5,] -0.336  0.164 -0.088  0.427  0.637 -0.069  0.211  0.459 -0.067
 [6,] -0.335 -0.261  0.001 -0.499  0.125 -0.609  0.403 -0.127  0.077
 [7,] -0.346 -0.228 -0.213 -0.013 -0.228 -0.299 -0.700  0.384 -0.062
 [8,] -0.336  0.034 -0.134  0.417 -0.690 -0.022  0.460  0.074  0.022
 [9,] -0.230  0.906  0.080 -0.259 -0.105 -0.148 -0.132 -0.054 -0.007
> # 谱分解的结果验证: Gamma * Lambda * t(Gamma)
> round(Gamma %*% Lambda %*% t(Gamma), digits = 3)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9]
 [1,] 1.000 0.642 0.653 0.488 0.524 0.593 0.554 0.534 0.351
```

```
 [2,] 0.642 1.000 0.907 0.707 0.754 0.692 0.756 0.719 0.461
 [3,] 0.653 0.907 1.000 0.686 0.722 0.714 0.735 0.718 0.441
 [4,] 0.488 0.707 0.686 1.000 0.595 0.671 0.669 0.603 0.419
 [5,] 0.524 0.754 0.722 0.595 1.000 0.586 0.618 0.629 0.481
 [6,] 0.593 0.692 0.714 0.671 0.586 1.000 0.681 0.584 0.339
 [7,] 0.554 0.756 0.735 0.669 0.618 0.681 1.000 0.666 0.346
 [8,] 0.534 0.719 0.718 0.603 0.629 0.584 0.666 1.000 0.434
 [9,] 0.351 0.461 0.441 0.419 0.481 0.339 0.346 0.434 1.000
> round(x.cor, digits = 3)
        Thick  Size Shape    MA  SECS    BN Bland Normal Mitosis
Thick   1.000 0.642 0.653 0.488 0.524 0.593 0.554  0.534   0.351
Size    0.642 1.000 0.907 0.707 0.754 0.692 0.756  0.719   0.461
Shape   0.653 0.907 1.000 0.686 0.722 0.714 0.735  0.718   0.441
MA      0.488 0.707 0.686 1.000 0.595 0.671 0.669  0.603   0.419
SECS    0.524 0.754 0.722 0.595 1.000 0.586 0.618  0.629   0.481
BN      0.593 0.692 0.714 0.671 0.586 1.000 0.681  0.584   0.339
Bland   0.554 0.756 0.735 0.669 0.618 0.681 1.000  0.666   0.346
Normal  0.534 0.719 0.718 0.603 0.629 0.584 0.666  1.000   0.434
Mitosis 0.351 0.461 0.441 0.419 0.481 0.339 0.346  0.434   1.000
```

谱分解的结果如下：相关矩阵的特征值为

$$\lambda_1 = 5.899,\ \lambda_2 = 0.776,\ \lambda_3 = 0.539,\ \lambda_4 = 0.460,\ \lambda_5 = 0.380,$$
$$\lambda_6 = 0.302,\ \lambda_7 = 0.294,\ \lambda_8 = 0.261,\ \lambda_9 = 0.088$$

对应的特征向量构成的矩阵为

$$\Gamma = \begin{pmatrix} -0.302 & -0.141 & 0.866 & -0.108 & -0.080 & 0.243 & -0.009 & 0.248 & 0.003 \\ -0.381 & -0.047 & -0.020 & 0.204 & 0.146 & 0.139 & -0.205 & -0.436 & 0.733 \\ -0.378 & -0.082 & 0.034 & 0.176 & 0.108 & 0.075 & -0.127 & -0.583 & -0.667 \\ -0.333 & -0.052 & -0.413 & -0.493 & 0.020 & 0.655 & 0.124 & 0.163 & -0.046 \\ -0.336 & 0.164 & -0.088 & 0.427 & 0.637 & -0.069 & 0.211 & 0.459 & -0.067 \\ -0.335 & -0.261 & 0.001 & -0.499 & 0.125 & -0.609 & 0.403 & -0.127 & 0.077 \\ -0.346 & -0.228 & -0.213 & -0.013 & -0.228 & -0.299 & -0.700 & 0.384 & -0.062 \\ -0.336 & 0.034 & -0.134 & 0.417 & -0.690 & -0.022 & 0.460 & 0.074 & 0.022 \\ -0.230 & 0.906 & 0.080 & -0.259 & -0.105 & -0.148 & -0.132 & -0.054 & -0.007 \end{pmatrix}$$

谱分解验证的结果如图9.11所示.

(e) [2 分] 画碎石图，问各个主成分的贡献率是多少？前 2 个主成分的累积贡献率是多少，前 3 个主成分的累积贡献率又是多少.

【解】 碎石图如图9.12所示，各个主成分的贡献率以及累积贡献率如图9.13所示. 我们看到，前 2

```
> # 谱分解的结果验证: Gamma * Lambda * t(Gamma)
> round(Gamma %*% Lambda %*% t(Gamma), digits = 3)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9]
 [1,] 1.000 0.642 0.653 0.488 0.524 0.593 0.554 0.534 0.351
 [2,] 0.642 1.000 0.907 0.707 0.754 0.692 0.756 0.719 0.461
 [3,] 0.653 0.907 1.000 0.686 0.722 0.714 0.735 0.718 0.441
 [4,] 0.488 0.707 0.686 1.000 0.595 0.671 0.669 0.603 0.419
 [5,] 0.524 0.754 0.722 0.595 1.000 0.586 0.618 0.629 0.481
 [6,] 0.593 0.692 0.714 0.671 0.586 1.000 0.681 0.584 0.339
 [7,] 0.554 0.756 0.735 0.669 0.618 0.681 1.000 0.666 0.346
 [8,] 0.534 0.719 0.718 0.603 0.629 0.584 0.666 1.000 0.434
 [9,] 0.351 0.461 0.441 0.419 0.481 0.339 0.346 0.434 1.000
> round(x.cor, digits = 3)
        Thick  Size Shape    MA  SECS    BN Bland Normal Mitosis
Thick   1.000 0.642 0.653 0.488 0.524 0.593 0.554  0.534   0.351
Size    0.642 1.000 0.907 0.707 0.754 0.692 0.756  0.719   0.461
Shape   0.653 0.907 1.000 0.686 0.722 0.714 0.735  0.718   0.441
MA      0.488 0.707 0.686 1.000 0.595 0.671 0.669  0.603   0.419
SECS    0.524 0.754 0.722 0.595 1.000 0.586 0.618  0.629   0.481
BN      0.593 0.692 0.714 0.671 0.586 1.000 0.681  0.584   0.339
Bland   0.554 0.756 0.735 0.669 0.618 0.681 1.000  0.666   0.346
Normal  0.534 0.719 0.718 0.603 0.629 0.584 0.666  1.000   0.434
Mitosis 0.351 0.461 0.441 0.419 0.481 0.339 0.346  0.434   1.000
```

图 9.11: 相关矩阵谱分解结果的验证.

个主成分的累积贡献率是 74.2%，前 3 个主成分的累积贡献率是 80.2%. 所用 R 代码如下:

```
> plot(lambda, axes = FALSE, pch = 16, type = "b", xlab = '', ylab = '',
+      xlim = c(0, 10), ylim = c(0, 7), main = 'Scree Plot')
> arrows(0, 0, 0, 7, length = 0.1)
> arrows(0, 0, 10, 0, length = 0.1)
> axis(1, at = 1:9, pos = 0)
> axis(2, at = 1:6, pos = 0)

> round(lambda / sum(lambda), digits = 3) # 贡献率
[1] 0.655 0.086 0.060 0.051 0.042 0.034 0.033 0.029 0.010
> round(cumsum(lambda) / sum(lambda), digits = 3) # 累积贡献率
[1] 0.655 0.742 0.802 0.853 0.895 0.928 0.961 0.990 1.000
```

(f) [2 分] 计算初始变量与前 2 个主成分的相关系数并给出结果. 前 2 个主成分对每个变量解释的比例是多少?

【解】 初始变量与前 2 个主成分的相关系数、前 2 个主成分对每个变量解释的比例如图9.14所示. 相应的 R 代码如下:

```
a = Gamma[, 1:2] # 前两大特征值对应的特征向量
round(a, digits = 3)
# 前两大特征值开方的矩阵 b
b = sqrt(matrix(lambda[1:2], nrow(a), ncol(a), byrow = TRUE))
round(b, digits = 3)
```

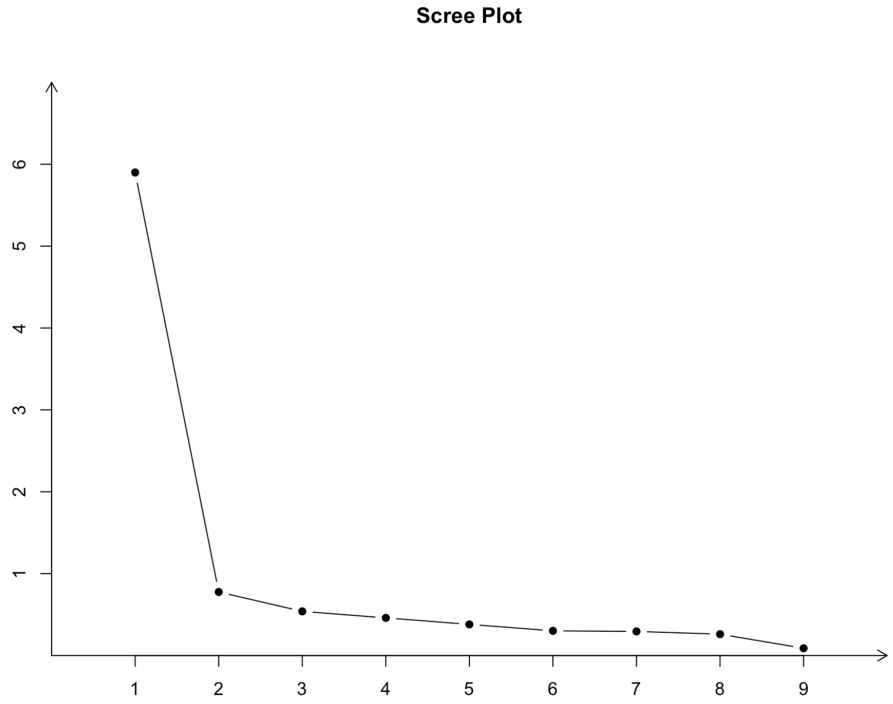

图 9.12: 碎石图.

```
> round(lambda / sum(lambda), digits = 3) # 贡献率
[1] 0.655 0.086 0.060 0.051 0.042 0.034 0.033 0.029 0.010
> round(cumsum(lambda) / sum(lambda), digits = 3) # 累积贡献率
[1] 0.655 0.742 0.802 0.853 0.895 0.928 0.961 0.990 1.000
```

图 9.13: 各主成分的贡献率与累积贡献率.

```
w = a * b # 变量与前两个主成分的相关系数
round(w, digits = 3)
c = as.matrix(apply(w^2, 1, sum)) # 各变量与前两个NPC相关系数的平方和
round(c, digits = 3)
# 将上述计算结果放在一个数据框中
NPC = data.frame(NPC1 = w[, 1], NPC2 = w[, 2], SS = c[, 1])
rownames(NPC) = colnames(x.new)
round(NPC, digits = 3)
```

(g) [2 分] 作变量在前 2 个主成分平面上的散点图，对结果进行解释.

【解】 变量在前 2 个主成分平面上的散点图如图9.15所示. 从中可以看到，所有变量点的位置都接近于单位圆的圆周，说明前两个主成分对每一个变量的解释程度都比较高. 变量 $X_2 \sim X_8$ 在前 2 个主成分平面上的散点图相对比较集中、夹角很小，说明这些变量之间具有较高的正相关. 变量 X_{10} (有丝分裂) 与这八个变量的夹角几乎垂直，说明这两组变量之间的相关性很小. 第一主成分除与变量 X_{10} (有丝分裂) 的相关系数较低，与其余变量相关系数的绝对值都在 0.7 以上，说明第一主成分综合反映了除变量 X_{10} (有丝分裂) 之外的其余变量的信息. 第二主成分与变量 X_{10} (有丝分裂) 的相关系数接近 0.8，与其余变量的相关系数都很小，说明第二主成分主要反映

```
> round(NPC, digits = 3)
           NPC1   NPC2    SS
Thick    -0.734 -0.124 0.554
Size     -0.925 -0.041 0.857
Shape    -0.917 -0.073 0.846
MA       -0.808 -0.046 0.655
SECS     -0.817  0.145 0.688
BN       -0.814 -0.230 0.715
Bland    -0.840 -0.201 0.746
Normal   -0.815  0.030 0.665
Mitosis -0.559  0.798 0.949
```

图 9.14: 初始变量与前 2 个主成分的相关系数.

了变量 X_{10} (有丝分裂) 的信息.

相应的 R 代码如下:

```
ucircle = cbind(cos((0:360)/180 * pi), sin((0:360)/180 * pi))
plot(ucircle, type = "l", lty = "solid", col = "blue",
     xlim = c(-1.05, 1.05), ylim = c(-1.05, 1.05),
     xlab = "NPC1", ylab = "NPC2",
     cex.lab = 1.2, cex.axis = 1.2, cex.main = 1.8,
     lwd = 2, asp = 1)
abline(h = 0, v = 0)
label = rownames(NPC)
points(NPC[, 1:2], pch = 16, col = 'red', cex = 1)
text(NPC[, 1:2], label)
```

(h) [2 分] 变量 Class (X_{11}) 是分类变量，分别对应于良性与恶性. 作观测数据在前 2 个主成分平面上的散点图，能否看出良性与恶性的表现存在不同?

【解】 观测数据在前 2 个主成分平面上的散点图如图9.16所示. 从该散点图可以看到，良性肿瘤与恶性肿瘤表现出了明显的不同. 虽然有一些点的位置相互重叠，但它们所占比例很小.

相应的 R 代码如下:

```
y = x.standard %*% Gamma # 观测数据在各主成分上的投影值
head(y)
graphics.off()
plot(y[, 1], y[, 2], pch = 16, col = x.clean$Class, xlab = "PC1",
     ylab = "PC2", main = "First vs. Second PC",
     cex.axis = 1.2, cex.lab = 1.2, cex.main = 1.6, asp = 1)
abline(h = 0, v = 0, lty = 2, col = 'cyan2')
```

(i) [2 分] 是否有必要考虑第 3 个主成分?

【解】 从上述分析可见，前两个主成分的散点图关于良性肿瘤与恶性肿瘤的划分已有较为明显的效果，可以不必考虑第 3 个主成分. 引入第 3 个主成分后对于这种区分的效果如果有显著的提

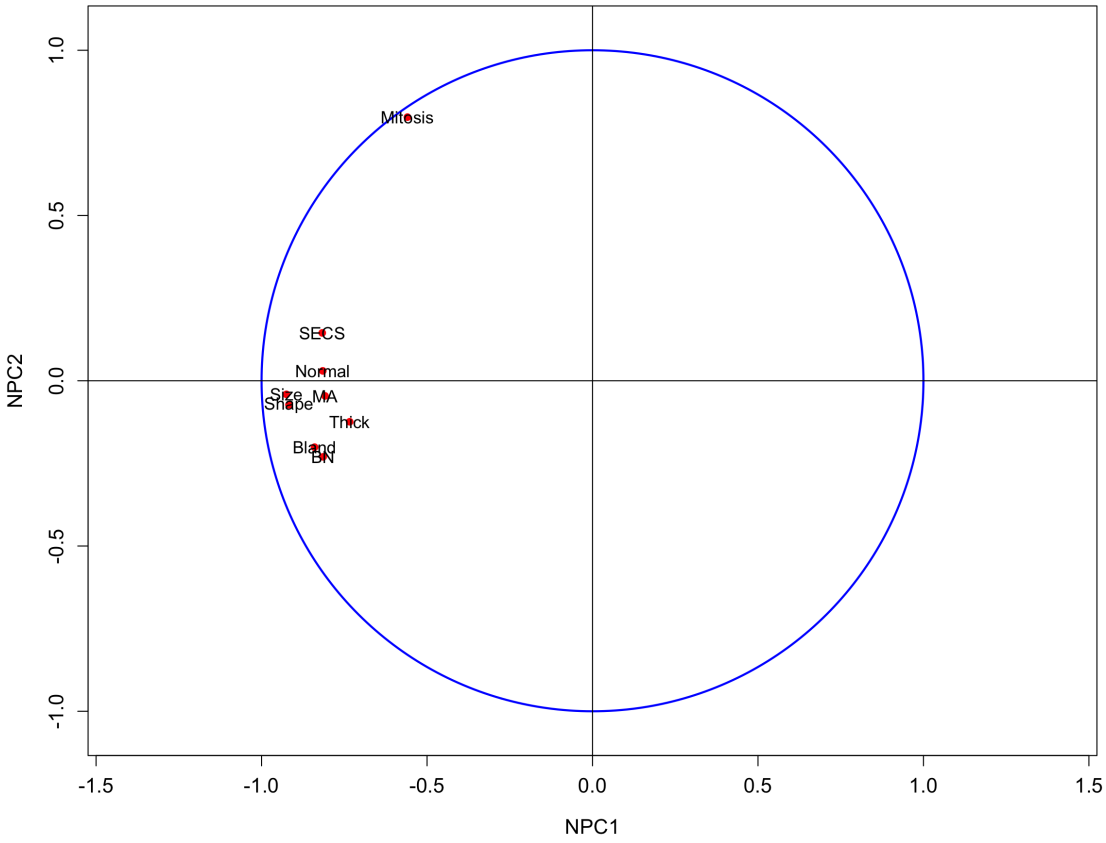

图 9.15: 变量在前 2 个主成分平面上的散点图.

升，则可以考虑引入它，具体的分析这里就不再赘述了，感兴趣的同学可以自己完成分析.

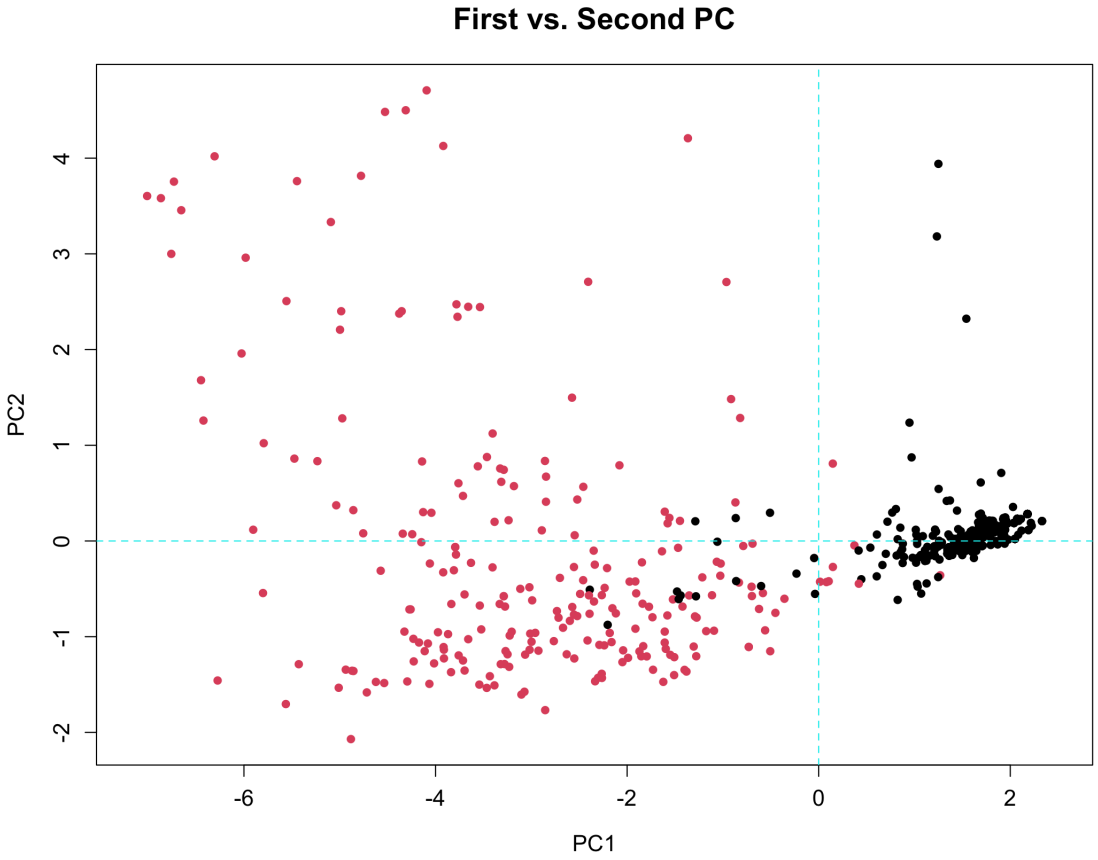

图 9.16: 观测数据在前 2 个主成分平面上的散点图 (红色代表良性，黑色代表恶性).

Chapter 10

第 10 周作业参考答案

1. 数据文件 2018−Mean Expenditure of Urban Residents.csv 包含了 2018 年中国 31 个省 (市) 的城镇居民人均生活费用，各变量的含义如下：Province 省市名称，Food 食品，Cloth 衣着，Residential 居住，Expenditure 生活用品与服务，Trans-Com (交通通讯), Education (教育娱乐), Healthcare (医疗保健), and Others (其它). 作上述变量的因子分析.

   (a) [2 分] 读入数据，从相关矩阵出发作因子分析.

   【解】 利用以下 R 代码计算相关矩阵、作因子分析. 因子数取 $k = 2$，原因见 (b).

