Chapter 15

1

	θ_1	θ_2	θ_3
a_1	2	1	5
a_2	0	4	6
a_3	4	0	2

	θ_1	θ_2	θ_3
x_1	0.4	0.5	0.3
x_2	0.6	0.5	0.7

(a) Let d_i be a decision rule, $i = 1, 2, 3$.Consider

d_i	x_1	x_2
d_1	a_1	a_2
d_2	a_2	a_3
d_3	a_3	a_1

(Note. You might consider other decisions.)

For $j = 1, 2, 3$, and i-th decision function $d_i, i = 1, 2, 3$,

$R(\theta_j, d_i) = \mathbb{E}_X\left[L(\theta_j, d_i(X))\right] = \sum_{k=1}^{2} L(\theta_j, d_i(x_k))P(X = x_k \mid \theta_j)$.

Thus,

$$R(\theta_1, d_1) = 0.4 \times 2 + 0.6 \times 0 = 0.8$$
$$R(\theta_2, d_1) = 0.5 \times 1 + 0.5 \times 4 = 2.5$$
$$R(\theta_3, d_1) = 0.3 \times 5 + 0.7 \times 6 = 5.7$$
$$R(\theta_1, d_2) = 0.4 \times 0 + 0.6 \times 4 = 2.4$$
$$R(\theta_2, d_2) = 0.5 \times 4 + 0.5 \times 0 = 2$$
$$R(\theta_3, d_2) = 0.3 \times 6 + 0.7 \times 2 = 3.2$$
$$R(\theta_1, d_3) = 0.4 \times 4 + 0.6 \times 2 = 2.8$$
$$R(\theta_2, d_3) = 0.5 \times 0 + 0.5 \times 1 = 0.5$$
$$R(\theta_3, d_3) = 0.3 \times 2 + 0.7 \times 5 = 4.1$$

(b)

$$\max_j R(\theta_j, d_1) = 5.7$$
$$\max_j R(\theta_j, d_2) = 3.2 \quad \Rightarrow \quad d_2 \text{ is the minimax rule.}$$
$$\max_j R(\theta_j, d_3) = 4.1$$

(c) Prior： $g(\theta_j) = \frac{1}{3}$ for $j = 1, 2, 3$

Bayes risk： $B(d) = R(\theta_1, d)g(\theta_1) + R(\theta_2, d)g(\theta_2) + R(\theta_3, d)g(\theta_3)$

Thus,

$$B(d_1) = R(\theta_1, d_1)g(\theta_1) + R(\theta_2, d_1)g(\theta_2) + R(\theta_3, d_1)g(\theta_3)$$
$$= 0.8 \times \tfrac{1}{3} + 2.5 \times \tfrac{1}{3} + 5.7 \times \tfrac{1}{3} = 3$$
$$B(d_2) = 2.4 \times \tfrac{1}{3} + 2 \times \tfrac{1}{3} + 3.2 \times \tfrac{1}{3} = \tfrac{38}{15}$$
$$B(d_3) = 2.8 \times \tfrac{1}{3} + 0.5 \times \tfrac{1}{3} + 4.1 \times \tfrac{1}{3} = \tfrac{37}{15}$$

$\Rightarrow$ d_3 is the Bayes rule corresponding to that prior.

(d)

$$h(\theta_1 \mid x_2) = \frac{f(x_2 \mid \theta_1)g(\theta_1)}{\sum_{j=1}^{3} f(x_2 \mid \theta_j)g(\theta_j)} = \frac{0.6 \times \frac{1}{3}}{(0.6 + 0.5 + 0.7) \times \frac{1}{3}} = \frac{0.2}{0.6} = \frac{1}{3}$$
$$h(\theta_2 \mid x_2) = \frac{0.5 \times \frac{1}{3}}{1.8 \times \frac{1}{3}} = \frac{5}{18}$$
$$h(\theta_3 \mid x_2) = 1 - \frac{1}{3} - \frac{5}{18} = \frac{7}{18}$$

