

Bayesian Analysis

Xiaozhu Zhang

Spring 2021

Contents

1	Introduction	2
2	The Prior Distributions	16
3	Bayesian Statistical Inference	30
4	Bayesian Decision Theory	55

1 Introduction

Ex. 1.1. Assume that θ is the defective rate of a batch of products. The prior for θ is

$$\pi(0.1) = 0.7, \quad \pi(0.2) = 0.3.$$

Suppose that 2 out of 8 randomly selected products from this batch are defective. Find the posterior distribution of θ .

Let the random variable X be the number of defective products, then $X \sim B(8, \theta)$. As a result,

$$\begin{aligned} \pi(\theta = 0.1|x) &= \frac{f(x|\theta = 0.1)\pi(\theta = 0.1)}{f(x|\theta = 0.1)\pi(\theta = 0.1) + f(x|\theta = 0.2)\pi(\theta = 0.2)} \\ &= \frac{C(8, 2) \cdot 0.1^2 \cdot 0.9^6 \cdot 0.7}{C(8, 2) \cdot 0.1^2 \cdot 0.9^6 \cdot 0.7 + C(8, 2) \cdot 0.2^2 \cdot 0.8^6 \cdot 0.3} \\ &= 0.5418. \end{aligned} \tag{1}$$

Similarly,

$$\pi(\theta = 0.2|x) = 0.4582.$$

□

Ex. 1.2. Assume that the number of defects on a tape follows a Poisson distribution $P(\lambda)$. The prior for λ is

$$\pi(1.0) = 0.4, \quad \pi(1.5) = 0.6.$$

Suppose that 3 defects are found on a tape. Find the posterior distribution of λ .

Let the random variable X be the number of defects on a tape, then $X \sim P(\lambda)$. As a result,

$$\begin{aligned} \pi(\lambda = 1.0|x) &= \frac{f(x|\lambda = 1.0)\pi(\lambda = 1.0)}{f(x|\lambda = 1.0)\pi(\lambda = 1.0) + f(x|\lambda = 1.5)\pi(\lambda = 1.5)} \\ &= \frac{(e^{-1.0} \cdot 1.0^3)/(3!) \cdot 0.4}{(e^{-1.0} \cdot 1.0^3)/(3!) \cdot 0.4 + (e^{-1.5} \cdot 1.5^3)/(3!) \cdot 0.6} \\ &= 0.2457. \end{aligned} \tag{2}$$

Similarly,

$$\pi(\lambda = 1.5|x) = 0.7543.$$

□

Ex. 1.3. Assume that θ is the defective rate of a batch of products. Suppose that 3 out of 8 randomly selected products from this batch are defective. The prior is:

(1) $\theta \sim U(0, 1)$;

(2) $\theta \sim \pi(\theta) = \begin{cases} 2(1 - \theta), & 0 < \theta < 1, \\ 0, & \text{otherwise.} \end{cases}$

Find the posterior distribution respectively.

Let the random variable X be the number of defective products, then $X \sim B(8, \theta)$. As a result,

(1)

$$\begin{aligned}\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{C(8,3)\theta^3(1-\theta)^5 \cdot 1}{\int_0^1 C(8,3)\theta^3(1-\theta)^5 \cdot 1d\theta} \\ &= \frac{\Gamma(10)}{\Gamma(4)\Gamma(6)}\theta^3(1-\theta)^5 \sim \text{Beta}(4,6).\end{aligned}\tag{3}$$

(2)

$$\begin{aligned}\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{C(8,3)\theta^3(1-\theta)^5 \cdot 2(1-\theta)}{\int_0^1 C(8,3)\theta^3(1-\theta)^5 \cdot 2(1-\theta)d\theta} \\ &= \frac{\Gamma(11)}{\Gamma(4)\Gamma(7)}\theta^3(1-\theta)^6 \sim \text{Beta}(4,7).\end{aligned}\tag{4}$$

□

Ex. 1.4. Suppose that X_1, \dots, X_n are samples from then density $p(x|\theta)$, and the prior is $\pi(\theta)$. Show that the posterior can be obtained as follows:

- (1) Find $\pi(\theta|x_1) \propto p(x_1|\theta)\pi(\theta)$ given $X_1 = x_1$;
- (2) Regard $\pi(\theta|x_1)$ as the prior and find $\pi(\theta|x_1, x_2) \propto p(x_2|\theta)\pi(\theta|x_1)$ given $X_2 = x_2$;
- (3) Repeat until we find $\pi(\theta|x_1, \dots, x_{n-1}) \propto p(x_n|\theta)\pi(\theta|x_1, \dots, x_{n-1})$ given $X_n = x_n$.

Let us go through this procedure:

$$\begin{aligned}\pi(\theta|x_1) &= \frac{p(x_1|\theta)\pi(\theta)}{p(x_1)}, \\ \pi(\theta|x_1, x_2) &= \frac{p(x_2|\theta)\pi(\theta|x_1)}{p(x_2)} = \frac{p(x_2|\theta)p(x_1|\theta)\pi(\theta)}{p(x_1)p(x_2)}, \\ &\vdots \\ \pi(\theta|x_1, \dots, x_{n-1}) &= \frac{p(x_1|\theta) \dots p(x_n|\theta)\pi(\theta)}{p(x_1) \dots p(x_n)} = \frac{p(x_1, \dots, x_n|\theta)\pi(\theta)}{p(x_1, \dots, x_n)}.\end{aligned}$$

□

Ex. 1.5. Suppose that the time someone spends waiting for bus every morning follows a uniform distribution $U(0, \theta)$. The prior for θ is

$$\pi(\theta) = \begin{cases} 192/\theta^4, & \theta \geq 4, \\ 0, & \theta < 4. \end{cases}$$

Say he waited 5, 5, and 8 minutes for the bus on three days. Find the posterior of θ .

Let the random variable X be the time spending to wait for bus. Then

$$\begin{aligned}
\pi(\theta|x_1, x_2, x_3) &= \frac{f(x_1, x_2, x_3|\theta)\pi(\theta)}{\int_4^\infty f(x_1, x_2, x_3|\theta)\pi(\theta)d\theta} \\
&= \frac{1/\theta^3 \cdot 192/\theta^4 \cdot I(\theta \geq 8)}{\int_4^\infty 1/\theta^3 \cdot 192/\theta^4 \cdot I(\theta \geq 8)d\theta} \\
&= \begin{cases} \frac{1/\theta^3 \cdot 192/\theta^4}{\int_8^\infty 1/\theta^3 \cdot 192/\theta^4 d\theta}, & \theta \geq 8, \\ 0, & \theta < 8, \end{cases} \\
&= \begin{cases} 6 \cdot 8^6/\theta^7, & \theta \geq 8, \\ 0, & \theta < 8. \end{cases}
\end{aligned} \tag{5}$$

□

Ex. 1.6. Suppose that the random variable X follows a uniform distribution $U(\theta - 1/2, \theta + 1/2)$. The prior for θ is $U(10, 20)$.

(1) If an observation of X is 12, find the posterior;

(2) If we got 6 observations of X : 12.0, 11.5, 11.7, 11.1, 11.4, 11.9, find the posterior.

(1)

$$\begin{aligned}
\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{10}^{20} f(x|\theta)\pi(\theta)d\theta} \\
&= \frac{1/10 \cdot I(11.5 \leq \theta \leq 12.5)}{\int_{10}^{20} 1/10 \cdot I(11.5 \leq \theta \leq 12.5)d\theta} \\
&= \begin{cases} \frac{1/10}{\int_{11.5}^{12.5} 1/10 d\theta}, & 11.5 \leq \theta \leq 12.5, \\ 0, & \text{otherwise,} \end{cases} \\
&= \begin{cases} 1, & 11.5 \leq \theta \leq 12.5, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{6}$$

(2)

$$\begin{aligned}
\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{10}^{20} f(x|\theta)\pi(\theta)d\theta} \\
&= \frac{1/10 \cdot I(11.5 \leq \theta \leq 11.6)}{\int_{10}^{20} 1/10 \cdot I(11.5 \leq \theta \leq 11.6)d\theta} \\
&= \begin{cases} \frac{1/10}{\int_{11.5}^{11.6} 1/10 d\theta}, & 11.5 \leq \theta \leq 11.6, \\ 0, & \text{otherwise,} \end{cases} \\
&= \begin{cases} 10, & 11.5 \leq \theta \leq 11.6, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{7}$$

□

Ex. 1.7. Suppose that the density of a random variable X is

$$p(x|\theta) = \frac{2x}{\theta^2} \quad (0 < x < \theta < 1).$$

- (1) If the prior for θ is $U(0, 1)$, find the posterior;
(2) If the prior for θ is $\pi(\theta) = 3\theta^2$ ($0 < \theta < 1$), find the posterior.

(1)

$$\begin{aligned}
\pi(\theta|x) &= \frac{p(x|\theta)\pi(\theta)}{\int_0^1 p(x|\theta)\pi(\theta)d\theta} \\
&= \frac{2x/\theta^2 \cdot 1 \cdot I(x < \theta)}{\int_0^1 2x/\theta^2 \cdot 1 \cdot I(x < \theta)d\theta} \\
&= \begin{cases} \frac{2x/\theta^2}{\int_x^1 2x/\theta^2 d\theta}, & x < \theta < 1, \\ 0, & \text{otherwise,} \end{cases} \\
&= \begin{cases} x/[\theta^2(1-x)], & x < \theta < 1, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{8}$$

(2)

$$\begin{aligned}
\pi(\theta|x) &= \frac{p(x|\theta)\pi(\theta)}{\int_0^1 p(x|\theta)\pi(\theta)d\theta} \\
&= \frac{2x/\theta^2 \cdot 3\theta^2 \cdot I(x < \theta)}{\int_0^1 2x/\theta^2 \cdot 3\theta^2 \cdot I(x < \theta)d\theta} \\
&= \begin{cases} \frac{6x}{\int_x^1 6xd\theta}, & x < \theta < 1, \\ 0, & \text{otherwise,} \end{cases} \\
&= \begin{cases} 1/(1-x), & x < \theta < 1, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{9}$$

□

Ex. 1.8. Suppose that 3 out of 100 randomly selected products are found to be defective. The prior for the defective rate θ is $Be(2, 200)$. Find the posterior.

Let the random variable X to be the number of defective products, then $X \sim B(100, \theta)$. As a result,

$$\begin{aligned}
\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} \\
&= \frac{C(100, 3)\theta^3(1-\theta)^{97} \cdot \frac{\Gamma(202)}{\Gamma(1)\Gamma(199)}\theta(1-\theta)^{199}}{\int_0^1 C(100, 3)\theta^3(1-\theta)^{97} \cdot \frac{\Gamma(202)}{\Gamma(1)\Gamma(199)}\theta(1-\theta)^{199}d\theta} \\
&= \frac{\Gamma(302)}{\Gamma(5)\Gamma(297)}\theta^4(1-\theta)^{296} \sim \text{Beta}(5, 297).
\end{aligned} \tag{10}$$

□

Ex. 1.9. Transform Poisson distribution and Gamma distribution into the natural form.