```
rm(list = ls(all = TRUE)) # 清除当前所有变量与对象
graphics.off() # 清除当前所有图形
library(data.table)
setwd("~/Desktop/2025_Multivariate Statistical Analysis/R_Code/Data")
Expenditure = fread("2018_Mean Expenditure of Urban Residents.csv",
    header = TRUE)
x = Expenditure[, 2:9]
head(x) # 拟分析数据的前 6 行
x.cor = cor(x) # 计算相关矩阵
round(x.cor, digits = 3)
# 以下作正交旋转的因子分析
x.FA = factanal(x, factors = 2, scores = "regression", rotation = "varimax")
str(x.FA)
```

相关矩阵为：

$$\mathcal{R} = \begin{pmatrix} 1.000 & 0.580 & 0.834 & 0.460 & 0.850 & 0.820 & 0.582 & 0.865 \\ 0.580 & 1.000 & 0.667 & 0.463 & 0.801 & 0.678 & 0.798 & 0.808 \\ 0.834 & 0.667 & 1.000 & 0.281 & 0.869 & 0.900 & 0.738 & 0.912 \\ 0.460 & 0.463 & 0.281 & 1.000 & 0.493 & 0.376 & 0.441 & 0.470 \\ 0.850 & 0.801 & 0.869 & 0.493 & 1.000 & 0.861 & 0.776 & 0.914 \\ 0.820 & 0.678 & 0.900 & 0.376 & 0.861 & 1.000 & 0.843 & 0.922 \\ 0.582 & 0.798 & 0.738 & 0.441 & 0.776 & 0.843 & 1.000 & 0.857 \\ 0.865 & 0.808 & 0.912 & 0.470 & 0.914 & 0.922 & 0.857 & 1.000 \end{pmatrix}$$

(b) [2 分] 确定公共因子数量并给出理由.

【解】 确定公共因子数量的 R 代码如下. 可以看到，前两个特征值的累计贡献率达到了 86.4%，所以，我们考虑 $k = 2$ 个公共因子的因子分析模型.

```
x.ee = eigen(x.cor) # 相关矩阵的谱分解
round(x.ee$values, digits = 3) # 相关矩阵的特征值
round(x.ee$values / sum(x.ee$values), digits = 3) # 各特征值的贡献率
round(cumsum(x.ee$values) / sum(x.ee$values), digits = 3) # 特征值的累计贡献率
```

特征值	6.087	0.825	0.550	0.271	0.099	0.086	0.056	0.026
贡献率	0.761	0.103	0.069	0.034	0.012	0.011	0.007	0.003
累计贡献率	0.761	0.864	0.933	0.967	0.979	0.990	0.997	1.000

(c) [2 分] 给出因子旋转之后的因子载荷矩阵.

【解】 用到的 R 代码如下，正交旋转后的因子载荷矩阵为

$$\widehat{\mathcal{Q}} = \begin{pmatrix} 0.902 & 0.302 \\ 0.451 & 0.691 \\ 0.777 & 0.512 \\ 0.315 & 0.357 \\ 0.746 & 0.562 \\ 0.678 & 0.656 \\ 0.330 & 0.941 \\ 0.740 & 0.651 \end{pmatrix}$$

```
x.ld = x.FA$loadings
x.ld
```

(d) [2 分] 作变量在公共因子平面上的散点图，对公共因子作出解释.

【解】 变量在公共因子平面上的散点图如图10.1所示，作图的 R 代码如下. 可以看到，对公共因子 1 影响最大的变量是 Food (食品)，还有变量 Residential (居住)、Trans-Com (交通通讯) 的载荷亦较高，因此公共因子 1 可以称作“生活质量”因子. 对公共因子 2 影响最大的变量是 Healthcare (医疗保健)，还有变量 Cloth (衣着)、Education (教育娱乐)、Others (其它) 的载荷亦较高，因此公共因子 2 可以称作“健康与教育”因子.

```
ucircle = cbind(cos((0:360)/180 * pi), sin((0:360)/180 * pi))
plot(ucircle, type = "l", lty = "solid", col = "blue",
     xlim = c(-1.05, 1.05), ylim = c(-1.05, 1.05),
     xlab = "Factor 1", ylab = "Factor 2",
     cex.lab = 1.2, cex.axis = 1.2, cex.main = 1.8,
     lwd = 2, asp = 1)
abline(h = 0, v = 0)
label = rownames(x.cor)
points(x.ld[, 1:2], pch = 16, col = 'red', cex = 1.5)
text(x.ld[, 1:2], label)
```

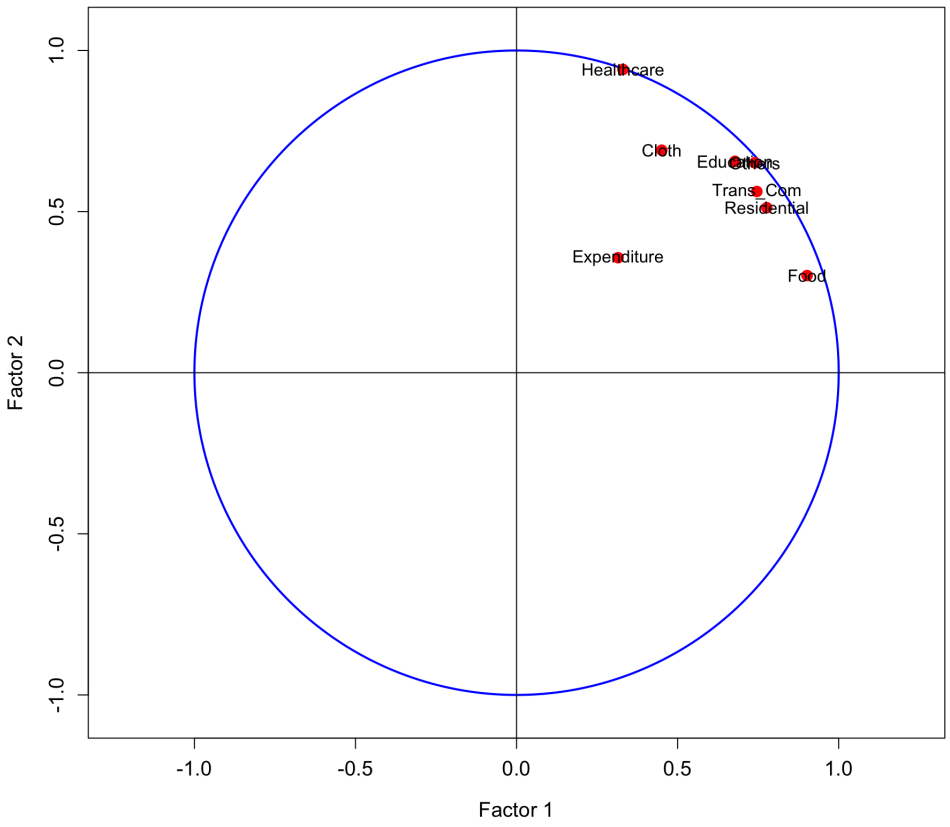

图 10.1: 变量在公共因子平面上的散点图.

(e) [2 分] 利用因子得分，作观测数据在公共因子平面上的散点图.

【解】 观测数据在公共因子平面上的散点图如图10.2所示，作图的 R 代码如下.