Then the posterior risk(PR) for a_1, a_2 and a_3 are

$$\mathrm{PR}(a_1 \mid x_2) = l(\theta_1, a_1)h(\theta_1 \mid x_2) + l(\theta_2, a_1) \cdot h(\theta_2 \mid x_2) + l(\theta_3, a_1)h(\theta_3 \mid x_2)$$
$$= 2 \times \frac{1}{3} + 1 \times \frac{5}{18} + 5 \times \frac{7}{18} = \frac{26}{9}$$
$$\mathrm{PR}(a_2 \mid x_2) = 0 \times \frac{1}{3} + 4 \times \frac{5}{18} + 6 \times \frac{7}{18} = \frac{31}{9}$$
$$\mathrm{PR}(a_3 \mid x_2) = 4 \times \frac{1}{3} + 0 \times \frac{5}{18} + 6 \times \frac{7}{18} = \frac{19}{9}.$$

Hence,a_3 has the smallest posterior risk, and is the Bayes action for x_2.

6

If X comes from class A $\sim N(0, 1)$,and if X comes from class B $\sim N(2, 1)$.

(a) We consider a classification problem between two classes, A and B:

- Class A: $X \mid A \sim N(0, 1) \Rightarrow f_A(x) = \phi(x)$
- Class B: $X \mid B \sim N(2, 1) \Rightarrow f_B(x) = \phi(x - 2)$
- Prior: $\pi_A = \pi_B = \frac{1}{2}$

- Loss function: 0—1 loss

Here, $\phi(x)$ is the probability density function of $N(0,1)$.

Then the general Bayes rule for classification under 0—1 loss is:

$$\delta(x) = \arg\max_{\theta_i} P(\theta_i \mid x)$$

That is, classify the observation to the class with the highest posterior probability.

Using Bayes' theorem, the posterior probabilities are:

$$P(A \mid x) = \frac{\pi_A f(x \mid A)}{\pi_A f(x \mid A) + \pi_B f(x \mid B)} = \frac{\phi(x)}{\phi(x) + \phi(x-2)}$$

$$P(B \mid x) = \frac{\phi(x-2)}{\phi(x) + \phi(x-2)}$$

Since the priors are equal and the loss is symmetric, the classification decision becomes:

$$P(A \mid x) > P(B \mid x) \iff \phi(x) > \phi(x-2)$$

Hence, the Bayes rule simplifies to:

$$\delta(x) = \begin{cases} A, & \text{if } \phi(x) > \phi(x-2) \\ B, & \text{otherwise} \end{cases}$$

In full generality, the posterior risk of classifying to class j is:

$$PR(j \mid x) = \sum_i l_{ij} \cdot P(\theta_i \mid x)$$

Under 0—1 loss:

$$l_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \Rightarrow PR(j \mid x) = 1 - P(\theta_j \mid x)$$

Therefore, minimizing posterior risk is equivalent to maximizing posterior probability:

$$\arg\min_j PR(j \mid x) = \arg\max_j P(\theta_j \mid x)$$

Hence,using the symmetry and unimodal shape of the standard normal density, we find:

- Decision threshold: $x = 1$
- If $x < 1$, choose A; if $x > 1$, choose B

(b) Given that $x = 1.5 > 1$, the Bayes rule classifies the observation as class B. We compute the probability of misclassification, which is $P(A \mid x = 1.5)$:

$$P(A \mid x) = \frac{\phi(1.5)}{\phi(1.5) + \phi(-0.5)}$$

From standard normal density tables:

$$\phi(1.5) \approx 0.1295, \quad \phi(-0.5) = \phi(0.5) \approx 0.3521$$

Hence,

$$P(A \mid x = 1.5) \approx \frac{0.1295}{0.1295 + 0.3521} \approx 0.269$$

This is the classification error probability at $x = 1.5$.

(c) According to the Bayes risk formula for 0–1 loss:

$$\text{Error} = \frac{1}{2} P_A(X > 1) + \frac{1}{2} P_B(X < 1)$$

From standard normal CDF Φ:

$$P_A(X > 1) = 1 - \Phi(1) \approx 0.1587, \quad P_B(X < 1) = \Phi(-1) \approx 0.1587$$

Therefore, the total Bayes error is:

$$\text{Error} \approx \frac{1}{2}(0.1587 + 0.1587) = 0.1587$$

(d) Assume that the cost of misclassifying class A as B is twice the cost of misclassifying class B as A.