(1) The Poisson distribution is of the form:

$$p(x, \theta) = \frac{e^{-\lambda} \lambda^x}{x!} = \exp(-\lambda) \exp(x \log \lambda) (x!)^{-1}.$$

Let $\phi = \log \lambda$, then $\lambda = \exp(\phi)$, and as a result the natural form is

$$p(x, \phi) = \exp(-\exp(\phi)) \exp(\phi x) (x!)^{-1}.$$

(2) The Gamma distribution is of the form:

$$f(x, \alpha, \beta) = \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha-1} e^{x/\beta} = \frac{1}{\Gamma(\alpha) \beta^\alpha} \exp \left[\frac{x}{\beta} + (\alpha - 1) \log x \right].$$

Let $\phi_1 = 1/\beta$ and $\phi_2 = \alpha - 1$, then $\beta = 1/\phi_1$ and $\alpha = \phi_2 + 1$. As a result the natural form is

$$f(x, \phi_1, \phi_2) = \frac{\phi_1^{\phi_2+1}}{\Gamma(\phi_2 + 1)} \exp [\phi_1 x + \phi_2 \log x].$$

□

Ex. 1.10. Denote the natural form of the exponential family by

$$f(x, \theta) = C(\theta) \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x).$$

Show that

$$E_\theta(T_j(x)) = -\frac{\partial \log C(\theta)}{\partial \theta_j} = -\frac{1}{C(\theta)} \frac{\partial C(\theta)}{\partial \theta_j},$$

$$\text{Cov}(T_j(x), T_s(x)) = -\frac{\partial^2 \log C(\theta)}{\partial \theta_j \partial \theta_s}.$$

(1) Since

$$1 = \int_{\mathcal{X}} C(\theta) \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x) dx,$$

if we take the first derivative *w.r.t.* θ_j , then

$$0 = \int_{\mathcal{X}} \frac{\partial C(\theta)}{\partial \theta_j} \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x) dx + \int_{\mathcal{X}} T_j(x) C(\theta) \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x) dx.$$

The second term on the R.H.S is exactly $E_\theta(T_j(x))$. As a result, we have

$$E_\theta(T_j(x)) = -\frac{\partial C(\theta)}{\partial \theta_j} \frac{1}{C(\theta)} \int_{\mathcal{X}} C(\theta) \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x) dx = -\frac{\partial C(\theta)}{\partial \theta_j} \frac{1}{C(\theta)} = -\frac{\partial \log C(\theta)}{\partial \theta_j}.$$

(2) Now we know that

$$-\frac{\partial \log C(\theta)}{\partial \theta_j} = \int_{\mathcal{X}} T_j(x) C(\theta) \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x) dx.$$

Take the first derivative *w.r.t.* θ_s , we have

$$\begin{aligned}
-\frac{\partial^2 \log C(\theta)}{\partial \theta_j \partial \theta_s} &= \int_{\mathcal{X}} T_j(x) \frac{\partial C(\theta)}{\partial \theta_s} \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x) dx + \int_{\mathcal{X}} T_j(x) T_s(x) C(\theta) \exp \left\{ \sum_{j=1}^k \theta_j T_j(x) \right\} h(x) dx \\
&= \frac{\partial C(\theta)}{\partial \theta_s} \frac{1}{C(\theta)} E_{\theta}(T_j(x)) + E_{\theta}(T_j(x) T_s(x)) \\
&= E_{\theta}(T_j(x) T_s(x)) - E_{\theta}(T_j(x)) E_{\theta}(T_s(x)) \\
&= \text{Cov}(T_j(x), T_s(x)).
\end{aligned} \tag{11}$$

□

Ex. 1.11. Let $T = T(X)$ be a sufficient statistic, and $S(X) = G(T(X))$. Assume that $S = G(T)$ is a bijective mapping. Show that S is also a sufficient statistic.

Since $T(X)$ is a sufficient statistic, by factorization theorem, we know

$$f(x, \theta) = g(T(x), \theta) \cdot h(x) = g(G^{-1} \circ S(x), \theta) \cdot h(x) = g^*(S(x), \theta) \cdot h(x).$$

Therefore, $S(X)$ is also a sufficient statistic.

□

Ex. 1.12. Suppose that $X = (X_1, \dots, X_n) \sim^{\text{i.i.d.}} P(\lambda)$. Show that $T(X) = \sum_{i=1}^n X_i$ is a sufficient statistic by:

- (1) definition;
- (2) factorization theorem.

(1) Note that $T(X) \sim P(n\lambda)$. The conditional distribution of X given $T(X)$ is

$$\begin{aligned}
\Pr(x_1 = x_1, \dots, X_n = x_n | T(X) = t; \lambda) &= \frac{\Pr(x_1 = x_1, \dots, X_n = x_n, T(X) = t; \lambda)}{\Pr(T(X) = t; \lambda)} \\
&= \frac{\Pr(x_1 = x_1, \dots, X_n = x_n; \lambda)}{\Pr(T(X) = t; \lambda)} \\
&= \left(\frac{e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \right) \left(\frac{e^{-n\lambda} \cdot (n\lambda)^t}{t!} \right)^{-1} \\
&= \frac{(\sum_{i=1}^n x_i)!}{n^{\sum_{i=1}^n x_i} \prod_{i=1}^n x_i!}.
\end{aligned} \tag{12}$$

The last equation holds due to the fact $t = \sum_{i=1}^n x_i$. Since the conditional distribution is relative constant to λ , we can conclude that $T(X)$ is a sufficient statistic.

(2) The joint distribution of X is

$$p(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.$$

Then $g(T(x), \lambda) = e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^n x_i}$ and $h(x) = 1/(\prod_{i=1}^n x_i!)$, which implies that $T(X)$ is a sufficient statistic.

□

Ex. 1.13. Suppose that $X = (X_1, \dots, X_n) \sim^{\text{i.i.d.}} \text{Geometric}(p)$. Show that $T(X) = \sum_{i=1}^n X_i$ is a sufficient statistic by:

- (1) definition;
- (2) factorization theorem.

(1) Note that $T(X) \sim NB(n, p)$. The conditional distribution of X given $T(X)$ is

$$\begin{aligned} \Pr(x_1 = x_1, \dots, X_n = x_n | T(X) = t; p) &= \frac{\Pr(x_1 = x_1, \dots, X_n = x_n, T(X) = t; p)}{\Pr(T(X) = t; p)} \\ &= \frac{\Pr(x_1 = x_1, \dots, X_n = x_n; p)}{\Pr(T(X) = t; p)} \\ &= \left[p^n (1-p)^{\sum_{i=1}^n x_i - n} \right] [C(t-1, n-1) p^n (1-p)^{t-n}]^{-1} \\ &= \frac{1}{C(t-1, n-1)}. \end{aligned} \tag{13}$$

The last equation holds due to the fact $t = \sum_{i=1}^n x_i$. Since the conditional distribution is relative constant to p , we can conclude that $T(X)$ is a sufficient statistic.

(2) The joint distribution of X is

$$p(x_1, \dots, x_n; p) = \prod_{i=1}^n p(1-p)^{x_i-1} = p^n (1-p)^{\sum_{i=1}^n x_i - n}.$$

Then $g(T(x), p) = p^n (1-p)^{\sum_{i=1}^n x_i - n}$ and $h(x) = 1$, which implies that $T(X)$ is a sufficient statistic. \square

Ex. 1.14. Suppose that $X = (X_1, \dots, X_n) \sim^{\text{i.i.d.}} U(\theta - 1/2, \theta + 1/2)$. Show that $(X_{(1)}, X_{(n)})$ is a sufficient statistic.

The joint distribution of X is

$$f(x_1, \dots, x_n; \theta) = 1 \cdot I\left(X_{(n)} - \frac{1}{2} \leq \theta \leq X_{(1)} + \frac{1}{2}\right).$$

Then $g(X_{(1)}, X_{(n)}, \theta)$ is the indicator function above, and $h(x) = 1$. Therefore, $(X_{(1)}, X_{(n)})$ is a sufficient statistic. \square

Ex. 1.15. Suppose that $X_1, \dots, X_m \sim^{\text{i.i.d.}} N(a, \sigma^2)$, $Y_1, \dots, Y_n \sim^{\text{i.i.d.}} N(b, \sigma^2)$, and X_i 's and Y_j 's are independent as well. Let

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j, \quad S^2 = \frac{1}{n+m-2} \left[\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2 \right].$$

Show that (\bar{X}, \bar{Y}, S^2) is a complete sufficient statistic.

The joint distribution of X_i 's and Y_j 's is

$$\begin{aligned} &f(x_1, \dots, x_m, y_1, \dots, y_n; a, b, \sigma) \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^{m+n}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^m (X_i - a)^2 + \sum_{j=1}^n (Y_j - b)^2 \right] \right\} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^{m+n}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^m X_i^2 - 2am\bar{X} + ma^2 + \sum_{j=1}^n Y_j^2 - 2bn\bar{Y} + nb^2 \right] \right\} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^{m+n}} \exp \left\{ -\frac{ma^2 + nb^2}{2\sigma^2} \right\} \exp \left\{ -\frac{1}{2\sigma^2} \left(\sum_{i=1}^m X_i^2 + \sum_{j=1}^n Y_j^2 \right) + \frac{am}{\sigma^2} \bar{X} + \frac{bn}{\sigma^2} \bar{Y} \right\}. \end{aligned} \tag{14}$$

As a result, the sampling distribution is a member of exponential family, and $(\sum_{i=1}^m X_i^2 + \sum_{j=1}^n Y_j^2, \bar{X}, \bar{Y})$ is a sufficient statistic. Now let $\phi_1 = -1/(2\sigma^2)$, $\phi_2 = am/\sigma^2$ and $\phi_3 = bn/\sigma^2$. The natural parameter space is

$$\Theta^* = \{(\phi_1, \phi_2, \phi_3) : -\infty < \phi_1 < 0, -\infty < \phi_2, \phi_3 < +\infty\},$$

so there exists interior points in it. We can conclude that $(\sum_{i=1}^m X_i^2 + \sum_{j=1}^n Y_j^2, \bar{X}, \bar{Y})$ is a complete sufficient statistic, and furthermore (\bar{X}, \bar{Y}, S^2) is also a complete sufficient statistic. \square

Ex. 1.16. Suppose that

$$X_1, \dots, X_n \sim \text{i.i.d.} \quad f(x, \theta) = \frac{1}{2\theta} e^{-|x|/\theta} \quad (-\infty < x < +\infty, \theta > 0).$$

Show that $T = \sum_{i=1}^n |X_i|$ is a complete sufficient statistic of θ .

The joint distribution of (X_1, \dots, X_n) is

$$f(x_1, \dots, x_n; \theta) = \frac{1}{(2\theta)^n} \exp \left\{ -\frac{1}{\theta} \sum_{i=1}^n |X_i| \right\}.$$

As a result, the sampling distribution is a member of exponential family, and T is a sufficient statistic. Let $\phi = -1/\theta$, then the natural parameter space is

$$\Theta^* = \{\phi : -\infty < \theta < 0\},$$

so there exists interior points in it. We can conclude that $T = \sum_{i=1}^n |X_i|$ is a complete sufficient statistic of θ . \square

Ex. 1.17. Suppose that $X_1, \dots, X_n \sim \text{i.i.d.} B(1, p)$, where $p \in (0, 1)$ is unknown. Assume that $s \in (0, n)$ is an integer. Find

- (1) The UMVUE of p^s ;
- (2) The UMVUE of $p^s + (1 - p)^{n-s}$.

(1) It is acknowledged that $T(X) = \sum_{i=1}^n X_i \sim B(n, p)$ is a complete sufficient statistic. Now we want an unbiased estimator $\delta(T)$ of p^s :

$$\sum_{t=0}^n C(n, t) \delta(t) p^t (1 - p)^{n-t} = p^s.$$

Solve the equation and we have

$$\delta(T) = \frac{T(T-1) \cdots (T-s+1)}{n(n-1) \cdots (n-s+1)}.$$

Since the unbiased estimator $\delta(T)$ is a function of T , we can conclude that it is the UMVUE of p^s .

(2) Similarly, we want an unbiased estimator $\delta^*(T)$ of $p^s + (1 - p)^{n-s}$. First we need to solve for the unbiased estimator $\delta'(t)$ of $(1 - p)^{n-s}$:

$$\sum_{t=0}^n C(n, t) \delta'(t) p^t (1 - p)^{n-t} = (1 - p)^{n-s}.$$

The equation gives

$$\delta'(t) = \frac{(n-T)(n-T-1) \cdots (s-T+1)}{n(n-1) \cdots (s+1)}.$$

As a result, the unbiased estimator $\delta^*(T) = \delta(T) + \delta'(T)$ is a function of T . We conclude that it is the UMVUE of $p^s + (1-p)^{n-s}$. \square

Ex. 1.18. Suppose that $X_1, \dots, X_m \sim^{\text{i.i.d.}} N(a, \sigma^2)$, $Y_1, \dots, Y_n \sim^{\text{i.i.d.}} N(a, 2\sigma^2)$, and X_i 's and Y_j 's are independent as well. Find the UMVUEs of a and σ^2 .

