Data Science Homework 2

R11922066 資工碩一 林庭安

Problem 1 Goodness of Estimation

(1)

Set $c_1 + c_2 = 1$

$$
\begin{aligned}
Var(c_1\theta_1 + c_2\theta_2) &= c_1^2 Var(\theta_1) + c_2^2 Var(t\theta_2) + 2c_1c_2Cov(\theta_1, \theta_2) \\
&= c_1^2 + 2c_2^2 + \frac{1}{2}c_1c_2 \\
&= c_1^2 + 2(1-c_1)^2 + \frac{1}{2}c_1(1-c_1) \\
&= \frac{5}{2}(c_1 - \frac{7}{10})^2 + \frac{31}{40}
\end{aligned}
$$

Minimizing this expression, we have $c_1 = \frac{7}{10}, c_2 = \frac{3}{10}$

$Var(c_1\theta_1 + c_2\theta_2) = \frac{31}{40}$

The minimum variance estimator of this unbiased class is $\frac{7}{10}\theta_1 + \frac{3}{10}\theta_2$

(2)

Set $c_1 + c_2 = 1$

$$
\begin{aligned}
Var(c_1\theta_3 + c_2\theta_4) &= c_1^2 Var(\theta_3) + c_2^2 Var(\theta_4) + 2c_1c_2Cov(\theta_3, \theta_4) \\
&= c_1^2 + 2c_2^2 + \frac{3}{2}c_1c_2 \\
&= c_1^2 + 2(1-c_1)^2 + \frac{3}{2}c_1(1-c_1) \\
&= \frac{3}{2}(c_1 - \frac{5}{6})^2 + \frac{23}{24}
\end{aligned}
$$

Minimizing this expression, we have $c_1 = \frac{5}{6}, c_2 = \frac{1}{6}$

$Var(c_1\theta_3 + c_2\theta_4) = \frac{23}{24}$

The minimum variance estimator of this unbiased class is $\frac{5}{6}\theta_3 + \frac{1}{6}\theta_4$

Problem 2 Interval Estimation

(1)

Let Y = X - θ , then

$f(x;\theta) = e^{-(x-\theta)}$, if $\theta < x < \infty$

$= g(y) = e^{-y}$, if $0 < y < \infty$

$\Longrightarrow Y \sim Exp(1)$

$$Q = X(1) - \theta$$
$$= min(X_1, X_2, ..., X_n) - \theta$$
$$= min(Y_1, Y_2, ..., Y_n)$$

Let $F_Q(q)$ be the cumulative distribution function of Q. Then

$$F_Q(q) = P(Q \leq q)$$
$$= P(min(Y_1, Y_2, ..., Y_n) \leq q)$$
$$= 1 - P(min(Y_1, Y_2, ..., Y_n) > q)$$
$$= 1 - P(Y_1 > q)^n$$
$$= 1 - [1 - F_{Y_1}(q)]^n$$

Thus, the probability density function of Q is

$$f_Q(q) = n[1 - F_{Y_1}(q)]^{n-1} f_{Y_1}(q)$$
$$= n[1 - \int_0^q e^{-y_1} dy_1]^{n-1} e^{-q}$$
$$= n[1 - (-e^{-y_1}\Big|_0^q\ )]^{n-1} e^{-q}$$
$$= ne^{-q(n-1)} e^{-q}$$
$$= ne^{-nq}, \text{ if } 0 < q < \infty$$
$$\implies Q \sim Exp(n)$$

$\because f_Q(q)$ is independent of θ

$\therefore$ Q is a pivotal quantity

(2)

$$P(a \leq Q \leq b) = 1 - \alpha$$

$$\begin{cases} \int_0^a ne^{-nq}\ dq = \frac{\alpha}{2} \\ \int_b^\infty ne^{-nq}\ dq = \frac{\alpha}{2} \end{cases} \implies \begin{cases} a = \frac{ln(1 - \frac{\alpha}{2})}{-n} \\ b = \frac{ln(\frac{\alpha}{2})}{-n} \end{cases}$$

$$1 - \alpha = P(a \leq Q \leq b)$$
$$= P(a \leq X(1) - \theta \leq b)$$
$$= P(-b \leq \theta - X(1) \leq -a)$$
$$= P(X(1) - b \leq \theta \leq X(1) - a)$$
$$= P(X(1) + \frac{ln(\frac{\alpha}{2})}{n} \leq \theta \leq X(1) + \frac{ln(1 - \frac{\alpha}{2})}{n})$$