- Loss of misclassifying A: $L(A \to B) = 2$
- Loss of misclassifying B: $L(B \to A) = 1$

Then redo parts (a)–(c) under this assumption.

(a) With equal priors $\pi_A = \pi_B = \frac{1}{2}$, the Bayes rule minimizes posterior risk. The decision rule becomes:

$$\frac{P(A \mid x)}{P(B \mid x)} > \frac{1}{2} \Rightarrow \frac{\phi(x)}{\phi(x-2)} > \frac{1}{2}$$

Recall that:

$$\frac{\phi(x)}{\phi(x-2)} = \exp(2x - 2) \Rightarrow \exp(2x - 2) > 2 \Rightarrow 2x - 2 > \log(2) \Rightarrow x > \frac{2 + \log(2)}{2} \approx 1.3466$$

Decision Rule:

$$\delta(x) = \begin{cases} A, & x < 1.3466 \\ B, & x > 1.3466 \end{cases}$$

(b) Since $x = 1.5 > 1.3466$, classify as class B. The misclassification probability is:

$$P(A \mid x = 1.5) = \frac{\phi(1.5)}{\phi(1.5) + \phi(-0.5)} \approx \frac{0.1295}{0.1295 + 0.3521} \approx 0.269$$

(c) Using the new decision boundary $x = 1.3466$, the total classification error is:

$$\text{Error} = \frac{1}{2} \cdot P_A(X > 1.3466) + \frac{1}{2} \cdot P_B(X < 1.3466)$$

From standard normal CDF:

$$P_A(X > 1.3466) = 1 - \Phi(1.3466) \approx 1 - 0.9102 = 0.0898$$
$$P_B(X < 1.3466) = \Phi(1.3466 - 2) = \Phi(-0.6534) \approx 0.256$$

$$\text{Total Error} = \frac{1}{2}(0.0898 + 0.256) = 0.173$$

(e) Redo parts (a)–(c) with changed prior: $\pi_A = \frac{2}{3}, \pi_B = \frac{1}{3}$

(a)

$$\frac{\pi_A \cdot \phi(x)}{\pi_B \cdot \phi(x-2)} > 1 \Rightarrow \frac{2}{3} \cdot \phi(x) > \frac{1}{3} \cdot \phi(x-2) \Rightarrow \frac{\phi(x)}{\phi(x-2)} > \frac{1}{2}$$

This is the same inequality as in part (d), so the classification boundary remains:

$$x > \frac{2 + \log 2}{2} \approx 1.347$$

(b) Since $x = 1.5 > 1.347$, classify as class B.

$$P(A \mid x = 1.5) = \frac{\frac{2}{3} \cdot \phi(1.5)}{\frac{2}{3} \cdot \phi(1.5) + \frac{1}{3} \cdot \phi(-0.5)} = \frac{0.0863}{0.0863 + 0.1174} \approx 0.424$$

(c)

$$\text{Error} = \frac{2}{3} \cdot P_A(X > 1.347) + \frac{1}{3} \cdot P_B(X < 1.347)$$

Using standard normal values:

$$P_A(X > 1.347) = 1 - \Phi(1.347) \approx 0.0888$$
$$P_B(X < 1.347) = \Phi(1.347 - 2) = \Phi(-0.653) \approx 0.256$$

$$\text{Error} = \frac{2}{3} \cdot 0.0888 + \frac{1}{3} \cdot 0.256 \approx 0.145$$

9

	$H : p \leq \frac{1}{2}$	$K : p > \frac{1}{2}$
$a_1 : chooseH$	0	1
$a_2 : chooseK$	2	0

	$P(X \mid p)$
$X=0$	$(1-p)^3$
$X=1$	$3p(1-p)^2$
$X=2$	$3p^2(1-p)$
$X=3$	p^3

	$X=0$	$X=1$	$X=2$	$X=3$
d_1	a_1	a_2	a_2	a_2
d_2	a_1	a_1	a_2	a_2
d_3	a_1	a_1	a_1	a_2

For a given decision rule d, the risk function is defined as:

Let $R(p,d) = \begin{cases} 1 \text{ P(d choose H} \mid p) \text{ if } p > \frac{1}{2} \\ 2 \text{ P(d choose K} \mid p) \text{ if } p \leq \frac{1}{2} \end{cases}$

The Bayes risk is computed by integrating $R(p,d)$ over $p \in (0,1)$.