The joint distribution of X_i 's and Y_j 's is

$$\begin{aligned} & f(x_1, \dots, x_m, y_1, \dots, y_n; a, \sigma) \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^{m+n}(\sqrt{2})^n} \exp \left\{ -\frac{1}{4\sigma^2} \left[\sum_{i=1}^m 2(X_i - a)^2 + \sum_{j=1}^n (Y_j - a)^2 \right] \right\} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^{m+n}(\sqrt{2})^n} \exp \left\{ -\frac{1}{4\sigma^2} \left[\sum_{i=1}^m 2X_i^2 - 4a \sum_{i=1}^m X_i + 2ma^2 + \sum_{j=1}^n Y_j^2 - 2a \sum_{j=1}^n Y_j + na^2 \right] \right\} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^{m+n}(\sqrt{2})^n} \exp \left\{ -\frac{(2m+n)a^2}{4\sigma^2} \right\} \exp \left\{ -\frac{1}{4\sigma^2} \left(2 \sum_{i=1}^m X_i^2 + \sum_{j=1}^n Y_j^2 \right) + \frac{a}{2\sigma^2} \left(2 \sum_{i=1}^m X_i + \sum_{j=1}^n Y_j \right) \right\}. \end{aligned} \quad (15)$$

As a result, the sufficient statistic is $(2 \sum_{i=1}^m X_i^2 + \sum_{j=1}^n Y_j^2, 2 \sum_{i=1}^m X_i + \sum_{j=1}^n Y_j)$. Now let $\phi_1 = -1/(4\sigma^2)$, and $\phi_2 = a/2\sigma^2$. The natural parameter space is

$$\Theta^* = \{(\phi_1, \phi_2) : -\infty < \phi_1 < 0, -\infty < \phi_2 < +\infty\},$$

so there exists interior points in it. We can conclude that $(2 \sum_{i=1}^m X_i^2 + \sum_{j=1}^n Y_j^2, 2 \sum_{i=1}^m X_i + \sum_{j=1}^n Y_j)$ is a complete sufficient statistic.

Now consider

$$\begin{aligned} \hat{a} &= \frac{2 \sum_{i=1}^m X_i + \sum_{j=1}^n Y_j}{2m+n}, \\ \hat{\sigma}^2 &= \frac{1}{2m+2n-2} \left[2 \sum_{i=1}^m X_i^2 + \sum_{j=1}^n Y_j^2 - \frac{(2 \sum_{i=1}^m X_i + \sum_{j=1}^n Y_j)^2}{2m+n} \right]. \end{aligned}$$

By simple algebra we know $E(\hat{a}) = a$ and $E(\hat{\sigma}^2) = \sigma^2$. Since the unbiased estimators are functions of complete sufficient statistic, we can conclude that \hat{a} and $\hat{\sigma}^2$ the UMVUEs of their expectations respectively. \square

Ex. 1.19. Suppose that $X_1, \dots, X_n \sim^{\text{i.i.d.}} B(k, \theta)$. Find the UMVUE of $\theta(1-\theta)$ by Lehmann-Scheffe, find the variance of the UMVUE, and compare it with the Cramer-Rao bound.

(1) It is acknowledged that $T(X) = \sum_{i=1}^n X_i \sim B(nk, \theta)$ is a complete sufficient statistic. Now we want an unbiased estimator $\delta(T)$ of $\theta(1-\theta)$:

$$\sum_{t=0}^{nk} C(nk, t) \delta(t) \theta^t (1-\theta)^{nk-t} = \theta(1-\theta).$$

Solve the equation and we have

$$\delta(T) = \frac{T(nk - T)}{nk(nk - 1)}.$$

Since it is a function of T , we can conclude that it is the UMVUE of $\theta(1 - \theta)$.

(2) Since $T \sim B(nk, \theta)$, we know:

- $E(T) = nk\theta$;
- $E(T^2) = \text{Var}(T) + E(T)^2 = nk\theta(1 - \theta) + (nk\theta)^2$;
- If we want $E(T^3)$, consider

$$\sum_{t=0}^{nk} t^2 C(nk, t) \theta^t (1 - \theta)^{nk-t} = nk\theta(1 - \theta) + (nk\theta)^2.$$

Take derivative w.r.t. θ , and by simple algebra, we have

$$E(T^3) = \theta(1 - \theta)(1 - 2\theta)(nk) + 3\theta^2(1 - \theta)(nk)^2 + \theta^3(nk)^3.$$

- If we want $E(T^4)$, similarly, consider

$$\sum_{t=0}^{nk} t^3 C(nk, t) \theta^t (1 - \theta)^{nk-t} = \theta(1 - \theta)(1 - 2\theta)(nk) + 3\theta^2(1 - \theta)(nk)^2 + \theta^3(nk)^3.$$

Take derivative w.r.t. θ , and by simple algebra, we have

$$\begin{aligned} E(T^4) &= \theta(nk)E(T^3) \\ &+ [\theta(1 - \theta)^2(1 - 2\theta) - \theta^2(1 - \theta)(1 - 2\theta) - 2\theta^2(1 - \theta)^2] (nk) \\ &+ [6\theta^2(1 - \theta)^2 - 3\theta^3(1 - \theta)] (nk)^2 \\ &+ 3\theta^3(1 - \theta)(nk)^3. \end{aligned} \tag{16}$$

As a result,

$$\begin{aligned} \text{Var}[T(nk - T)] &= \text{Var}[nkT - T^2] \\ &= (nk)^2 \text{Var}(T) + \text{Var}(T^2) - 2(nk) \text{Cov}(T, T^2) \\ &= (nk)^2 \text{Var}(T) + E(T^4) - E(T^2)^2 - 2(nk)(E(T^3) - E(T)E(T^2)). \end{aligned} \tag{17}$$

By some cumbersome calculations,

$$\begin{aligned} \text{Var}[T(nk - T)] &= (-6\theta^4 + 12\theta^2 - 7\theta^2 + \theta)(nk) \\ &+ (10\theta^4 - 2\theta^3 + 10\theta^2 - 2\theta)(nk)^2 \\ &+ (-4\theta^4 + 10\theta^3 - 7\theta^2 + \theta)(nk)^3. \end{aligned} \tag{18}$$

Furthermore,

$$\text{Var}[\delta(T)] = \frac{\text{Var}[T(nk - T)]}{(nk)^2(nk - 1)^2}.$$

(3) The density function of X_i is

$$p(x; \theta) = C(k, x)\theta^x(1 - \theta)^{k-x}.$$

Therefore, the information number is

$$\begin{aligned} I(\theta) &= -E \left[\frac{\partial^2 \log p(X; \theta)}{\partial \theta^2} \right] \\ &= -E \left[\frac{\partial^2}{\partial \theta^2} (X \log \theta + (k - X) \log(1 - \theta)) \right] \\ &= -E \left[\frac{\partial}{\partial \theta} \left(\frac{X}{\theta} - \frac{k - X}{1 - \theta} \right) \right] \\ &= \frac{k(1 + 2\theta)}{\theta(\theta + 1)}. \end{aligned} \tag{19}$$

The C-R bound is

$$\frac{[(\theta(1 - \theta))'']^2}{nI(\theta)} = \frac{(1 - 2\theta)^2\theta(1 + \theta)}{nk(1 + 2\theta)}.$$

Ex. 1.20. Suppose that

$$X_1, \dots, X_n \sim \text{i.i.d.} \quad f(x, \theta) = \begin{cases} \theta^{-1}e^{-x/\theta}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Assume that θ is an unknown parameter. Find the UMVUE by C-R inequality.

From the density function we know that $X_i \sim \text{Exp}(\theta)$, $E(X_i) = \theta$, and $\text{Var}(X_i) = \theta^2$. As a result,

$$\hat{\theta} = \frac{\sum_{i=1}^n X_i}{n}$$

is an unbiased estimator of θ . The variance of it is θ^2/n .

On the other hand, the information number is

$$\begin{aligned} I(\theta) &= -E \left[\frac{\partial^2 \log f(X; \theta)}{\partial \theta^2} \right] \\ &= -E \left[\frac{\partial^2}{\partial \theta^2} \left(-\log \theta - \frac{X}{\theta} \right) \right] \\ &= -E \left[\frac{\partial}{\partial \theta} \left(-\frac{1}{\theta} + \frac{X}{\theta^2} \right) \right] \\ &= \frac{1}{\theta^2}. \end{aligned} \tag{20}$$

Thus the C-R bound is

$$\frac{1}{nI(\theta)} = \frac{\theta^2}{n}.$$

Since the lower bound is attained by $\hat{\theta}$, we can conclude that the estimator (sample mean) is the UMVUE.

Ex. 1.21. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} \Gamma(\alpha, \lambda),$$

where α is known, and $\lambda > 0$. Find the UMVUE of $g(\lambda) = 1/\lambda$ by C-R bound.

We can construct an estimator

$$\hat{g}(\lambda) = \frac{1}{n} \sum_{i=1}^n \frac{X_i}{\alpha}.$$

Then $E[\hat{g}(\lambda)] = E[X_i]/\alpha = 1/\lambda$, and $\text{Var}[\hat{g}(\lambda)] = \text{Var}[X_i]/(n\alpha^2) = 1/(n\alpha\lambda^2)$. The estimator is indeed an unbiased estimator.

On the other hand, since the density function of X_i is

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},$$

the information number is

$$\begin{aligned} I(\lambda) &= -E \left[\frac{\partial^2 \log f(X; \lambda)}{\partial \lambda^2} \right] \\ &= -E \left[\frac{\partial^2}{\partial \lambda^2} (\alpha \log \lambda + (\alpha - 1) \log X - \lambda X) \right] \\ &= -E \left[\frac{\partial}{\partial \lambda} \left(\frac{\alpha}{\lambda} - X \right) \right] \\ &= \frac{\alpha}{\lambda^2}. \end{aligned} \tag{21}$$

As a result, the C-R bound is

$$\frac{1/\lambda^4}{nI(\lambda)} = \frac{1}{n\alpha\lambda^2}.$$

The lower bound is attained by $\hat{g}(\lambda)$, so we conclude that it is the UMVUE.

Ex. 1.22. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} N(0, \sigma^2).$$

Find the level α UMPT for

$$H_0 : \sigma^2 \leq \sigma_0^2 \longleftrightarrow H_1 : \sigma^2 > \sigma_0^2.$$

The joint density is

$$f(x_1, \dots, x_n; \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 \right\}.$$

Therefore,

$$Q(\sigma^2) = -\frac{1}{2\sigma^2}, \quad T(x) = \sum_{i=1}^n x_i^2.$$

Since $Q(\sigma^2)$ is a strictly increasing function, we can construct the test function

$$\phi(x) = \begin{cases} 1, & T(x) > c, \\ 0, & T(x) \leq c. \end{cases}$$

Let

$$E[\phi(X)|\sigma = \sigma_0] = \Pr(T(X) > c|\sigma = \sigma_0) = \Pr\left(\frac{\sum_{i=1}^n X_i^2}{\sigma_0^2} > \frac{c}{\sigma_0^2}\right) = \alpha,$$

and as a result, we have $\chi_\alpha^2(n) = c/\sigma_0^2$, indicating that $c = \sigma_0^2 \chi_\alpha^2(n)$.

In conclusion, the level α UMPT is

$$\phi(x) = \begin{cases} 1, & \sum_{i=1}^n x_i^2 > \sigma_0^2 \chi_\alpha^2(n), \\ 0, & \sum_{i=1}^n x_i^2 \leq \sigma_0^2 \chi_\alpha^2(n). \end{cases}$$

Ex. 1.23. Given an integer k . In order to test for the probability of an even ($p \leq p_0$ or not), we independently repeat our experiment until the event occurs k times. Let the random variable X be the total number of experiments when we stop.