The $100(1-\alpha)\%$ confidence interval for θ is $[X(1) + \frac{ln(\frac{\alpha}{2})}{n})\ ,\ X(1) + \frac{ln(1 - \frac{\alpha}{2})}{n}]$

Problem 3 Type I and Type II Errors

$$\begin{aligned} P(\text{Type I Error}) &= P(\text{Reject } H_o \mid H_o \text{ is true}) \\ &= P(X > 0.92 \mid H_o : \theta = 1) \\ &= \int_{0.92}^{1} \frac{1}{\theta} dx \\ &= \int_{0.92}^{1} 1 dx \\ &= 0.08 \end{aligned}$$

$$\begin{aligned} P(\text{Type II Error}) &= P(\text{Accept } H_o \mid H_a \text{ is true}) \\ &= P(X \leq 0.92 \mid H_a : \theta = 2) \\ &= \int_{0}^{0.92} \frac{1}{\theta} dx \\ &= \int_{0}^{0.92} \frac{1}{2} dx \\ &= 0.46 \end{aligned}$$

Problem 4 Hypothesis Testing

Since each observation is $X \sim Pois(\lambda = 0.5)$

The sum of observation is $\sim Pois(\lambda = 4)$

(1)

$$\begin{aligned} \alpha &= P(\text{Type I Error}) \\ &= P(\text{Reject } H_o \mid H_o \text{ is true}) \\ &= P(\sum_{i=1}^{8} x_i \geq 8 \mid H_o : \lambda = 0.5) \\ &= 1 - P(\sum_{i=1}^{8} x_i < 8 \mid H_o : \lambda = 0.5) \\ &= 1 - \sum_{k=0}^{7} \frac{e^{-4} 4^k}{k!} \\ &= 1 - 0.9489 \\ &= 0.0511 \end{aligned}$$

(2)

$$\begin{aligned} P(\text{Type II Error}) &= P(\text{Accept } H_o \mid H_a \text{ is true}) \\ &= P(\sum_{i=1}^{8} x_i < 8 \mid \lambda \neq 0.5) \\ &= \sum_{k=0}^{7} \frac{e^{\lambda} \lambda^k}{k!} \end{aligned}$$

$$\beta(\lambda) = \begin{cases} P(\text{Type I Error}) & \text{if } \lambda = 0.5 \\ 1 - P(\text{Type II Error}) & \text{if } \lambda \neq 0.5 \end{cases}$$

$$= \begin{cases} 0.0511 & \text{if } \lambda = 0.5 \\ 1 - \sum_{k=0}^{7} \frac{e^{\lambda}\lambda^{k}}{k!} & \text{if } \lambda \neq 0.5 \end{cases}$$

Problem 5 LDA

```
→  HW2 python3 main.py
eigenvalues:
 [ 8.68732138e-01 -1.92627984e-17]

eigenvectors:
 [[ 0.99130435 -0.37139068]
 [-0.13158907  0.92847669]]
```

```python
import numpy as np
c1 = np.array([[5, 3], [3, 5], [3, 4], [4, 5], [4, 7], [5, 6]])
c2 = np.array([[9, 10], [7, 7], [8, 5], [8, 8], [7, 2], [10, 8]])

def getMean(c):
    return np.array([np.mean(c[:, 0]), np.mean(c[:, 1])])

def WithinMatrix(c, m):
    return ( np.matmul(  np.transpose(c-m), c-m ) )

mean1 = getMean(c1)
mean2 = getMean(c2)
mean = (mean1 + mean2) / 2

sw1 = WithinMatrix(c1, mean1)
sw2 = WithinMatrix(c2, mean2)
sw = sw1 + sw2

sb1 = np.matmul( (mean1 - mean).reshape([2,1]), np.transpose((
   mean1 - mean).reshape([2,1])) )
sb2 = np.matmul( (mean2 - mean).reshape([2,1]), np.transpose((
   mean2 - mean).reshape([2,1])) )
```

```
sb = sb1 + sb2

eigval, eigvec = np.linalg.eig( np.matmul(np.linalg.inv(sw), sb) )

print("eigenvalues:\n",eigval,"\n")
print("eigenvectors:\n",eigvec)
```