Because $X \sim \text{Bin}(3,p)$, we have

$$P(X=0) = (1-p)^3$$
$$P(X=1) = 3p(1-p)^2$$
$$P(X=2) = 3p^2(1-p)$$
$$P(X=3) = p^3$$

Rule d_1: Choose K if $X > 0$, otherwise H

- Choose H only when $X = 0$
- If $p > 1/2$: risk $= 1(1 - P(X > 0)) = (1-p)^3$
- If $p \leq 1/2$: risk $= 2P(X > 0) = 2[1 - (1-p)^3]$

Bayes risk:

$$B(d_1) = \int_0^{1/2} 2[1-(1-p)^3]\,dp + \int_{1/2}^1 (1-p)^3\,dp = 0.53125 + 0.015625 = 0.546875$$

Rule d_2: Choose H if $X \leq 1$, otherwise K

- Choose H if $X = 0$ or $X = 1$
- If $p > 1/2$: risk $= 1[(1-p)^3 + 3p(1-p)^2]$
- If $p \leq 1/2$: risk $= 2[P(X=2) + P(X=3)] = 2[3p^2(1-p) + p^3]$

Bayes risk:

$$B(d_2) = \int_0^{1/2} 2[3p^2(1-p) + p^3]\,dp + \int_{1/2}^1 [(1-p)^3 + 3p(1-p)^2]\,dp = 0.1875 + 0.09375 = 0.28125$$

Rule d_3: Choose H if $X \leq 2$, otherwise K

- Choose K only when $X = 3$
- If $p \leq 1/2$: risk $= 2p^3$
- If $p > 1/2$: risk $= 1[(1-p)^3 + 3p(1-p)^2 + 3p^2(1-p)]$

Bayes risk:

$$B(d_3) = \int_0^{1/2} 2p^3\,dp + \int_{1/2}^1 \left[(1-p)^3 + 3p(1-p)^2 + 3p^2(1-p)\right] dp = 0.03125 + 0.265625 = 0.296875$$

Comparison:

Rule	Bayes Risk
d_1	0.546875
d_2	0.28125
d_3	0.296875

Hence, Rule d_2 yields the smallest Bayes risk and is thus the optimal decision rule among the three.

11

Suppose $X \sim \text{Bin}(2, p)$. We compare the risk functions of the following three estimators of p, under squared error loss:

$$\hat{p}_1 = \frac{X}{2}, \quad \hat{p}_2 = \frac{X+1}{3}, \quad \hat{p}_3 = \frac{X+1}{4}$$

The loss function is the squared error loss:

$$L(p, \hat{p}) = (\hat{p} - p)^2$$

The risk function is defined as the expected loss:

$$R(p) = \mathbb{E}_p[(\hat{p}(X) - p)^2] = \sum_{x=0}^{2} (\hat{p}(x) - p)^2 \cdot \binom{2}{x} p^x (1-p)^{2-x}$$

Under square error loss, risk = mean square error = $\text{Var}(\hat{p_i}) + bias(\hat{p_i})^2$

- $\hat{p}_1 = \frac{X}{2}$

This is the maximum likelihood estimator (MLE), which is unbiased. Its risk function is:

$$R_1(p) = \mathrm{Var}_p\left(\frac{X}{2}\right) = \frac{1}{4}\cdot \mathrm{Var}(X) = \frac{1}{4}\cdot 2p(1-p) = \frac{p(1-p)}{2}$$

- $\hat{p}_2 = \frac{X+1}{3}$

This is a Bayes estimator under a Beta(1,1) prior (uniform prior). The risk is:

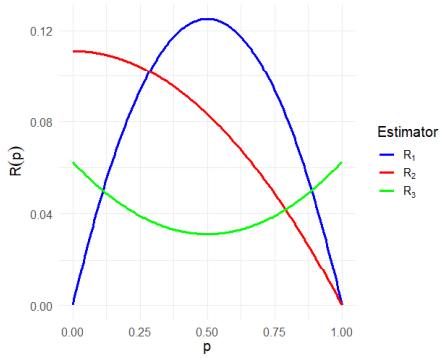
$$R_2(p) = \sum_{x=0}^{2}\left(\frac{x+1}{3} - p\right)^2 \cdot \binom{2}{x} p^x (1-p)^{2-x} = Var(\frac{X+1}{3}) + (E(\frac{X+1}{3}) - p)^2 = \frac{2p(1-p)}{9} + (\frac{2p+1}{3} - p)^2$$

- $\hat{p}_3 = \frac{X+1}{4}$

A shrinkage estimator with more bias but potentially lower variance. Risk function:

$$R_3(p) = \sum_{x=0}^{2}\left(\frac{x+1}{4} - p\right)^2 \cdot \binom{2}{x} p^x (1-p)^{2-x} = Var(\frac{X+1}{4}) + (E(\frac{X+1}{34}) - p)^2 = \frac{2p(1-p)}{16} + (\frac{2p+1}{4} - p)^2$$

Using numerical evaluation and plots, we compare the risk functions $R_1(p), R_2(p), R_3(p)$ over $p \in [0,1]$.



Based on the risk function plots, we observe that:

Each estimator $\hat{p}_1, \hat{p}_2, \hat{p}_3$ performs better in different intervals of $p \in [0,1]$, but none of them has a uniformly lower risk function than the others.

Therefore:

$$\forall i \neq j, \quad \nexists\, R(p,\hat{p}_i) \leq R(p,\hat{p}_j) \text{ for all } p \in [0,1] \text{ and strict for some } p$$

$$\Rightarrow \text{No estimator dominates another. All three are admissible.}$$

13

$$k \sim Bin(n,p),\ \pi(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0,1,...,n$$

Let $X = \begin{cases} 1, \text{ if the item is defective} \\ 0, \text{ otherwise} \end{cases}$

$$P(X = x|k) = (\frac{k}{n})^x (1 - \frac{k}{n})^{1-x}$$

$$P(k|X = x) = \frac{P(X = x|k)\pi(k)}{P(X = x)} = \frac{(\frac{k}{n})^x (1 - \frac{k}{n})^{1-x} \binom{n}{k} p^k (1-p)^{n-k}}{\sum_{k=0}^{n} P(X = x|k) \times \pi(k)}$$

(a)

$$\begin{aligned} P(k|X = 1) &= \frac{(\frac{k}{n})^1 (1 - \frac{k}{n})^0 \binom{n}{k} p^k (1-p)^{n-k}}{P(X = 1)} \\ &= \frac{(\frac{k}{n}) \binom{n}{k} p^k (1-p)^{n-k}}{\sum_{k=0}^{n} \frac{k}{n} \cdot \pi(k)} \\ &= \frac{(\frac{k}{n}) \binom{n}{k} p^k (1-p)^{n-k}}{E(\frac{k}{n})} \\ &= \frac{(\frac{k}{n}) \binom{n}{k} p^k (1-p)^{n-k}}{p} \\ &= \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \end{aligned}$$

(b)

$$\begin{aligned}
P(k|X=0) &= \frac{(\frac{k}{n})^0(1-\frac{k}{n})^1\binom{n}{k}p^k(1-p)^{n-k}}{P(X=0)} \\
&= \frac{(1-\frac{k}{n})\binom{n}{k}p^k(1-p)^{n-k}}{\sum_{k=0}^{n}(1-\frac{k}{n})\pi(k)} \\
&= \frac{(1-\frac{k}{n})\binom{n}{k}p^k(1-p)^{n-k}}{E(1-\frac{k}{n})} \\
&= \frac{(1-\frac{k}{n})\binom{n}{k}p^k(1-p)^{n-k}}{1-p} \\
&= \binom{n-1}{k}p^k(1-p)^{n-k-1}
\end{aligned}$$