(1) Show that the probability mass function of X is

$$\Pr(X = x) = C(x-1, k-1)p^k(1-p)^{x-k}$$

where $x = k, k+1, \dots$

(2) Given p_0 and α , find the level α UMPT for

$$H_0 : p \leq p_0 \longleftrightarrow H_1 : p > p_0.$$

(1) Suppose that our realization of X is x . Among the x experiments, k are considered as successes (if we call the occurrence of the event a success), and $x - k$ are considered as failures. The last experiment must be a success, so we need to choose another $k - 1$ successes among the first $x - 1$ experiments. Similar to the binomial experiment, we justify that the p.m.f. of X is

$$\Pr(X = x) = C(x-1, k-1)p^k(1-p)^{x-k},$$

and that the total number of experiments must be at least the number of successes. [In fact, the distribution is called the *negative binomial distribution*, and $X \sim NB(k, p)$]

(2) The p.m.f. of X is

$$\begin{aligned} p(x; p) &= C(x-1, k-1) \left(\frac{p}{1-p}\right)^k (1-p)^x \\ &= C(x-1, k-1) \left(\frac{p}{1-p}\right)^k \exp[\log(1-p)x]. \end{aligned} \tag{22}$$

Since $Q(p) = \log(1-p)$ is a strictly decreasing function, we can construct the test function

$$\phi(x) = \begin{cases} 1, & x < c, \\ r, & x = c, \\ 0, & x \geq c. \end{cases}$$

We can determine c by

$$\alpha_1 = \sum_{x=k}^{c-1} C(x-1, k-1)p_0^k(1-p_0)^{x-k} \leq \alpha \leq \sum_{x=k}^c C(x-1, k-1)p_0^k(1-p_0)^{x-k}.$$

Let

$$E[\phi(X)|p = p_0] = \Pr(X < c) + r \cdot \Pr(X = c) = \alpha,$$

then we can determine

$$r = \frac{\alpha - \alpha_1}{C(c-1, k-1)p_0^k(1-p_0)^{c-k}}.$$

Ex. 1.24. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} \text{Exp}(\lambda),$$

where $\lambda > 0$ is unknown. Given λ_0 and α , find the level α UMPT for

$$H_0 : \lambda \geq \lambda_0 \longleftrightarrow H_1 : \lambda < \lambda_0.$$

The joint density is

$$f(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\}.$$

Since $Q(\lambda) = -\lambda$ is a strictly decreasing function, and $T(x) = \sum_{i=1}^n x_i$, we can construct the test function

$$\phi(x) = \begin{cases} 1, & T(x) > c, \\ 0, & T(x) \leq c. \end{cases}$$

By $T(X) \sim \Gamma(n, \lambda)$ and

$$E[\phi(X)|\lambda = \lambda_0] = \Pr(T(X) > c|\lambda = \lambda_0) = \alpha,$$

we know that $c = \Gamma_\alpha(n, \lambda)$. Thus the level α UMPT is

$$\phi(x) = \begin{cases} 1, & \sum_{i=1}^n x_i > \Gamma_\alpha(n, \lambda), \\ 0, & \sum_{i=1}^n x_i \leq \Gamma_\alpha(n, \lambda). \end{cases}$$

2 The Prior Distributions

Ex. 2.1. *Not working on it.*

Ex. 2.2. *Not working on it.*

Ex. 2.3. *Suppose that the prior distribution for parameter θ is $Be(\alpha, \beta)$. The expectation and variance derived from the prior information are $1/3$ and $1/45$, respectively. Please find the prior distribution.*

If $X \sim Be(\alpha, \beta)$, then

$$E(X) = \frac{\alpha}{\alpha + \beta}, \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Let $E(X) = 1/3$ and $\text{Var}(X) = 1/45$, we have $\alpha = 3$ and $\beta = 6$. Thus, the prior distribution is $Be(3, 6)$. \square

Ex. 2.4. *Not working on it.*

Ex. 2.5. *Not working on it.*

Ex. 2.6. *The failure time of an electronic component follows the exponential distribution:*

$$f(x|\theta) = \theta^{-1}e^{-x/\theta} \quad (x > 0).$$

The prior distribution of the unknown parameter θ follows $\Gamma^{-1}(1, 100)$. Find the marginal probability of the component failing before time 200.

Consider:

$$\begin{aligned} \Pr(X < 200) &= \int_0^{200} f(x)dx \\ &= \int_0^{200} \int_0^{+\infty} f(x|\theta)f(\theta)d\theta dx \\ &= \int_0^{200} \int_0^{+\infty} \theta^{-1}e^{-x/\theta} \frac{100}{\Gamma(1)} \theta^{-2}e^{-100/\theta} d\theta dx \\ &= 100 \int_0^{200} \int_0^{+\infty} \theta^{-3}e^{-\frac{x+100}{\theta}} d\theta dx \\ &= 100 \int_0^{200} \frac{\Gamma(2)}{(x+100)^2} dx \\ &= \frac{2}{3}. \end{aligned} \tag{23}$$

\square

Ex. 2.7. *Suppose that*

$$X_1, \dots, X_n \sim \text{i.i.d. } P(\theta_i) \quad i = 1, \dots, n.$$

If $\theta_1, \dots, \theta_n$ are samples from the Gamma distribution $\Gamma(r, \lambda)$, find the joint marginal distribution $m(\mathbf{x})$ for $\mathbf{X} = (X_1, \dots, X_n)$.

Consider:

$$\begin{aligned}
m(\mathbf{x}) &= \prod_{i=1}^n f(x_i) \\
&= \prod_{i=1}^n \int_0^{+\infty} f(x_i|\theta_i) f(\theta_i|r, \lambda) d\theta_i \\
&= \prod_{i=1}^n \int_0^{+\infty} \frac{e^{-\theta_i} \theta_i^{x_i}}{x_i!} \frac{\lambda^r}{\Gamma(r)} \theta_i^{r-1} e^{-\lambda \theta_i} d\theta_i \\
&= \prod_{i=1}^n \frac{\lambda^r}{x_i! \Gamma(r)} \int_0^{+\infty} \theta_i^{x_i+r-1} e^{-(\lambda+1)\theta_i} d\theta_i \\
&= \prod_{i=1}^n \frac{\lambda^r}{x_i! \Gamma(r)} \frac{\Gamma(x_i+r)}{(\lambda+1)^{x_i+r}}.
\end{aligned} \tag{24}$$

□

Ex. 2.8. Cont'd Ex. 2.7. Let $n = 3$, $x_1 = 3$, $x_2 = 0$, $x_3 = 5$. Find the ML-II prior.

Let $\ell = \log m(\mathbf{x})$, then

$$\begin{aligned}
\ell &= \log \prod_{i=1}^n \frac{\lambda^r}{x_i! \Gamma(r)} \frac{\Gamma(x_i+r)}{(\lambda+1)^{x_i+r}} \\
&= \sum_{i=1}^n [r \log \lambda - \log x_i! - \log \Gamma(r) + \log \Gamma(x_i+r) - (x_i+r) \log(\lambda+1)] \\
&= \sum_{i=1}^n \left[r \log \lambda - \log x_i! + \sum_{j=0}^{x_i-1} \log(r+j) - (x_i+r) \log(\lambda+1) \right].
\end{aligned} \tag{25}$$

To solve for λ , let

$$\frac{\partial \ell}{\partial \lambda} = \frac{nr}{\lambda} - \sum_{i=1}^n \frac{x_i+r}{\lambda+1} = 0,$$

and we have $r = \lambda \bar{x}$. To solve for r , let

$$\frac{\partial \ell}{\partial r} = n \log \lambda + \sum_{i=1}^n \sum_{j=0}^{x_i-1} \frac{1}{r+j} - n \log(\lambda+1) = 0,$$

then we have

$$n \log \frac{\lambda+1}{\lambda} = \sum_{i=1}^n \sum_{j=0}^{x_i-1} \frac{1}{\lambda \bar{x} + j},$$

which is

$$3 \log \frac{\lambda+1}{\lambda} = \frac{2}{\lambda \bar{x} + 0} + \frac{2}{\lambda \bar{x} + 1} + \frac{2}{\lambda \bar{x} + 2} + \frac{1}{\lambda \bar{x} + 3} + \frac{1}{\lambda \bar{x} + 4},$$

where $\bar{x} = 8/3$. By numerical methods, we know $\lambda = 0.7018048$, and $r = 1.87147957$. Therefore, The ML-II prior is $\Gamma(1.87147957, 0.7018048)$. □

Ex. 2.9. Cont'd Ex. 2.7. Show that

$$\hat{r} = \frac{\bar{x}^2}{S^2 - \bar{x}}, \quad \hat{\lambda} = \frac{\bar{x}}{S^2 - \bar{x}}$$

using the moment method.

Since $\mu(\theta) = \theta$ and $\sigma^2(\theta) = \theta$, we know

$$\begin{aligned} E^{\theta|r,\lambda}[\mu(\theta)] &= E^{\theta|r,\lambda}[\theta] = \frac{r}{\lambda}, \\ E^{\theta|r,\lambda}[\sigma^2(\theta)] &= E^{\theta|r,\lambda}[\theta] = \frac{r}{\lambda}, \\ E^{\theta|r,\lambda} \left\{ [\mu(\theta) - \mu_m(\lambda)]^2 \right\} &= E^{\theta|r,\lambda} \left\{ \left[\theta - \frac{r}{\lambda} \right]^2 \right\} = \frac{r}{\lambda^2}. \end{aligned}$$

As a result,

$$\begin{aligned} \bar{x} &= E^{\theta|r,\lambda}[\mu(\theta)] = \frac{r}{\lambda}, \\ S^2 &= E^{\theta|r,\lambda}[\sigma^2(\theta)] + E^{\theta|r,\lambda} \left\{ [\mu(\theta) - \mu_m(\lambda)]^2 \right\} = \frac{r}{\lambda} + \frac{r}{\lambda^2}, \end{aligned}$$

which implies that

$$\hat{r} = \frac{\bar{x}^2}{S^2 - \bar{x}}, \quad \hat{\lambda} = \frac{\bar{x}}{S^2 - \bar{x}}.$$

□

Ex. 2.10. Suppose that X follows the exponential distribution $\text{Exp}(\theta)$, and the prior distribution for θ is $\Gamma(\alpha, \lambda)$. Let

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} m(x|\alpha, \lambda),$$

the sample mean be $\bar{X} = 2$, and the sample variance be $S^2 = 8$. Find the prior distribution using the moment method.

Here we assume that

$$f(X|\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}.$$

Since $\mu(\theta) = \theta$ and $\sigma^2(\theta) = \theta^2$, we know

$$\begin{aligned} E^{\theta|\alpha,\lambda}[\mu(\theta)] &= E^{\theta|\alpha,\lambda}[\theta] = \alpha\lambda, \\ E^{\theta|\alpha,\lambda}[\sigma^2(\theta)] &= E^{\theta|\alpha,\lambda}[\theta^2] = \alpha\lambda^2 + \alpha^2\lambda^2, \\ E^{\theta|\alpha,\lambda} \left\{ [\mu(\theta) - \mu_m(\lambda)]^2 \right\} &= E^{\theta|\alpha,\lambda} \left\{ [\theta - \alpha\lambda]^2 \right\} = \alpha\lambda^2. \end{aligned}$$

As a result,

$$\begin{aligned} \bar{x} &= E^{\theta|\alpha,\lambda}[\mu(\theta)] = \alpha\lambda = 2, \\ S^2 &= E^{\theta|\alpha,\lambda}[\sigma^2(\theta)] + E^{\theta|\alpha,\lambda} \left\{ [\mu(\theta) - \mu_m(\lambda)]^2 \right\} = 2\alpha\lambda^2 + \alpha^2\lambda^2 = 8, \end{aligned}$$

which implies that $\alpha = 2$, $\lambda = 1$. In other words,

$$f(\theta) = \theta e^{-\theta}.$$

□

Ex. 2.11. Determine whether the following distribution families are location parameter families, scale parameter families or neither. Find a non-informative prior for each of them.

- (1) The uniform distribution $U(\theta - 1, \theta + 1)$;
- (2) The Cauchy distribution $C(0, \beta)$;
- (3) The T distribution $T(n, \mu, \sigma^2)$, where n is fixed;
- (4) The Pareto distribution $Pa(x_0, \alpha)$ where α is fixed.

(1) If $X \sim U(\theta - 1, \theta + 1)$, then

$$f(x) = \frac{1}{2} \cdot I(\theta - 1 < x < \theta + 1) = \frac{1}{2} \cdot I(-1 < x - \theta < 1) = g(x - \theta),$$

where $g(t) = 1/2 \cdot I(-1 < t < 1)$. So the uniform distribution family is a location parameter family, and the uninformative prior is $\pi(\theta) \equiv 1$.

(2) If $X \sim C(0, \beta)$, then

$$f(x) = \frac{1}{\pi} \cdot \frac{\beta}{\beta^2 + x^2} = \frac{1}{\pi} \cdot \frac{1}{\beta(1 + (x/\beta)^2)} = \beta^{-1} \phi\left(\frac{x}{\beta}\right),$$

where $\phi(t) = (1/\pi) \cdot (1 + t^2)^{-1}$, and $\beta > 0$. So the Cauchy distribution family is a scale parameter family, and the uninformative prior is $\pi(\beta) = 1/\beta$.

(3) If $X \sim T(n, \mu, \sigma^2)$, then

$$f(x) = \sigma^{-1} g\left(\frac{x - \mu}{\sigma}\right),$$

where

$$g(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

The T distribution family is a location-scale parameter distribution family.

If we assume that the non-informative priors of μ and of σ are independent, then the joint non-informative prior is $\pi(\mu, \sigma) = 1/\sigma$.