```
y = x.FA$scores
y
plot(y, type = "n", xlim = c(-2, 4), ylim = c(-3, 3), asp = 1)
abline(h = 0, v = 0, lty = 2, col = "cyan")
label2 = Expenditure$Province
points(y, pch = 16, col = 'orange', cex = 1.5)
text(y, label2)
```

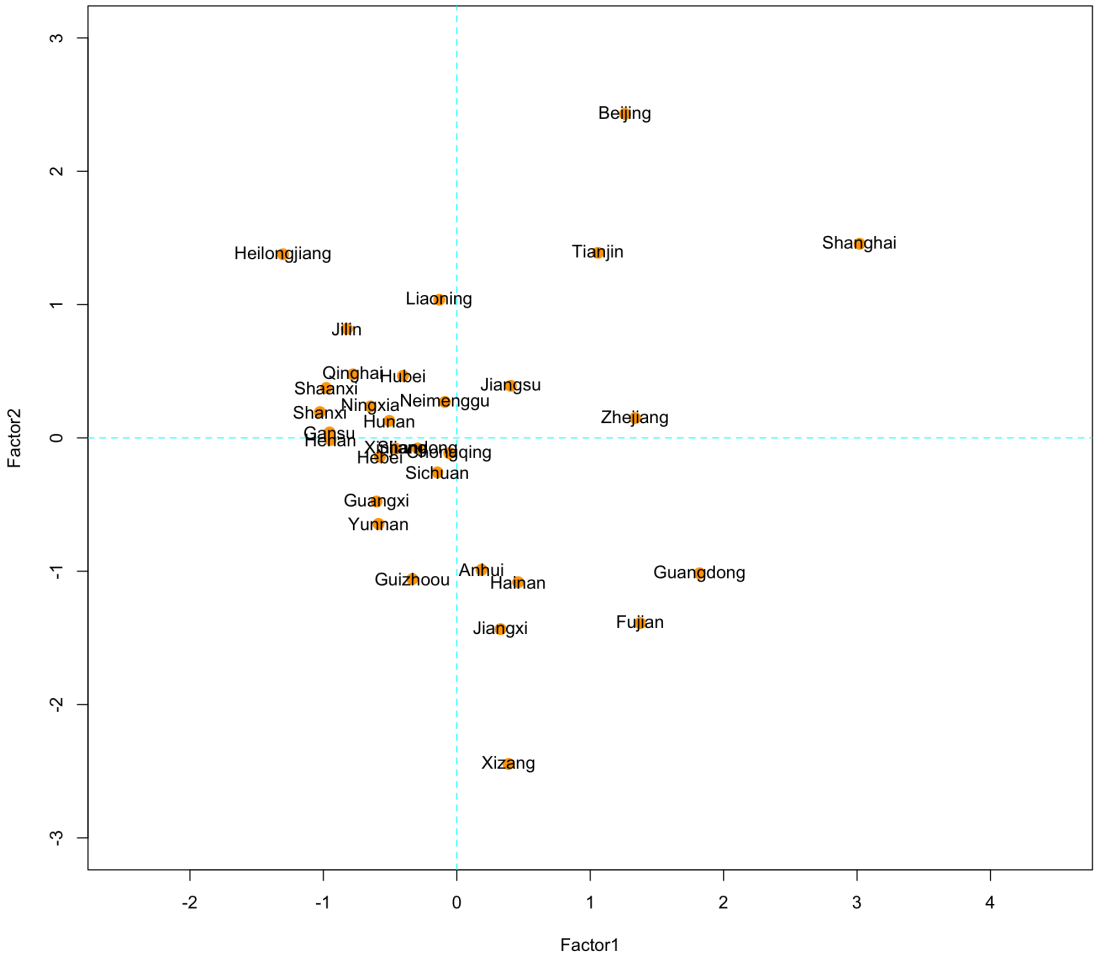

图 10.2: 观测数据在公共因子平面上的散点图.

(f) [2 分] 利用第 1 个公共因子的得分对我国 31 个省 (市) 进行排序并作简要分析.

【解】 我国 31 个省 (市) 第 1 个公共因子的得分如图10.3所示，公共因子 1 是“生活质量”因子，从城镇居民人均生活费用数据来分析，生活质量因子得分较高的省市有：上海、广东、福建、浙江、北京、天津；生活质量因子得分较低的省市有：黑龙江、山西、陕西、甘肃、河南、吉林、青海、宁夏、广西. 其它省市的取值居中. 提取数据的 R 代码如下：

```
fact1_score = y[, 1]
names(fact1_score) = Expenditure$Province
```

```
fact1_sort = round(sort(fact1_score, decreasing = TRUE), digits = 2)
fact1_sort = as.data.frame(fact1_sort)
fact1_sort
```

```
> fact1_sort
             fact1_sort
Shanghai           3.02
Guangdong          1.82
Fujian             1.37
Zhejiang           1.34
Beijing            1.26
Tianjin            1.06
Hainan             0.46
Jiangsu            0.40
Xizang             0.39
Jiangxi            0.33
Anhui              0.19
Chongqing         -0.06
Neimenggu         -0.09
Liaoning          -0.13
Sichuan           -0.15
Shandong          -0.29
Guizhoou          -0.33
Hubei             -0.40
Xinjiang          -0.46
Hunan             -0.51
Hebei             -0.58
Yunnan            -0.59
Guangxi           -0.60
Ningxia           -0.65
Qinghai           -0.78
Jilin             -0.82
Henan             -0.95
Gansu             -0.95
Shaanxi           -0.98
Shanxi            -1.03
Heilongjiang      -1.30
```

图 10.3: 31 个省 (市) 第 1 个公共因子的得分.

(g) [2 分] 利用第 2 个公共因子的得分对我国 31 个省 (市) 进行排序并作简要分析.

【解】 我国 31 个省 (市) 第 2 个公共因子的得分如图10.4所示，公共因子 2 是“健康与教育”因子，从城镇居民人均生活费用数据来分析，健康与教育因子得分较高的省市有：北京、上海、天津、黑龙江、辽宁；健康与教育因子得分较低的省市有：西藏、江西、福建、海南、贵州、广东. 其它省市的取值居中. 提取数据的 R 代码如下：

```
fact2_score = y[, 2]
names(fact2_score) = Expenditure$Province
fact2_sort = round(sort(fact2_score, decreasing = TRUE), digits = 2)
fact2_sort = as.data.frame(fact2_sort)
```