15

	θ_1	θ_2	θ_3
x_1	0.1	0.2	0.4
x_2	0.1	0.2	0.2
x_3	0.2	0.2	0.2
x_4	0.6	0.4	0.2

(a)

$$\begin{aligned}
f(\theta_1|x_2) &= \frac{f(x_2|\theta_1)p(\theta_1)}{\sum_{i=1}^{3}f(x_2|\theta_i)p(\theta_i)} = \frac{0.1\times 0.5}{0.1\times 0.5+0.2\times 0.25+0.2\times 0.25} = \frac{1}{3} \\
f(\theta_2|x_2) &= \frac{f(x_2|\theta_2)p(\theta_2)}{\sum_{i=1}^{3}f(x_2|\theta_i)p(\theta_i)} = \frac{1}{3} \\
f(\theta_3|x_2) &= \frac{f(x_2|\theta_3)p(\theta_3)}{\sum_{i=1}^{3}f(x_2|\theta_i)p(\theta_i)} = \frac{1}{3}
\end{aligned}$$

(b) Under squared error loss, the Bayesian rule to minimize the posterior risk is：

$$\hat{\theta} = E(\Theta|X=x_2) = (1+10+20)\times\frac{1}{3} = \frac{31}{3}$$

提醒：當$X=x_2$時$\hat{\theta}$並不為$\theta_1,\theta_2,\theta_3$中的某個值。

(c) Under absolute error loss, the Bayesian rule to minimize the posterior risk is：

$$\hat{\theta} = (\text{the median of } \Theta \text{ given } X = x_2) = \theta_2 = 10$$

(d)

$$f(\theta_1|x_2, x_1) = \frac{f(x_2, x_1|\theta_1)p(\theta_1)}{\sum_{i=1}^{3} f(x_2, x_1|\theta_i)p(\theta_i)} = \frac{0.1 \times 0.1 \times 0.5}{0.1 \times 0.1 \times 0.5 + 0.2 \times 0.2 \times 0.25 + 0.4 \times 0.2 \times 0.25} = \frac{1}{7}$$
$$f(\theta_2|x_2, x_1) = \frac{f(x_2, x_1|\theta_2)p(\theta_2)}{\sum_{i=1}^{3} f(x_2, x_1|\theta_i)p(\theta_i)} = \frac{2}{7}$$
$$f(\theta_3|x_2, x_1) = \frac{f(x_2, x_1|\theta_3)p(\theta_3)}{\sum_{i=1}^{3} f(x_2, x_1|\theta_i)p(\theta_i)} = \frac{4}{7}$$

提醒：也可以用(a)中的θ分配當prior，對數據x_1做update後求得，例如：

$$f(\theta_1|x_2, x_1) = \frac{\frac{1}{3} \times 0.1}{\frac{1}{3} \times 0.1 + \frac{1}{3} \times 0.2 + \frac{1}{3} \times 0.4} = \frac{1}{7}$$

23

(a)

$$X_1, ..., X_9|\mu \overset{iid}{\sim} N(\mu, 9), and\ \mu \sim N(0, 1)$$

$$\begin{aligned}
f(x_1, ..., x_9) &= \int_{-\infty}^{\infty} f(x_1, ..., x_9|\mu)\pi(\mu)d\mu \\
&= (\frac{1}{\sqrt{18\pi}})^9 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp(-\frac{\sum_{i=1}^{9}(x_i - \mu)^2}{18} - \frac{\mu^2}{2})d\mu \\
&= (\frac{1}{\sqrt{18\pi}})^9 \frac{1}{\sqrt{2\pi}} exp(-\frac{\sum_{i=1}^{9} x_i^2}{18}) \int_{-\infty}^{\infty} exp(-\mu^2 + \bar{x}\mu)d\mu \\
&= (\frac{1}{\sqrt{18\pi}})^9 \frac{1}{\sqrt{2\pi}} exp(-\frac{\sum_{i=1}^{9} x_i^2}{18} + \frac{1}{4}\bar{x}^2) \int_{-\infty}^{\infty} exp(-(\mu - \frac{1}{2}\bar{x})^2)d\mu \\
&= (\frac{1}{\sqrt{18\pi}})^9 \frac{1}{\sqrt{2}} exp(-\frac{\sum_{i=1}^{9} x_i^2}{18} + \frac{1}{4}\bar{x}^2)
\end{aligned}$$