(4) If $X \sim Pa(x_0, \alpha)$, then

$$f(x) = \begin{cases} \frac{\alpha x_0^\alpha}{x^{\alpha+1}}, & x \geq x_0, \\ 0, & x < x_0. \end{cases}$$

It is clear that the Pareto distribution family (w.r.t. x_0) is not a location parameter family, since it is not the form of $g(x - x_0)$. If we consider the Jeffreys prior, the problem arise since the probability density function of Pareto distribution does not satisfy the second regularity condition; more specifically, the support set depends on parameter x_0 . A good paper Li, Sun and Peng (2019) gives the conclusion: $\pi(x_0) = 1/x_0$. \square

Ex. 2.12. Find the Jeffreys priors for the following distributions:

- (1) Poisson distribution $P(\lambda)$;
- (2) Binomial distribution $B(n, \theta)$ (n is known);
- (3) Negative binomial distribution $Nb(r, \theta)$ (r is known);
- (4) Exponential distribution $Exp(1/\lambda)$;
- (5) Gamma distribution $\Gamma(\alpha, \lambda)$ (α is known);
- (6) Multinomial distribution $M(n, \mathbf{p})$, $\mathbf{p} = (p_1, \dots, p_k)$ (n is known).

(1) Poisson distribution $P(\lambda)$.

Step 1: the log-likelihood:

$$\ell = \sum_{i=1}^n \log f(x_i|\lambda) = \sum_{i=1}^n \log \frac{e^\lambda \lambda^{x_i}}{x_i!} = \sum_{i=1}^n [\lambda + x_i \log \lambda - \log x_i!].$$

Step 2: the Fisher information matrix:

$$I(\lambda) = E_{X|\lambda} \left[-\frac{\partial^2 \ell}{\partial \lambda^2} \right] = E_{X|\lambda} \left[\sum_{i=1}^n \frac{X_i}{\lambda^2} \right] = \frac{n}{\lambda}.$$

Step 3: the non-informative prior:

$$\pi(\lambda) = I(\lambda)^{1/2} \propto \frac{1}{\sqrt{\lambda}}.$$

(2) Binomial distribution $B(n, \theta)$ (n is known).

Step 1: the log-likelihood:

$$\begin{aligned} \ell &= \sum_{i=1}^N \log f(x_i|\theta) \\ &= \sum_{i=1}^N \log C(n, x_i) \theta^{x_i} (1 - \theta)^{n-x_i} \\ &= \sum_{i=1}^N [\log C(n, x_i) + x_i \log \theta + (n - x_i) \log(1 - \theta)]. \end{aligned} \tag{26}$$

Step 2: the Fisher information matrix:

$$I(\theta) = E_{X|\theta} \left[-\frac{\partial^2 \ell}{\partial \theta^2} \right] = E_{X|\theta} \sum_{i=1}^N \left[\frac{X_i}{\theta^2} + \frac{n - X_i}{(1 - \theta)^2} \right] = \frac{nN}{\theta(1 - \theta)}.$$

Step 3: the non-informative prior:

$$\pi(\theta) = I(\theta)^{1/2} \propto \frac{1}{\theta^{1/2}(1 - \theta)^{1/2}}.$$

(3) Negative binomial distribution $Nb(r, \theta)$ (r is known).

Step 1: the log-likelihood:

$$\begin{aligned} \ell &= \sum_{i=1}^n \log f(x_i|\theta) \\ &= \sum_{i=1}^n \log C(x_i + r - 1, x_i) \theta^r (1 - \theta)^{x_i} \\ &= \sum_{i=1}^n [\log C(x_i + r - 1, x_i) + r \log \theta + x_i \log(1 - \theta)]. \end{aligned} \tag{27}$$

Step 2: the Fisher information matrix:

$$I(\theta) = E_{X|\theta} \left[-\frac{\partial^2 \ell}{\partial \theta^2} \right] = E_{X|\theta} \sum_{i=1}^n \left[\frac{r}{\theta^2} + \frac{X_i}{(1-\theta)^2} \right] = \frac{nr}{\theta^2(1-\theta)}.$$

Step 3: the non-informative prior:

$$\pi(\theta) = I(\theta)^{1/2} \propto \frac{1}{\theta(1-\theta)^{1/2}}.$$

(4) Exponential distribution $Exp(1/\lambda)$.

Step 1: the log-likelihood:

$$\ell = \sum_{i=1}^n \log f(x_i|\lambda) = \sum_{i=1}^n \log \frac{1}{\lambda} e^{-x_i/\lambda} = \sum_{i=1}^n \left[-\log \lambda - \frac{x_i}{\lambda} \right].$$

Step 2: the Fisher information matrix:

$$I(\lambda) = E_{X|\lambda} \left[-\frac{\partial^2 \ell}{\partial \lambda^2} \right] = E_{X|\lambda} \left[\sum_{i=1}^n \frac{X_i}{\lambda^2} \right] = \frac{n}{\lambda^2}.$$

Step 3: the non-informative prior:

$$\pi(\lambda) = I(\lambda)^{1/2} \propto \frac{1}{\lambda}.$$

(5) Gamma distribution $\Gamma(\alpha, \lambda)$ (α is known).

Step 1: the log-likelihood:

$$\begin{aligned} \ell &= \sum_{i=1}^n \log f(x_i|\lambda) \\ &= \sum_{i=1}^n \log \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i} \\ &= \sum_{i=1}^n [\alpha \log \lambda - \log \Gamma(\alpha) + (\alpha-1) \log x_i - \lambda x_i]. \end{aligned} \tag{28}$$

Step 2: the Fisher information matrix:

$$I(\lambda) = E_{X|\lambda} \left[-\frac{\partial^2 \ell}{\partial \lambda^2} \right] = E_{X|\lambda} \left[\frac{n\alpha}{\lambda^2} \right] = \frac{n\alpha}{\lambda^2}.$$

Step 3: the non-informative prior:

$$\pi(\lambda) = I(\lambda)^{1/2} \propto \frac{1}{\lambda}.$$

(6) Multinomial distribution $M(n, \mathbf{p})$, $\mathbf{p} = (p_1, \dots, p_k)$ (n is known).

Step 1: the log-likelihood:

$$\begin{aligned}
\ell &= \sum_{i=1}^N \log f(\mathbf{x}_i | \mathbf{p}) \\
&= \sum_{i=1}^N \log \left(\prod_{j=1}^k p_j^{x_{ij}} \right) \left(1 - \sum_{j=1}^k p_j \right)^{n - \sum_{j=1}^k x_{ij}} \\
&= \sum_{i=1}^N \left[\sum_{j=1}^k x_{ij} \log p_j + \left(n - \sum_{j=1}^k x_{ij} \right) \log \left(1 - \sum_{j=1}^k p_j \right) \right]
\end{aligned} \tag{29}$$

Step 2: the Fisher information matrix:

$$\begin{aligned}
I_{ab}(\mathbf{p}) &= E_{\mathbf{X}|\mathbf{p}} \left[-\frac{\partial^2 \ell}{\partial p_a \partial p_b} \right] \\
&= E_{\mathbf{X}|\mathbf{p}} \sum_{i=1}^N \frac{n - \sum_{j=1}^k x_{ij}}{(1 - \sum_{j=1}^k p_j)^2} \\
&= \frac{nN}{1 - \sum_{j=1}^k p_j};
\end{aligned} \tag{30}$$

$$\begin{aligned}
I_{aa}(\mathbf{p}) &= E_{\mathbf{X}|\mathbf{p}} \left[-\frac{\partial^2 \ell}{\partial p_a^2} \right] \\
&= E_{\mathbf{X}|\mathbf{p}} \sum_{i=1}^N \left[\frac{x_{ia}}{p_a^2} + \frac{n - \sum_{j=1}^k x_{ij}}{(1 - \sum_{j=1}^k p_j)^2} \right] \\
&= nN \left(\frac{1}{p_a} + \frac{1}{1 - \sum_{j=1}^k p_j} \right).
\end{aligned} \tag{31}$$

Step 3: the non-informative prior:

$$\begin{aligned}
\pi(\mathbf{p})^2 &= [\det \mathbf{I}(\mathbf{p})] \\
&= nN \begin{vmatrix} p_1^{-1} + (1 - \sum_{j=1}^k p_j)^{-1} & (1 - \sum_{j=1}^k p_j)^{-1} & \cdots & (1 - \sum_{j=1}^k p_j)^{-1} \\ (1 - \sum_{j=1}^k p_j)^{-1} & p_2^{-1} + (1 - \sum_{j=1}^k p_j)^{-1} & \cdots & (1 - \sum_{j=1}^k p_j)^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ (1 - \sum_{j=1}^k p_j)^{-1} & (1 - \sum_{j=1}^k p_j)^{-1} & \cdots & p_k^{-1} + (1 - \sum_{j=1}^k p_j)^{-1} \end{vmatrix} \\
&\propto (p_1 p_2 \cdots p_k)^{-1}.
\end{aligned} \tag{32}$$

As a result,

$$\pi(\mathbf{p})^2 = (p_1 p_2 \cdots p_k)^{-1/2}.$$

□

Ex. 2.13. Suppose that $X_j \sim f(x_j | \theta_j)$, and the Jeffreys priors for θ_j are $\pi_j(\theta_j)$ ($j = 1, \dots, k$). If the X_j 's are independent, show that the Jeffreys prior for $\theta = (\theta_1, \dots, \theta_k)$ is $\pi(\theta) = \prod_{j=1}^k \pi_j(\theta_j)$.

The key to the problem is that for $\mathbf{X} = (X_1, \dots, X_k)$:

$$f(\mathbf{x}|\theta) = \prod_{j=1}^k f(x_j|\theta_j).$$

First, we know that the Jeffrey prior for θ_j can be calculated as

$$\pi_j(\theta_j)^2 = \sum_{i=1}^n E_{X_j|\theta_j} \left[\frac{\partial^2}{\partial \theta_j^2} \log f(x_{ij}|\theta_j) \right].$$

Next, consider the log-likelihood of \mathbf{X} :

$$\begin{aligned} \ell &= \sum_{i=1}^n \log f(\mathbf{x}|\theta) \\ &= \sum_{i=1}^n \log \prod_{j=1}^k f(x_{ij}|\theta_j) \\ &= \sum_{i=1}^n \sum_{j=1}^k \log f(x_{ij}|\theta_j). \end{aligned} \tag{33}$$

On the one hand, since $\partial^2 \ell / (\partial \theta_a \partial \theta_b) = 0$, we know $I_{ab}(\theta) = 0$. On the other hand,

$$I_{aa}(\theta) = \sum_{i=1}^n E_{X_a|\theta_a} \left[\frac{\partial^2}{\partial \theta_a^2} \log f(x_{ia}|\theta_a) \right],$$

which is exactly $\pi_a(\theta_a)^2$. Since the Fisher information matrix is a diagonal matrix, we know

$$\det \mathbf{I}(\theta) = \prod_{a=1}^k I_{aa}(\theta) = \prod_{a=1}^k \pi_a(\theta_a)^2.$$

As a result, the non-informative prior for θ is

$$[\det \mathbf{I}(\theta)]^{1/2} = \sqrt{\prod_{j=1}^k \pi_j(\theta_j)^2} = \prod_{j=1}^k \pi_j(\theta_j).$$

□

Ex. 2.14. A location-scale density is of the form $\sigma^{-1}f((x - \theta)/\sigma)$. Show that $\pi(\theta, \sigma) = 1/\sigma^2$ is the non-informative prior for the location-scale parameter (θ, σ) , using the invariance of the transformation group. The transformations are: $Y = cX + b$, $\eta = c\theta + b$ and $\xi = c\sigma$ ($b \in \mathbb{R}$, $c > 0$).

The transformations can be described as

$$\begin{aligned} X &\rightarrow Y = cX + b, \\ (\theta, \sigma) &\rightarrow (\eta, \xi) = (c\theta + b, c\sigma). \end{aligned}$$

Let π and π^* be the non-informative priors for (θ, σ) and (η, ξ) , respectively. Due to the invariance of the translation-scale transformation group, we have

$$\pi(\tau) = \pi^*(\tau).$$

On the other hand, we know

$$\pi^*(\eta, \xi) = \pi(\theta, \sigma)|_{\theta=(\eta-b)/c, \sigma=\xi/c} \cdot |J| = \frac{1}{c^2} \pi\left(\frac{\eta-b}{c}, \frac{\xi}{c}\right),$$

where

$$J = \begin{bmatrix} \partial\theta/\partial\eta & \partial\theta/\partial\xi \\ \partial\sigma/\partial\eta & \partial\sigma/\partial\xi \end{bmatrix} = \begin{bmatrix} 1/c & 0 \\ 0 & 1/c \end{bmatrix}.$$

Put the results together, we have

$$\pi(\eta, \xi) = \pi^*(\eta, \xi) = \frac{1}{c^2} \pi\left(\frac{\eta-b}{c}, \frac{\xi}{c}\right).$$

without losing of generality, take $\eta = b$ and $\xi = c$,

$$\pi(b, c) = \frac{1}{c^2} \pi(0, 1).$$

Let $\pi(0, 1) = 1$, we conclude that

$$\pi(\theta, \sigma) = \frac{1}{\sigma^2} \quad (\sigma > 0).$$

□

Ex. 2.15. Suppose that X follows the negative binomial distribution:

$$f(x|p) = C(x-1, k-1)p^k(1-p)^{x-k} \quad (x = k, k+1, \dots).$$

Show that the conjugate prior for p is of the Beta distribution.