```
fact2_sort
```

```
> fact2_sort
             fact2_sort
Beijing            2.43
Shanghai           1.46
Tianjin            1.39
Heilongjiang       1.38
Liaoning           1.04
Jilin              0.82
Qinghai            0.48
Hubei              0.46
Jiangsu            0.39
Shaanxi            0.37
Neimenggu          0.27
Ningxia            0.24
Shanxi             0.19
Zhejiang           0.15
Hunan              0.13
Gansu              0.04
Henan             -0.02
Shandong          -0.08
Xinjiang          -0.08
Chongqing         -0.11
Hebei             -0.14
Sichuan           -0.26
Guangxi           -0.48
Yunnan            -0.65
Anhui             -0.99
Guangdong         -1.02
Guizhoou          -1.06
Hainan            -1.08
Fujian            -1.39
Jiangxi           -1.43
Xizang            -2.45
```

图 10.4: 31 个省 (市) 第 2 个公共因子的得分.

(h) [2 分] 利用第 1、第 2 两个公共因子的得分对我国 31 个省 (市) 城镇居民人均生活费用进行综合排序，并作简要分析.

【解】 因为第 1 个公共因子的贡献率为 76.09%，第 2 个公共因子的贡献率为 10.31%，利用贡献率作为因子得分的权重系数对第 1、2 公共因子的得分值进行加权求和，作为综合排序的依据，排序的结果如图10.5所示. 从城镇居民人均生活费用数据来分析，综合健康、教育、生活质量等因素，可以看到综合得分较高的省市有：上海、广东、北京、浙江、天津、福建；综合得分较低的省市有：黑龙江、山西、河南、甘肃、陕西. 其它省市的取值相对居中. 用到的 R 代码如下：

```
a = x.ee$values / sum(x.ee$values)
round(a, digits = 4)[1:2] # 第1、2公共因子的贡献率
z = a[1] * y[, 1] + a[2] * y[, 2]
```

```
names(z) = Expenditure$Province
fact12_sort = round(sort(z, decreasing = TRUE), digits = 2)
fact12_sort = as.data.frame(fact12_sort)
fact12_sort
```

```
> fact12_sort
             fact12_sort
Shanghai            2.45
Guangdong           1.28
Beijing             1.21
Zhejiang            1.03
Tianjin             0.95
Fujian              0.90
Jiangsu             0.35
Hainan              0.24
Jiangxi             0.10
Xizang              0.04
Anhui               0.04
Liaoning            0.01
Neimenggu          -0.04
Chongqing          -0.05
Sichuan            -0.14
Shandong           -0.23
Hubei              -0.26
Xinjiang           -0.36
Guizhoou           -0.36
Hunan              -0.37
Hebei              -0.45
Ningxia            -0.47
Guangxi            -0.51
Yunnan             -0.51
Jilin              -0.54
Qinghai            -0.54
Shaanxi            -0.71
Gansu              -0.72
Henan              -0.72
Shanxi             -0.76
Heilongjiang       -0.85
```

图 10.5: 31 个省 (市) 第 1、2 公共因子得分的综合排序.

2. 有 48 人申请到某公司就业. 该公司对申请者的 15 项指标进行打分，这 15 项指标分别是：FL (求职信的形式)，APP (外貌)，AA (专业能力)，LA (讨人喜欢程度)，SC (自信心)，LC (洞察力)，HON (诚实度)，SMS (推销能力)，EXP (经验)，DRV (驾驶水平)，AMB (事业心)，GSP (理解能力)，POT (潜在能力)，KJ (社交能力)，SUIT (适应能力). 结果见数据文件 Applicants.csv .

   (a) [2 分] 读入数据，从相关矩阵出发作因子分析.

   【解】 读入数据、计算相关矩阵、作因子分析的 R 代码如下，本题中我们使用 psych 包中的 fa()

函数来拟合因子分析模型. 相关矩阵的结果如图10.6所示. 取因子个数 $k = 2$, 原因见 (b). 拟合带有正交旋转的因子分析模型.

```
rm(list = ls(all = TRUE)) # 清除当前所有变量与对象
graphics.off() # 清除当前所有图形
library(data.table)
setwd("~/Desktop/2025_Multivariate Statistical Analysis/R_Code/Data")
Applicants = fread("Applicants.csv", header = TRUE)
head(Applicants)
x = as.matrix(Applicants[, 2:16])
head(x) # 拟分析数据的前 6 行
x.cor = cor(x) # 计算相关矩阵
round(x.cor, digits = 3)
# 利用 psych 包中的 fa 函数拟合正交旋转的因子分析模型
library(psych)
x.fa = fa(x, nfactors = 2, rotate = "varimax", fm = "pa", scores = "regression")
str(x.fa)
```

```
> round(x.cor, digits = 3)
        FL   APP    AA    LA    SC    LC    HON   SMS    EXP   DRV   AMB   GSP   POT    KJ  SUIT
FL   1.000 0.199 0.204 0.258 0.050 0.124 -0.067 0.203  0.505 0.299 0.221 0.237 0.284 0.371 0.514
APP  0.199 1.000 0.208 0.412 0.436 0.463  0.262 0.529  0.211 0.378 0.579 0.588 0.570 0.393 0.462
AA   0.204 0.208 1.000 0.098 0.188 0.146  0.048 0.237  0.107 0.211 0.164 0.139 0.186 0.048 0.219
LA   0.258 0.412 0.098 1.000 0.302 0.483  0.645 0.362  0.141 0.393 0.347 0.503 0.606 0.685 0.327
SC   0.050 0.436 0.188 0.302 1.000 0.808  0.410 0.800  0.015 0.704 0.842 0.721 0.672 0.482 0.250
LC   0.124 0.463 0.146 0.483 0.808 1.000  0.356 0.818  0.147 0.698 0.758 0.883 0.777 0.527 0.416
HON -0.067 0.262 0.048 0.645 0.410 0.356  1.000 0.240 -0.156 0.280 0.215 0.386 0.416 0.448 0.003
SMS  0.203 0.529 0.237 0.362 0.800 0.818  0.240 1.000  0.255 0.815 0.860 0.782 0.754 0.563 0.558
EXP  0.505 0.211 0.107 0.141 0.015 0.147 -0.156 0.255  1.000 0.337 0.195 0.299 0.348 0.215 0.693
DRV  0.299 0.378 0.211 0.393 0.704 0.698  0.280 0.815  0.337 1.000 0.780 0.714 0.788 0.613 0.623
AMB  0.221 0.579 0.164 0.347 0.842 0.758  0.215 0.860  0.195 0.780 1.000 0.784 0.769 0.547 0.435
GSP  0.237 0.588 0.139 0.503 0.721 0.883  0.386 0.782  0.299 0.714 0.784 1.000 0.876 0.549 0.528
POT  0.284 0.570 0.186 0.606 0.672 0.777  0.416 0.754  0.348 0.788 0.769 0.876 1.000 0.539 0.574
KJ   0.371 0.393 0.048 0.685 0.482 0.527  0.448 0.563  0.215 0.613 0.547 0.549 0.539 1.000 0.396
SUIT 0.514 0.462 0.219 0.327 0.250 0.416  0.003 0.558  0.693 0.623 0.435 0.528 0.574 0.396 1.000
```

图 10.6: 求职数据的相关矩阵.

(b) [2 分] 确定公共因子数量并给出理由.

【解】 使用 psych 包中的 fa.parallel() 函数作碎石图, 结果如图10.7所示, 相应的 R 代码如下, 从中可以看出, 取因子个数 $k = 2$ 是合理的.