以上這段可以不用算，只需使用$f(\mu|\mathbf{x}) \propto f(\mathbf{x}|\mu) \cdot \pi(\mu)$，再觀察$f(\mathbf{x}|\mu) \cdot \pi(\mu)$中$\mu$的函數形式來找出對應的distribution即可。

$$\therefore f(\mu|x_1, ..., x_9) = \frac{f(x_1, .., x_9|\mu)\pi(\mu)}{f(x_1, ..., x_9)} = \frac{1}{\sqrt{\pi}} exp(-(\mu - \frac{1}{2}\bar{x})^2)$$
$$\Rightarrow \mu|x_1, .., x_9 \sim N(\frac{\bar{x}}{2}, \frac{1}{2})$$

precision$=1/Var(\mu|x_1, ...x_9) = 2$

(b) For squared error loss, the Bayes estimate of μ is $E(\mu|x_1,..,x_9)=\frac{\bar{x}}{2}$

$$\begin{aligned}E_\Theta(R(\mu,\frac{\bar{x}}{2}))&=\int_{-\infty}^{\infty}E((\frac{1}{2}\bar{x}-\mu)^2|\mu)\pi(\mu)d\mu\\&=\int_{-\infty}^{\infty}(Var(\frac{1}{2}\bar{x}|\mu)+E(\frac{1}{2}\bar{x}-\mu)^2)\pi(\mu)d\mu\\&=\int_{-\infty}^{\infty}(\frac{1}{4}+\frac{1}{4}\mu^2)\pi(\mu)d\mu\\&=\frac{1}{2}\end{aligned}$$

(c) A credibility interval is posterior probability interval,
so 95% credibility interval for μ is $\left[\frac{\bar{x}}{2}-Z_{0.025}\sqrt{\frac{1}{2}},\frac{\bar{x}}{2}+Z_{0.025}\sqrt{\frac{1}{2}}\right]$

24

$$X|\mu\sim N(\mu,0.25),and\ \mu\sim N(0,1)$$

$$\begin{aligned}f(\mu|x)&=\frac{f(X|\mu)\pi(\mu)}{f(x)}\\&=\frac{f(X|\mu)\pi(\mu)}{\int_{-\infty}^{\infty}f(X|\mu)\pi(\mu)d\mu}\\&=\frac{\frac{1}{\pi}exp(-\frac{\mu^2}{2}-\frac{(x-\mu)^2}{0.5})}{\int_{-\infty}^{\infty}\frac{1}{\pi}exp(-\frac{\mu^2}{2}-\frac{(x-\mu)^2}{0.5})d\mu}\\&=\frac{1}{\sqrt{2\pi(1/5)}}exp(-\frac{1}{2(1/5)}(\mu-\frac{4}{5}x)^2)\\&\Rightarrow\mu|x\sim N(\frac{4}{5}x,\frac{1}{5})\end{aligned}$$

提醒：與上一題相同，只需使用$f(\mu|\mathbf{x})\propto f(\mathbf{x}|\mu)\cdot\pi(\mu)$，再觀察$f(\mathbf{x}|\mu)\cdot\pi(\mu)$中$\mu$的函數形式來找出對應的distribution即可。

For squared error loss, the Bayes estimate of μ is $E(\mu|x)=\frac{4}{5}x$

注意：在square error loss下，risk function=MSE=Var+$bias^2$

$$\begin{aligned} R(\mu, \frac{4}{5}x) &= E(\frac{4}{5}x - \mu)^2 \\ &= Var(\frac{4}{5}x) + [E(\frac{4}{5}x) - \mu]^2 \\ &= \frac{16}{25} \times \frac{1}{4} + \frac{1}{25}\mu^2 \\ &= \frac{4 + \mu^2}{25} \end{aligned}$$

$$\begin{aligned} E_\Theta(R(\mu, \frac{4}{5}x)) &= \int_{-\infty}^{\infty} E((\frac{4}{5}x - \mu)^2|\mu)\pi(\mu)d\mu \\ &= \int_{-\infty}^{\infty} (Var(\frac{4}{5}x|\mu) + E(\frac{4}{5}x - \mu|\mu)^2)\pi(\mu)d\mu \\ &= \int_{-\infty}^{\infty} (\frac{16}{25} \times \frac{1}{4} + \frac{1}{25}\mu^2)\pi(\mu)d\mu \\ &= \frac{1}{5} \end{aligned}$$