Assume that

$$\pi(p|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}.$$

Then the posterior distribution of p is

$$\begin{aligned} \pi(p|x, \alpha, \beta) &= \frac{f(x|p)\pi(p|\alpha, \beta)}{\int_0^1 f(x|p)\pi(p|\alpha, \beta)dp} \\ &= \frac{p^k(1-p)^{x-k}p^{\alpha-1}(1-p)^{\beta-1}}{\int_0^1 p^k(1-p)^{x-k}p^{\alpha-1}(1-p)^{\beta-1}dp} \\ &= \frac{\Gamma(\alpha + x + \beta)}{\Gamma(k + \alpha)\Gamma(x - k + \beta)} p^{k+\alpha-1}(1-p)^{x-k+\beta-1}, \end{aligned} \tag{34}$$

indicating that $p|x, \alpha, \beta \sim \text{Beta}(\alpha + k, \beta + x - k)$. Since both the prior and the posterior are of Beta distribution, we conclude that Beta distribution family is the conjugate prior for p . □

Ex. 2.16. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} \text{Exp}(\theta),$$

and the prior distribution of θ is $\Gamma(r, \lambda)$.

(1) If we know that the expectation of the prior is 0.0002, and the variance is 0.0001. Determine the values for hyper-parameters.

(2) Show that $\Gamma(r, \lambda)$ is the conjugate prior distribution family for θ .

(1) Consider

$$\pi(\theta|r, \lambda) = \frac{\lambda^r}{\Gamma(r)} \theta^{r-1} e^{-\lambda\theta}.$$

Since $E(\theta) = r/\lambda = 0.0002$, and $\text{Var}(\theta) = r/\lambda^2 = 0.0001$, we have

$$r = 0.0004, \quad \lambda = 2.$$

(2) The likelihood is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} = \theta^n e^{-n\theta\bar{x}}.$$

The posterior distribution of θ is

$$\begin{aligned} \pi(\theta|\mathbf{x}, r, \lambda) &= \frac{f(\mathbf{x}|\theta)\pi(\theta|r, \lambda)}{\int_0^{+\infty} f(\mathbf{x}|\theta)\pi(\theta|r, \lambda)d\theta} \\ &= \frac{\theta^n e^{-n\theta\bar{x}} \theta^{r-1} e^{-\lambda\theta}}{\int_0^{+\infty} \theta^n e^{-n\theta\bar{x}} \theta^{r-1} e^{-\lambda\theta} d\theta} \\ &= \frac{(n\bar{x} + \lambda)^{n+r-1}}{\Gamma(n+r-1)} \theta^{n+r-1} e^{-(n\bar{x}+\lambda)\theta}, \end{aligned} \tag{35}$$

indicating that $\theta|\mathbf{x}, r, \lambda \sim \Gamma(n+r-1, n\bar{x}+\lambda)$. Since both the prior and the posterior are of Gamma distribution, we conclude that Gamma distribution family is the conjugate prior for θ . \square

Ex. 2.17. Suppose that X follows a distribution from the exponential family, with the density

$$f(x|\theta) = \exp\{a(\theta)b(x) + c(\theta) + d(x)\}.$$

Show that

$$h(\theta) = A \exp\{k_1 a(\theta) + k_2 c(\theta)\}$$

is the conjugate prior distribution of θ , where A is a constant and k_1, k_2 are independent of θ .

The posterior distribution is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)h(\theta)}{\int_{\Theta} f(x|\theta)h(\theta)d\theta} \\ &= \frac{\exp\{a(\theta)b(x) + c(\theta) + d(x)\} A \exp\{k_1 a(\theta) + k_2 c(\theta)\}}{\int_{\Theta} \exp\{a(\theta)b(x) + c(\theta) + d(x)\} A \exp\{k_1 a(\theta) + k_2 c(\theta)\} d\theta} \\ &= \frac{\exp\{(b(x) + k_1)a(\theta) + (1 + k_2)c(\theta)\}}{\int_{\Theta} \exp\{(b(x) + k_1)a(\theta) + (1 + k_2)c(\theta)\} d\theta}. \end{aligned} \tag{36}$$

It we denote

$$A^* = \left[\int_{\Theta} \exp\{(b(x) + k_1)a(\theta) + (1 + k_2)c(\theta)\} d\theta \right]^{-1},$$

then posterior is of the form

$$A^* \exp\{k_1^* a(\theta) + k_2^* c(\theta)\}$$

where A^* , k_1^* , k_2^* are all independent of θ . □

Ex. 2.18. Suppose that X follows the distribution

$$f(x|\lambda) = \begin{cases} \lambda^{-1} e^{-x/\lambda}, & 0 < x < \infty, \\ 0, & t \leq 0. \end{cases}$$

Show that the conjugate prior distribution for λ is the Inverse Gamma family.

If the prior for λ is $IG(\alpha, \beta)$:

$$\pi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha-1} e^{-\beta/\lambda},$$

then the posterior is:

$$\begin{aligned} \pi(x|\lambda) &= \frac{f(x|\lambda)\pi(\lambda)}{\int_0^{+\infty} f(x|\lambda)\pi(\lambda)d\lambda} \\ &= \frac{\lambda^{-1} e^{-x/\lambda} \lambda^{-\alpha-1} e^{-\beta/\lambda}}{\int_0^{+\infty} \lambda^{-1} e^{-x/\lambda} \lambda^{-\alpha-1} e^{-\beta/\lambda} d\lambda} \\ &= \frac{(x + \beta)^{\alpha+1}}{\Gamma(\alpha + 1)} \lambda^{-\alpha-2} e^{-(x+\beta)/\lambda}, \end{aligned} \tag{37}$$

which is also a member in the Inverse Gamma family. □

Ex. 2.19. Suppose that

$$X_1, \dots, X_n \sim \text{i.i.d. } U(0, \theta),$$

and the prior distribution of θ is the Pareto:

$$\pi(\theta) = \begin{cases} \alpha \theta_0^\alpha / \theta^{\alpha+1}, & \theta > \theta_0, \\ 0, & \theta \leq \theta_0, \end{cases}$$

where $\theta_0 > 0$, $\alpha > 0$. Show that the Pareto distribution is the conjugate prior for θ .

The posterior is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta} \\ &= \frac{\prod_{i=1}^n \left[\frac{1}{\theta} I(0 \leq x_i \leq \theta) \right] \alpha \theta_0^\alpha / \theta^{\alpha+1} I(\theta > \theta_0)}{\int_{\Theta_0} \prod_{i=1}^n \left[\frac{1}{\theta} I(0 \leq x_i \leq \theta) \right] \alpha \theta_0^\alpha / \theta^{\alpha+1} I(\theta > \theta_0) d\theta} \\ &= \begin{cases} \frac{\alpha \theta_0^\alpha / \theta^{n+\alpha+1}}{\int_K^{+\infty} \alpha \theta_0^\alpha / \theta^{n+\alpha+1} d\theta}, & K \leq \theta, \\ 0, & \text{otherwise,} \end{cases} \\ &= \begin{cases} (n + \alpha) \theta_0^{n+\alpha} / \theta^{n+\alpha+1}, & K \leq \theta, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{38}$$

where $K = \max\{\theta_0, x_1, \dots, x_n\}$. Thus, the posterior is also of the form of the Pareto distribution. \square

Ex. 2.20. Suppose that $X \sim N(\theta, 1)$, where $\theta > 0$ and the prior expectation of θ is μ . Show that the maximum entropy prior is $\text{Exp}(1/\mu)$.

The constraint is

$$E^\pi(\theta) = \int_0^{+\infty} \theta \pi(\theta) d\theta = \mu.$$

According to the formula, the maximum entropy prior is

$$\tilde{\pi}(\theta) = \frac{\exp\{\lambda\theta\}}{\int_0^{+\infty} \exp\{\lambda\theta\}} = A^* \exp\{\lambda\theta\},$$

which is of the form of the exponential distribution $\text{Exp}(-\lambda)$. Therefore,

$$E^{\tilde{\pi}}(\theta) = -\frac{1}{\lambda}.$$

Let $\pi(\theta) = \tilde{\pi}(\theta)$, we have $\lambda = -1/\mu$, and the maximum entropy prior is $\text{Exp}(1/\mu)$. \square

Ex. 2.21. Suppose that θ is a scale parameter (and thus $\pi_0(\theta) = \theta^{-1}$). Assume that $\theta \in (a, b)$, and the median of the prior $\pi(\theta)$ is z :

$$\int_a^z \pi(\theta) d\theta = \int_z^b \pi(\theta) d\theta = \frac{1}{2}.$$

Show that the maximum entropy prior is

$$\pi(\theta) = \begin{cases} \frac{1}{\theta} (2 \log \frac{z}{a})^{-1}, & 0 < a < \theta < z, \\ \frac{1}{\theta} (2 \log \frac{b}{z})^{-1}, & z < \theta < b. \end{cases}$$

The constraints are

$$\int_a^b I(\theta < z) \pi(\theta) d\theta = \frac{1}{2}, \quad \int_a^b I(\theta > z) \pi(\theta) d\theta = \frac{1}{2}.$$

According to the formula, the maximum entropy prior is

$$\begin{aligned} \tilde{\pi}(\theta) &= \frac{\frac{1}{\theta} \exp\{\lambda_1 I(\theta < z) + \lambda_2 I(\theta > z)\}}{\int_a^b \frac{1}{\theta} \exp\{\lambda_1 I(\theta < z) + \lambda_2 I(\theta > z)\} d\theta} \\ &= A^* \frac{1}{\theta} \exp\{\lambda_1 I(\theta < z) + \lambda_2 I(\theta > z)\} \\ &= \begin{cases} A^*/\theta \exp\{\lambda_1\}, & 0 < a < \theta < z, \\ A^*/\theta \exp\{\lambda_2\}, & z < \theta < b, \end{cases} \end{aligned} \tag{39}$$

where

$$A^* = \left[\int_a^b \frac{1}{\theta} \exp\{\lambda_1 I(\theta < z) + \lambda_2 I(\theta > z)\} d\theta \right]^{-1}.$$

Now let $\pi(\theta) = \tilde{\pi}(\theta)$:

$$\begin{aligned}\int_a^b I(\theta < z) \tilde{\pi}(\theta) d\theta &= \int_a^z \frac{A^*}{\theta} \exp\{\lambda_1\} = A^* \exp\{\lambda_1\} \log \frac{z}{a} = \frac{1}{2}, \\ \int_a^b I(\theta > z) \tilde{\pi}(\theta) d\theta &= \int_z^b \frac{A^*}{\theta} \exp\{\lambda_2\} = A^* \exp\{\lambda_2\} \log \frac{b}{z} = \frac{1}{2},\end{aligned}$$

then we can solve for the unknowns:

$$A^* \lambda_1 = \frac{1}{2} \log \frac{a}{z}, \quad A^* \lambda_2 = \frac{1}{2} \log \frac{z}{b}.$$

As a result, the maximum entropy prior is

$$\pi(\theta) = \begin{cases} \frac{1}{\theta} (2 \log \frac{z}{a})^{-1}, & 0 < a < \theta < z, \\ \frac{1}{\theta} (2 \log \frac{b}{z})^{-1}, & z < \theta < b. \end{cases}$$

□

Ex. 2.22. Suppose that X_1 and X_2 are independent, and they follow the exponential distributions with expectations of μ_1 and μ_2 , respectively. Assume that the interested parameter is $\phi_1 = \mu_2/\mu_1$, while $\phi_2 = \mu_2 \cdot \mu_1$ is the nuisance parameter. Show that the Reference prior for (ϕ_1, ϕ_2) is $\pi(\phi_1, \phi_2) = (\phi_1 \phi_2)^{-1}$.