```
fa.parallel(x.cor, n.obs = 48, fa = "both", n.iter = 100,
        main = "Scree plots with parallel analysis")
```

(c) [2 分] 给出因子旋转之后的因子载荷矩阵.

【解】 用到的 R 代码如下:

```
x.ld = x.fa$loadings
x.ld
```

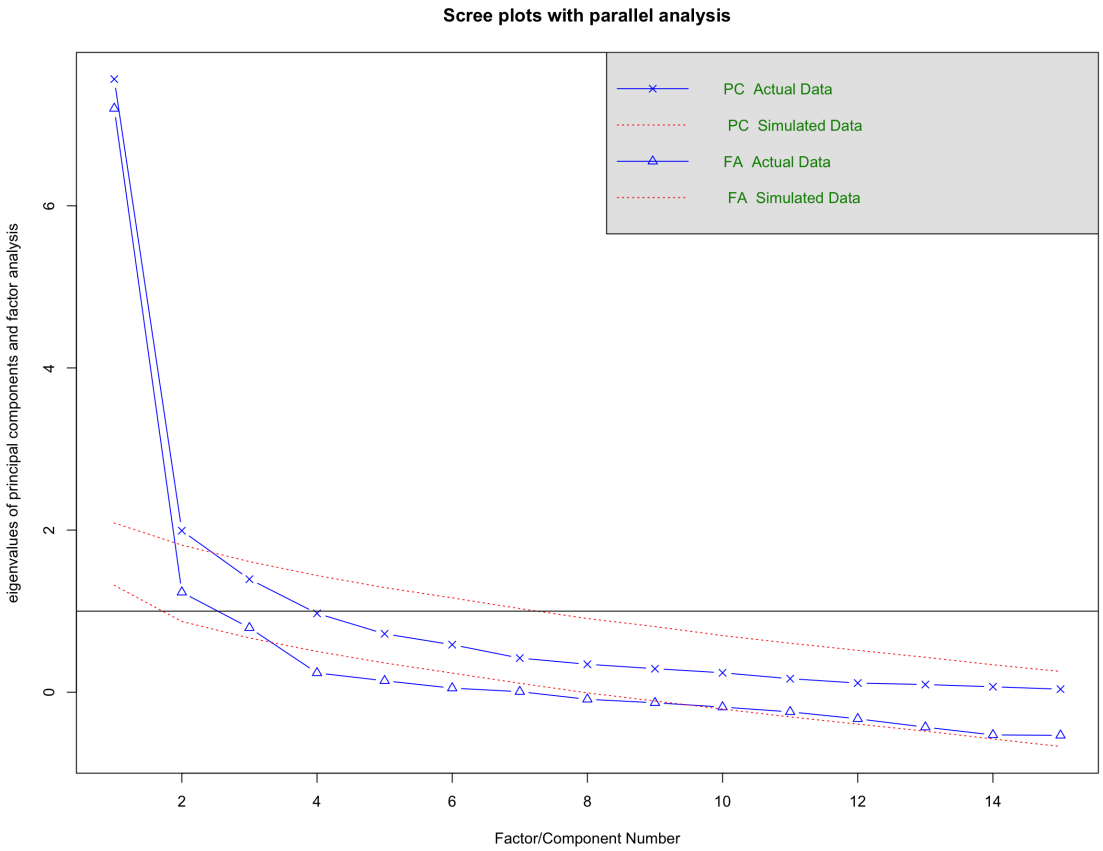

图 10.7: 求职数据的碎石图.

正交旋转后的因子载荷矩阵为 (载荷小于 0.1 的未显示数值):

$$\widehat{\mathcal{Q}} = \begin{pmatrix} & 0.616 \\ 0.537 & 0.280 \\ 0.155 & 0.189 \\ 0.547 & 0.173 \\ 0.878 & \\ 0.880 & 0.145 \\ 0.514 & -0.171 \\ 0.831 & 0.312 \\ & 0.807 \\ 0.747 & 0.411 \\ 0.838 & 0.243 \\ 0.856 & 0.304 \\ 0.826 & 0.376 \\ 0.617 & 0.270 \\ 0.320 & 0.850 \end{pmatrix}$$

(d) [2 分] 作变量在公共因子平面上的散点图，对公共因子作出解释.

【解】 变量在公共因子平面上的散点图如图10.8所示，作图的 R 代码如下.

```
ucircle = cbind(cos((0:360)/180 * pi), sin((0:360)/180 * pi))
plot(ucircle, type = "l", lty = "solid", col = "blue",
     xlim = c(-1.05, 1.05), ylim = c(-1.05, 1.05),
     xlab = "Factor 1", ylab = "Factor 2",
     cex.lab = 1.2, cex.axis = 1.2, cex.main = 1.8,
     lwd = 2, asp = 1)
abline(h = 0, v = 0)
label = rownames(x.cor)
points(x.ld[, 1:2], pch = 16, col = 'red', cex = 1.5)
text(x.ld[, 1:2], label)
```

可以看到，对公共因子 1 影响大的变量主要有：SC (自信心)，LC (洞察力)，SMS (推销能力)，AMB (事业心)，GSP (理解能力)，POT (潜在能力)，因此公共因子 1 可以称作“能力与潜力”因子. 对公共因子 2 影响大的变量主要有：SUIT (适应能力)，EXP (经验)，FL (求职信的形式)，因此公共因子 2 可以称作“经验与适应”因子.

(e) [2 分] 该公司准备录用其中 6 人，利用公共因子得分对该公司的录用结果给出建议.

【解】 先做 48 位求职者在公共因子平面上的散点图，结果如图10.9所示，作图的 R 代码如下.

```
y = x.fa$scores
y
plot(y, type = "n", xlim = c(-3, 2), ylim = c(-2, 3), asp = 1,
     xlab = "Factor 1", ylab = "Factor 2")
abline(h = 0, v = 0, lty = 2, col = "cyan")
label2 = as.character(Applicants$ID)
points(y, pch = 16, col = 'orange', cex = 1.5)
text(y, label2)
```

首先计算第 1、2 公共因子的贡献率，结果分别为 70.48%、15.48%，利用贡献率作为因子得分的权重系数对第 1、2 公共因子的得分值进行加权求和，作为综合排序的依据，排序的结果 (前 20 位) 如图10.10所示. 若录用其中 6 人，建议录用得分排名前 6 的求职者，即录用第 23、40、39、22、24、2 号求职人员.