The mle of μ is x

$$\begin{aligned} R(\mu, x) &= Var(x) + [E(x) - \mu]^2 \\ &= \frac{1}{4} \end{aligned}$$

$$E_\Theta(R(\mu, x)) = \int_{-\infty}^{\infty} E((x - \mu)^2|\mu)\pi(\mu)d\mu = \frac{1}{4}$$

When $-\frac{3}{2} < \mu < \frac{3}{2}$, Bayes estimator will have smaller risk than MLE.

When $\mu > \frac{3}{2}$ or $\mu < -\frac{3}{2}$, Bayes estimator will have larger risk than MLE.

29

$$X_1, ..., X_n|\lambda_0 \sim Exp(\lambda_0),\ and\ \lambda_0 \sim \Gamma(\alpha, \lambda)$$

$$\begin{aligned}
f(\lambda_0|x_1, ..., x_n) &= \frac{f(X_1, ..., X_n|\lambda_0)\pi(\lambda_0)}{f(x_1, ..., x_n)} \\
&= \frac{\lambda_0^n exp(-\lambda_0 \sum_{i=1}^n x_i)\frac{\lambda_0^{\alpha-1}exp(-\lambda\lambda_0)}{\Gamma(\alpha)(\frac{1}{\lambda})^\alpha}}{\int_0^\infty \lambda_0^n exp(-\lambda_0 \sum_{i=1}^n x_i)\frac{\lambda_0^{\alpha-1}exp(-\lambda\lambda_0)}{\Gamma(\alpha)(\frac{1}{\lambda})^\alpha} d\lambda_0} \\
&= \frac{\lambda_0^n exp(-\lambda_0 \sum_{i=1}^n x_i)\frac{\lambda_0^{\alpha-1}exp(-\lambda\lambda_0)}{\Gamma(\alpha)(\frac{1}{\lambda})^\alpha}}{\int_0^\infty \frac{\lambda_0^{n+\alpha-1}exp(-\lambda_0(\lambda+\sum_{i=1}^n x_i))}{\Gamma(\alpha)(\frac{1}{\lambda})^\alpha} d\lambda_0} \\
&= \frac{\lambda_0^n exp(-\lambda_0 \sum_{i=1}^n x_i)\frac{\lambda_0^{\alpha-1}exp(-\lambda\lambda_0)}{\Gamma(\alpha)(\frac{1}{\lambda})^\alpha}}{\frac{\Gamma(n+\alpha)(\frac{1}{\lambda+\sum_{i=1}^n x_i})^{n+\alpha}}{\Gamma(\alpha)(\frac{1}{\lambda})^\alpha}\int_0^\infty \frac{\lambda_0^{n+\alpha-1}exp(-\lambda_0(\lambda+\sum_{i=1}^n x_i))}{\Gamma(n+\alpha)(\frac{1}{\lambda+\sum_{i=1}^n x_i})^{n+\alpha}} d\lambda_0} \\
&= \frac{\lambda_0^{n+\alpha-1}exp(-\lambda_0(\lambda+\sum_{i=1}^n x_i))}{\Gamma(n+\alpha)(\frac{1}{\lambda+\sum_{i=1}^n x_i})^{n+\alpha}}
\end{aligned}$$

提醒：與上上一題相同，只需使用$f(\lambda_0|\mathbf{x}) \propto f(\mathbf{x}|\lambda_0) \cdot \pi(\lambda_0)$，再觀察$f(\mathbf{x}|\lambda_0) \cdot \pi(\lambda_0)$中$\lambda_0$的函數形式來找出對應的distribution即可。

Therefore, $\lambda_0|x_1, ..., x_n \sim \Gamma(n+\alpha, \lambda+\sum_{i=1}^n x_i)$，posterior 與 prior 都是 gamma distribution

⇒ gamma distribution is conjugate prior of the exponential distribution.