Before we start to find the reference prior, let us find the Fisher information matrix first. Since

$$X_1 \sim \frac{1}{\mu_1} e^{-\frac{x_1}{\mu_1}}, \quad X_2 \sim \frac{1}{\mu_2} e^{-\frac{x_2}{\mu_2}},$$

the log-likelihood is

$$\begin{aligned}\ell &= \log f(x_1) + \log f(x_2) \\ &= -\log \mu_1 - \frac{x_1}{\mu_1} - \log \mu_2 - \frac{x_2}{\mu_2} \\ &= \log \phi_2 - x_1 \sqrt{\frac{\phi_1}{\phi_2}} - x_2 \sqrt{\frac{1}{\phi_1 \phi_2}}.\end{aligned}\tag{40}$$

Consequently we can find the information matrix

$$\mathbf{I}(\phi_1, \phi_2) = \begin{bmatrix} 1/(2\phi_1^2) & 0 \\ 0 & 1/(2\phi_2^2) \end{bmatrix}$$

Step 1: The conditional Reference prior:

$$\pi(\phi_2|\phi_1) = \sqrt{\frac{1}{2\phi_2^2}} \propto \frac{1}{\phi_2}.$$

Step 2: Take the monotone-increasing subset $\Omega_i = L_i \times S_i$ from the parameter space $\Omega = \mathbb{R}^+ \times \mathbb{R}^+$, where $L_i = [l_{i1}, l_{i2}]$, $S_i = [s_{i1}, s_{i2}]$, such that $L_1 \subset L_2 \subset \dots$, $S_1 \subset S_2 \subset \dots$, $\bigcup_{i=1}^{\infty} L_i = \mathbb{R}^+$, and $\bigcup_{i=1}^{\infty} S_i = \mathbb{R}^+$. Let $\Omega_{i,\phi_1} = \{\phi_2 : (\phi_1, \phi_2) \in \Phi_i\} = S_i$, then

$$K_i(\phi_1) = \left[\int_{S_i} \pi(\phi_2|\phi_1) d\phi_2 \right]^{-1} = \left[\int_{s_{i1}}^{s_{i2}} \frac{1}{\phi_2} d\phi_2 \right]^{-1} = \frac{1}{\log s_{i2} - \log s_{i1}},$$

$$\pi_i(\phi_2|\phi_1) = K_i(\phi_1) \cdot \pi(\phi_2|\phi_1) \cdot I_{S_i}(\phi_2) = \frac{1}{(\log s_{i2} - \log s_{i1})\phi_2} (s_{i1} \leq \phi_2 \leq s_{i2}).$$

Step 3: The marginal Reference prior:

$$\begin{aligned} \pi_i(\phi_1) &= \exp \left\{ \frac{1}{2} \int_{S_i} \pi_i(\phi_2|\phi_1) \log \frac{|\mathbf{I}(\phi_1, \phi_2)|}{|I_{22}(\phi_1, \phi_2)|} d\phi_2 \right\} \\ &= \exp \left\{ \frac{1}{2} \int_{s_{i1}}^{s_{i2}} \frac{-\log 2\phi_1^2}{(\log s_{i2} - \log s_{i1})\phi_2} d\phi_2 \right\} \\ &= \exp \left\{ -\frac{1}{2} \log 2\phi_1^2 \right\} \\ &= \frac{1}{\sqrt{2}\phi_1}. \end{aligned} \tag{41}$$

Step 4: The limit:

$$\pi(\phi_1, \phi_2) = \lim_{i \rightarrow \infty} \frac{K_i(\phi_1)\pi_i(\phi_1)}{K_i(\phi_{10})\pi_i(\phi_{10})} \pi(\phi_2|\phi_1) \propto \pi_i(\phi_1)\pi(\phi_2|\phi_1) \propto (\phi_1\phi_2)^{-1}.$$

□

Ex. 2.23. Suppose that

$$X_i \sim^{\text{i.i.d.}} N(\theta_i, 900) \quad (i = 1, \dots, p).$$

We believe that θ_i 's are similar, and again assume that they are i.i.d samples form a distribution, and the mean of those θ_i 's is about 100, while the standard deviation of the guess of the mean is about 20. The variance of θ_i 's is an unknown constant, so it's set to be a non-informative constant prior. Find a reasonable multi-level prior model for the description above.

$$\begin{aligned} X_i &\sim^{\text{i.i.d.}} N(\theta_i, 900) \quad (i = 1, \dots, p), \\ \theta_i &\sim^{\text{i.i.d.}} N(k_i, C), \\ k_i &\sim^{\text{i.i.d.}} N(100, 20^2). \end{aligned}$$

□

3 Bayesian Statistical Inference

Ex. 3.1. Suppose that $X \sim B(n, \theta)$.

(1) If the prior is $\pi(\theta) = [\theta(1-\theta)]^{-1}I(0 < \theta < 1)$, find the posterior of θ given x , when $1 \leq x \leq n-1$.

(2) If $\pi(\theta) = I(0 < \theta < 1)$, find the posterior of θ given x .

(1) The posterior is

$$\begin{aligned}\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_0^1 f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{C(n, x)\theta^x(1-\theta)^{n-x}\theta^{-1}(1-\theta)^{-1}}{\int_0^1 C(n, x)\theta^x(1-\theta)^{n-x}\theta^{-1}(1-\theta)^{-1}d\theta} \\ &= \frac{\Gamma(n)}{\Gamma(x)\Gamma(n-x)}\theta^{x-1}(1-\theta)^{n-x-1}, \quad 1 \leq x \leq n-1.\end{aligned}\tag{42}$$

(2) The posterior is

$$\begin{aligned}\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_0^1 f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{C(n, x)\theta^x(1-\theta)^{n-x}}{\int_0^1 C(n, x)\theta^x(1-\theta)^{n-x}d\theta} \\ &= \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}, \quad 0 < \theta < 1.\end{aligned}\tag{43}$$

□

Ex. 3.2. Suppose that

$$X_1, \dots, X_n \sim \text{i.i.d. } P(\theta),$$

and the prior for θ is $\pi(\theta) = \theta^{-1}I(\theta > 0)$. Find the posterior of θ given \mathbf{x} , when $\mathbf{x} \neq (0, \dots, 0)$.

The likelihood is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\theta^{n\bar{x}}}{\prod_{i=1}^n x_i!}.$$

Then the posterior is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_0^{+\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta} \\ &= \frac{e^{-n\theta}\theta^{n\bar{x}}\theta^{-1}}{\int_0^{+\infty} e^{-n\theta}\theta^{n\bar{x}}\theta^{-1}d\theta} \\ &= \frac{n^{n\bar{x}}}{\Gamma(n\bar{x})}\theta^{n\bar{x}-1}e^{-n\theta}, \quad \theta > 0.\end{aligned}\tag{44}$$

□

Ex. 3.3. Suppose that

$$X_1, \dots, X_n \sim \text{i.i.d. } N(\theta, \sigma^2)$$

where θ and σ^2 are unknown. Assume that the joint prior for (θ, σ^2) is

$$\pi(\theta, \sigma^2) = \sigma^{-2} I(\sigma^2 > 0).$$

Show that

(1) The posterior of (θ, σ^2) is

$$\pi(\theta, \sigma^2 | \mathbf{x}) = \pi_1(\theta | \sigma^2, \mathbf{x}) \pi_2(\sigma^2 | \mathbf{x}),$$

where $\pi_1(\theta | \sigma^2, \mathbf{x})$ is $N(\bar{x}, \sigma^2/n)$, and $\pi_2(\sigma^2 | \mathbf{x})$ is of the Inverse Gamma distribution

$$\Gamma^{-1} \left(\frac{n-1}{2}, \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 \right).$$

(2) The marginal posterior of σ^2 is $\Gamma^{-1} \left(\frac{n-1}{2}, \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 \right)$ given \mathbf{x} .

(3) The marginal posterior of θ is

$$\mathcal{T} \left(n-1, \bar{x}, \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)} \right),$$

given \mathbf{x} .

(1) Since (\bar{X}, S^2) is the sufficient statistic of (θ, σ^2) , where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\theta, \frac{\sigma^2}{n}), \quad S^2 = \frac{1}{\nu} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \nu = n-1, \quad \frac{\nu S^2}{\sigma^2} \sim \chi_\nu^2,$$

the likelihood is

$$\begin{aligned} f(t|\theta, \sigma^2) &= f_1(\bar{x}|\theta, \sigma^2) f_2(s^2|\sigma^2) \\ &= \sqrt{\frac{n}{2\pi\sigma^2}} \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \theta)^2 \right\} \frac{\left(\frac{\nu}{2\sigma^2} \right)^{\nu/2}}{\Gamma(\nu/2)} (s^2)^{\nu/2-1} e^{-\nu s^2/2\sigma^2} \\ &\propto \sigma^{-\nu-1} \exp \left\{ -\frac{1}{2\sigma^2} [\nu s^2 + n(\bar{x} - \theta)^2] \right\}. \end{aligned} \quad (45)$$

Then the posterior is

$$\begin{aligned} \pi(\theta, \sigma^2 | t) &\propto f(t|\theta, \sigma^2) \pi(\theta, \sigma^2) \\ &\propto \sigma^{-\nu-3} \exp \left\{ -\frac{1}{2\sigma^2} [\nu s^2 + n(\bar{x} - \theta)^2] \right\}. \end{aligned} \quad (46)$$

In order to normalize the posterior, consider

$$\begin{aligned} K &= \int_0^{+\infty} \int_{-\infty}^{+\infty} \sigma^{-\nu-3} \exp \left\{ -\frac{1}{2\sigma^2} [\nu s^2 + n(\bar{x} - \theta)^2] \right\} d\theta d\sigma^2 \\ &= \int_0^{+\infty} \sigma^{-\nu-3} \exp \left\{ -\frac{\nu s^2}{2\sigma^2} \right\} \sqrt{\frac{2\pi\sigma^2}{n}} d\sigma^2 \\ &= \sqrt{\frac{2\pi}{n}} \frac{\Gamma(\nu/2)}{(vs^2/2)^{\nu/2}}, \end{aligned} \quad (47)$$

and therefore the posterior is

$$\pi(\theta, \sigma^2 | t) = K^{-1} \sigma^{-\nu-3} \exp \left\{ -\frac{1}{2\sigma^2} [\nu s^2 + n(\bar{x} - \theta)^2] \right\}.$$

Of course $\pi(\theta, \sigma^2 | \mathbf{x})$ can be written as $\pi_1(\theta | \sigma^2, \mathbf{x}) \pi_2(\sigma^2 | \mathbf{x})$. It is easy to find $\pi_1(\theta | \sigma^2, \mathbf{x})$ when the prior $\pi_1^*(\theta) \equiv 1$ and the σ^2 is known:

$$\begin{aligned} \pi_1(\theta | \sigma^2, \bar{x}) &= \frac{f(\bar{x} | \theta, \sigma^2) \pi_1^*(\theta)}{\int_{\Theta} f(\bar{x} | \theta, \sigma^2) \pi_1^*(\theta) d\theta} \\ &= \frac{\sqrt{n/(2\pi\sigma^2)} \exp \{ -n(\bar{x} - \theta)^2 / (2\sigma^2) \} \cdot 1}{\int_{-\infty}^{+\infty} \sqrt{n/(2\pi\sigma^2)} \exp \{ -n(\bar{x} - \theta)^2 / (2\sigma^2) \} \cdot 1 d\theta} \\ &= \sqrt{\frac{n}{2\pi\sigma^2}} \exp \left\{ -\frac{n}{2\sigma^2} (\theta - \bar{x})^2 \right\}, \end{aligned} \quad (48)$$

indicating that $\pi_1(\theta | \sigma^2, \mathbf{x})$ is $N(\bar{x}, \sigma^2/n)$. Consequently,

$$\begin{aligned} \pi_2(\sigma^2 | \mathbf{x}) &= \frac{\pi(\theta, \sigma^2 | t)}{\pi_1(\theta | \sigma^2, \bar{x})} \\ &= \frac{(\nu s^2/2)^{\nu/2}}{\Gamma(\nu/2)} (\sigma^2)^{-(\nu/2+1)} \exp \left\{ -\frac{\nu s^2}{2\sigma^2} \right\}, \end{aligned} \quad (49)$$

which is exactly $\Gamma^{-1}(\nu/2, \nu s^2/2)$. Plug the expressions of ν and s^2 in, we have

$$\Gamma^{-1} \left(\frac{n-1}{2}, \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 \right).$$

(2) The marginal posterior of σ^2 given \mathbf{x} is

$$\begin{aligned} \pi(\sigma^2 | \mathbf{x}) &= \int_{-\infty}^{+\infty} \pi(\theta, \sigma^2 | t) d\theta \\ &= \int_{-\infty}^{+\infty} K^{-1} \sigma^{-\nu-3} \exp \left\{ -\frac{1}{2\sigma^2} [\nu s^2 + n(\bar{x} - \theta)^2] \right\} d\theta \\ &= K^{-1} \sigma^{-\nu-3} \exp \left\{ -\frac{\nu s^2}{2\sigma^2} \right\} \sqrt{\frac{2\pi\sigma^2}{n}} \\ &= \frac{(\nu s^2/2)^{\nu/2}}{\Gamma(\nu/2)} (\sigma^2)^{-(\nu/2+1)} \exp \left\{ -\frac{\nu s^2}{2\sigma^2} \right\}, \end{aligned} \quad (50)$$

which is $\Gamma^{-1}(\frac{n-1}{2}, \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2)$ given \mathbf{x} .