```
a = x.fa$values # 公共因子特征值
b = sum(a > 0) # 特征值大于零的个数
r = a[1:b]
r = r / sum(a[1:b])
round(r, digits = 4)[1:2] # 第1、2公共因子的贡献率
z = r[1] * y[, 1] + r[2] * y[, 2]
names(z) = label2
results = round(sort(z, decreasing = TRUE), digits = 2)
results = as.data.frame(results)
results
```

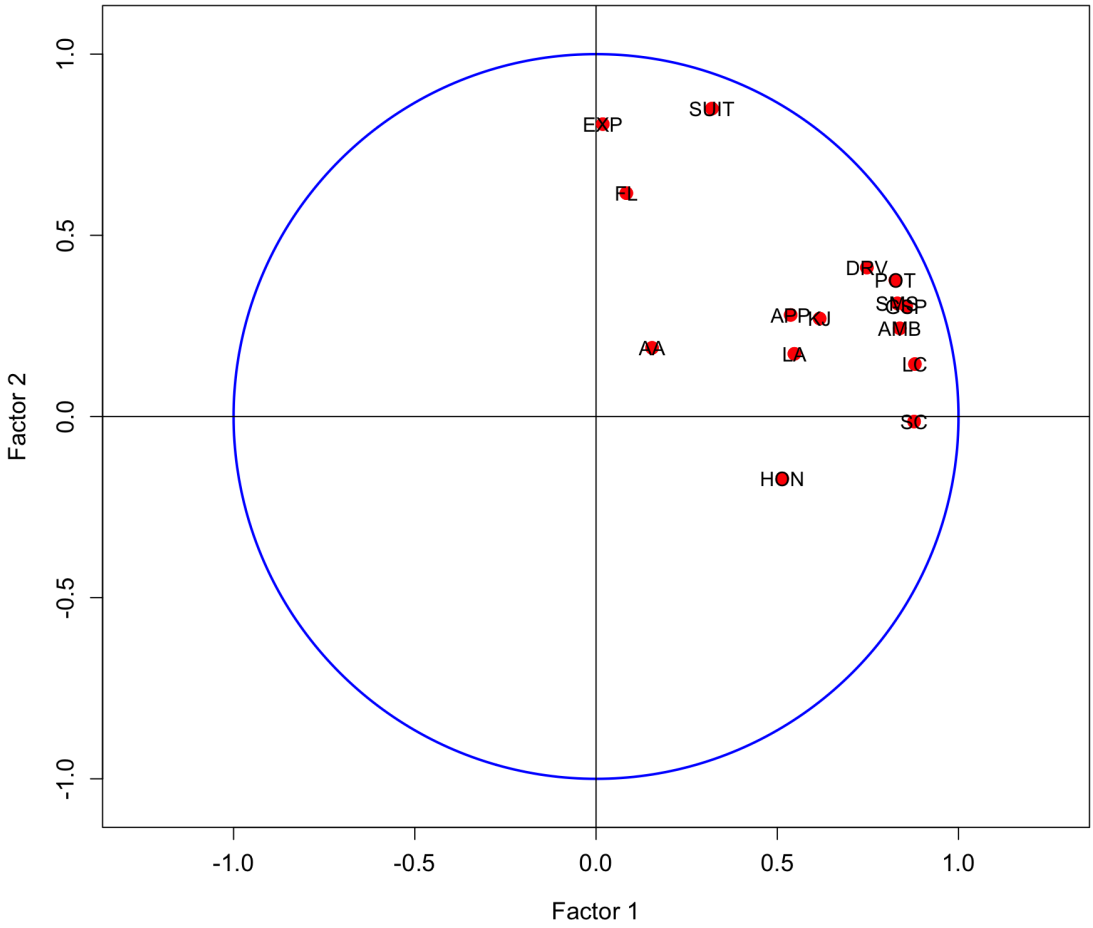

图 10.8: 变量在公共因子平面上的散点图.

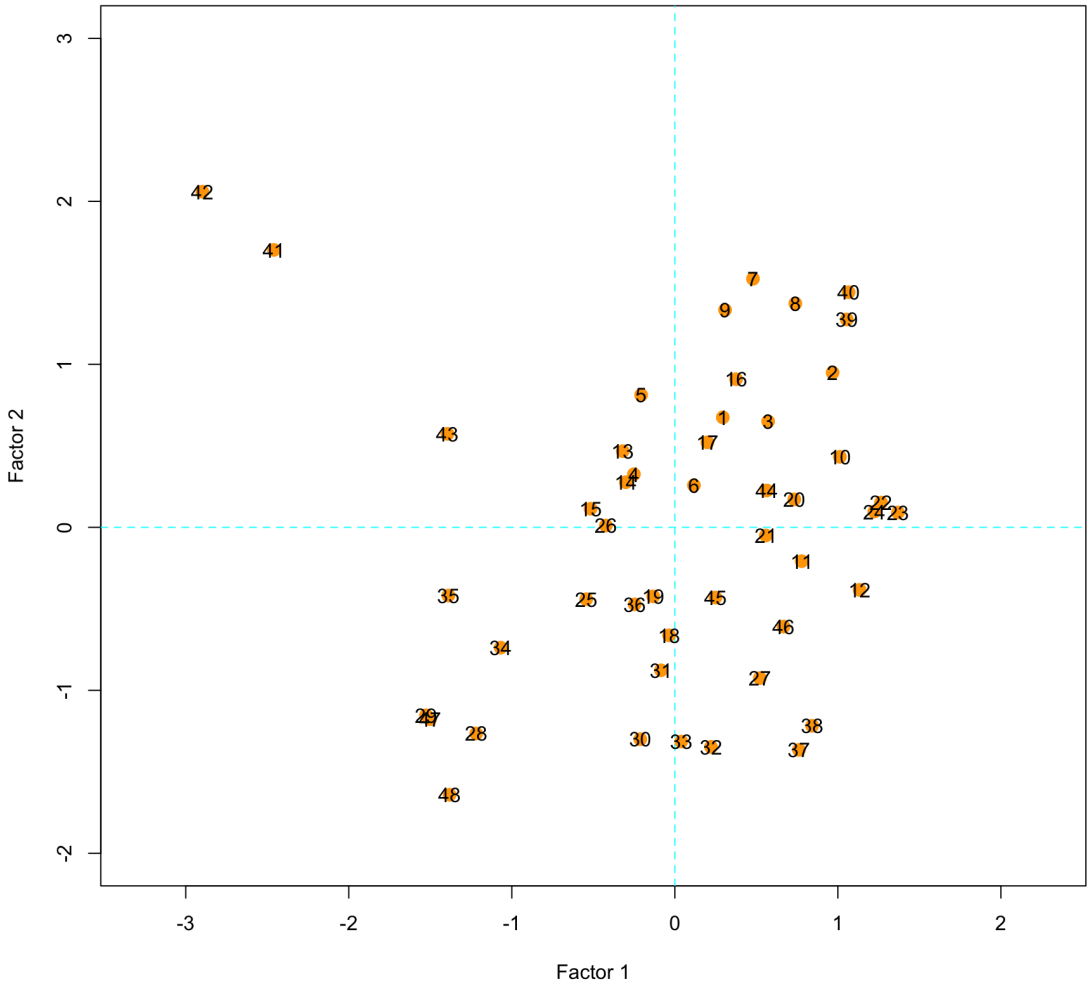

图 10.9: 48 位求职者在公共因子平面上的散点图.

```
> results
   results
23    0.98
40    0.97
39    0.94
22    0.91
24    0.88
2     0.83
10    0.78
12    0.74
8     0.73
7     0.57
20    0.54
11    0.52
3     0.50
44    0.43
9     0.42
16    0.41
38    0.40
21    0.39
46    0.38
37    0.32
```

图 10.10: 综合得分位于前 20 名的求职者.