(3) The marginal posterior of θ given \mathbf{x} is

$$\begin{aligned} \pi(\theta | \mathbf{x}) &= \int_0^{+\infty} \pi(\theta, \sigma^2 | t) d\sigma^2 \\ &= \int_0^{+\infty} K^{-1} \sigma^{-\nu-3} \exp \left\{ -\frac{1}{2\sigma^2} [\nu s^2 + n(\bar{x} - \theta)^2] \right\} d\sigma^2 \\ &= K^{-1} \Gamma\left(\frac{\nu+1}{2}\right) \left\{ \frac{\nu s^2 + n(\bar{x} - \theta)^2}{2} \right\}^{-\frac{\nu+1}{2}} \\ &= \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{1}{\sqrt{\nu\pi}} \frac{\sqrt{n(n-1)}}{\sqrt{\nu s^2}} \left\{ 1 + \frac{1}{\nu} \frac{(\theta - \bar{x})^2 n(n-1)}{\nu s^2} \right\}^{-\frac{\nu+1}{2}}, \end{aligned} \quad (51)$$

which is exactly

$$\mathcal{T}\left(n-1, \bar{x}, \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}\right).$$

□

Ex. 3.4. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} N(\theta, 2^2),$$

and the prior for θ is of Gaussian.

(1) If $n = 100$, show that the posterior standard deviation must be less than $1/5$ regardless of the prior standard deviation.

(2) If the standard deviation of the prior for θ is 1, find the minimum sample size to guarantee that the posterior variance is less than or equal to 0.1.

It is acknowledged that the posterior is $N(u(\mathbf{x}), \eta^2)$, where

$$u(\mathbf{x}) = \frac{4/n}{4/n + \tau^2} \mu + \frac{\tau^2}{4/n + \tau^2} \bar{x}, \quad \eta^2 = \frac{4\tau^2}{4 + n\tau^2}$$

if the prior is $N(\mu, \tau^2)$.

(1) If $n = 100$, then

$$\eta^2 = \frac{4\tau^2}{4 + n\tau^2} = \frac{4\tau^2}{4 + 100\tau^2} = \frac{1}{1/\tau^2 + 25} \leq \frac{1}{25},$$

so $\eta \leq 1/5$ for all τ .

(2) If $\tau = 1$, and let

$$\eta^2 = \frac{4\tau^2}{4 + n\tau^2} = \frac{4}{4 + n} \leq 0.1,$$

then $n \geq 36$.

□

Ex. 3.5. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} N(\theta_1, \sigma^2).$$

Let $\theta_2 = 1/(2\sigma^2)$, and (θ_1, θ_2) enjoys the following assumptions:

(1) $\theta_1 | \theta_2 \sim N(0, 1/(2\theta_2))$;

(2) $\theta_2 \sim \Gamma(\alpha, \lambda)$, where α and λ are known.

Find the joint posterior $\pi(\theta_1, \theta_2 | \mathbf{x})$.

Since (\bar{X}, S^2) is the sufficient statistic of (θ_1, σ^2) , where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\theta_1, \frac{\sigma^2}{n}\right), \quad S^2 = \frac{1}{\nu} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \nu = n-1, \quad \frac{\nu S^2}{\sigma^2} \sim \chi_\nu^2,$$

the likelihood is

$$\begin{aligned} f(t | \theta_1, \sigma^2) &= f_1(\bar{x} | \theta_1, \sigma^2) f_2(s^2 | \sigma^2) \\ &= \sqrt{\frac{n}{2\pi\sigma^2}} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x} - \theta_1)^2\right\} \frac{\left(\frac{\nu}{2\sigma^2}\right)^{\nu/2}}{\Gamma(\nu/2)} (s^2)^{\nu/2-1} e^{-\nu s^2/2\sigma^2} \\ &\propto \sigma^{-\nu-1} \exp\left\{-\frac{1}{2\sigma^2} [\nu s^2 + n(\bar{x} - \theta_1)^2]\right\} \\ &\propto \theta_2^{(\nu+1)/2} \exp\left\{-\theta_2 [\nu s^2 + n(\bar{x} - \theta_1)^2]\right\} \\ &\propto f(t | \theta_1, \theta_2). \end{aligned} \tag{52}$$

Consequently, the posterior is

$$\begin{aligned}\pi(\theta_1, \theta_2 | \mathbf{x}) &= \frac{f(t|\theta_1, \theta_2)\pi(\theta_1|\theta_2)\pi(\theta_2)}{\int_{\Theta_2} \int_{\Theta_1} f(t|\theta_1, \theta_2)\pi(\theta_1|\theta_2)\pi(\theta_2)d\theta_1 d\theta_2} \\ &= \frac{\theta_2^{(\nu+1)/2+1/2+\alpha-1} \exp\{-\theta_2 \nu s^2 - \lambda \theta_2\} \exp\{-\theta_2 n(\bar{x} - \theta_1)^2 - \theta_2 \theta_1^2\}}{\int_0^{+\infty} \int_{-\infty}^{+\infty} \theta_2^{(\nu+1)/2+1/2+\alpha-1} \exp\{-\theta_2 \nu s^2 - \lambda \theta_2\} \exp\{-\theta_2 n(\bar{x} - \theta_1)^2 - \theta_2 \theta_1^2\} d\theta_1 d\theta_2}.\end{aligned}\quad (53)$$

Since

$$\begin{aligned}& \int_{-\infty}^{+\infty} \exp\{-\theta_2 n(\bar{x} - \theta_1)^2 - \theta_2 \theta_1^2\} d\theta_1 \\ &= \int_{-\infty}^{+\infty} \exp\left\{-(n+1)\theta_2 \left[\theta_1 - \left(\frac{n\bar{x}}{n+1}\right)\right]^2\right\} \exp\left\{\left[\frac{(n\bar{x})^2}{n+1} - n\bar{x}^2\right] \theta_2\right\} d\theta_1 \\ &= \exp\left\{\left[\frac{(n\bar{x})^2}{n+1} - n\bar{x}^2\right] \theta_2\right\} \sqrt{\frac{\pi}{n+1}} \theta_2^{-1/2},\end{aligned}\quad (54)$$

the denominator of $\pi(\theta_1, \theta_2 | \mathbf{x})$ is

$$\begin{aligned}K &= \sqrt{\frac{\pi}{n+1}} \int_0^{+\infty} \theta_2^{(\nu+1)/2+\alpha-1} \exp\left\{-\left[\nu s^2 + \lambda + \frac{(n\bar{x})^2}{n+1} - n\bar{x}^2\right] \theta_2\right\} d\theta_2 \\ &= \sqrt{\frac{\pi}{n+1}} \frac{\left[\nu s^2 + \lambda + \frac{(n\bar{x})^2}{n+1} - n\bar{x}^2\right]^{(\nu+1)/2+\alpha}}{\Gamma((\nu+1)/2+\alpha)}.\end{aligned}\quad (55)$$

Therefore, the joint posterior is

$$\pi(\theta_1, \theta_2 | \mathbf{x}) = K^{-1} \theta_2^{(\nu+1)/2+1/2+\alpha-1} \exp\left\{-\left[\nu s^2 + \lambda + n(\bar{x} - \theta_1)^2 + \theta_1^2\right] \theta_2\right\}.$$

□

Ex. 3.6. Suppose that

$$\mathbf{X} = (X_1, \dots, X_n) \sim^{\text{i.i.d.}} \text{Exp}(1/\theta),$$

and the prior for θ is $\Gamma^{-1}(\alpha, \beta)$. Show that the posterior for θ is

$$\Gamma^{-1}\left(n + \alpha, \sum_{i=1}^n x_i + \beta\right),$$

given \mathbf{x} .

Since the likelihood is

$$f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-x_i/\theta} = \frac{1}{\theta^n} e^{-\sum_{i=1}^n x_i/\theta},$$

the posterior for θ is

$$\begin{aligned}\pi(\theta | \mathbf{x}) &= \frac{f(x_1, \dots, x_n | \theta)\pi(\theta)}{\int_{\mathbb{R}^+} f(x_1, \dots, x_n | \theta)\pi(\theta)d\theta} \\ &= \frac{\theta^{-n} \exp\left\{-\theta^{-1} \sum_{i=1}^n x_i\right\} \theta^{-(\alpha+1)} \exp\{\theta^{-1}\beta\}}{\int_{\mathbb{R}^+} \theta^{-n} \exp\left\{-\theta^{-1} \sum_{i=1}^n x_i\right\} \theta^{-(\alpha+1)} \exp\{\theta^{-1}\beta\} d\theta} \\ &= \frac{(\sum_{i=1}^n x_i + \beta)^{\alpha+n}}{\Gamma(\alpha+n)} \theta^{-(\alpha+n+1)} \exp\left\{\theta^{-1} \left(\sum_{i=1}^n x_i + \beta\right)\right\}.\end{aligned}\quad (56)$$

□

Ex. 3.7. Suppose that

$$\mathbf{X} = (X_1, \dots, X_n) \sim \text{i.i.d. } U(0, \theta),$$

and the prior for θ is $Pa(\theta_0, \alpha)$. Show that the posterior for θ is

$$Pa(\max\{\theta_0, x_1, \dots, x_n\}, n + \alpha),$$

given \mathbf{x} .

The posterior is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta} \\ &= \frac{\prod_{i=1}^n \left[\frac{1}{\theta} I(0 \leq x_i \leq \theta)\right] \alpha \theta_0^\alpha / \theta^{\alpha+1} I(\theta > \theta_0)}{\int_{\Theta_0} \prod_{i=1}^n \left[\frac{1}{\theta} I(0 \leq x_i \leq \theta)\right] \alpha \theta_0^\alpha / \theta^{\alpha+1} I(\theta > \theta_0) d\theta} \\ &= \begin{cases} \frac{\alpha \theta_0^\alpha / \theta^{n+\alpha+1}}{\int_K^{+\infty} \alpha \theta_0^\alpha / \theta^{n+\alpha+1} d\theta}, & K \leq \theta, \\ 0, & \text{otherwise,} \end{cases} \\ &= \begin{cases} (n + \alpha) K^{n+\alpha} / \theta^{n+\alpha+1}, & K \leq \theta, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (57)$$

where $K = \max\{\theta_0, x_1, \dots, x_n\}$. Thus, the posterior is exactly $Pa(\max\{\theta_0, x_1, \dots, x_n\}, n + \alpha)$. \square

Ex. 3.8. Suppose that $X \sim \Gamma(n/2, 1/(2\theta))$, and the prior for θ is $\Gamma^{-1}(\alpha, \beta/2)$. Show that the posterior for θ is $\Gamma^{-1}(n/2 + \alpha, x/2 + \beta/2)$ given \mathbf{x} .

The posterior is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{(2\theta)^{-n/2} \exp\{-x/(2\theta)\} \theta^{-(\alpha+1)} \exp\{-\beta/(2\theta)\}}{\int_{\mathbb{R}^+} (2\theta)^{-n/2} \exp\{-x/(2\theta)\} \theta^{-(\alpha+1)} \exp\{-\beta/(2\theta)\} d\theta} \\ &= \frac{(x/2 + \beta/2)^{n/2+\alpha}}{\Gamma(n/2 + \alpha)} \theta^{-(n/2+\alpha+1)} \exp\left\{-\frac{x + \beta}{2\theta}\right\}, \end{aligned} \quad (58)$$

which is exactly $\Gamma^{-1}(n/2 + \alpha, x/2 + \beta/2)$. \square

Ex. 3.9. Suppose that

$$\mathbf{X} = (X_1, \dots, X_n) \sim \text{i.i.d. } Nb(r, \theta),$$

and the prior for θ is $Be(\alpha, \beta)$. Show that the posterior for θ is $Be(\alpha + rn, \sum_{i=1}^n x_i - nr + \beta)$, given \mathbf{x} .

The likelihood is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n [C(x_i - 1, r - 1) \theta^r (1 - \theta)^{x_i - r}] = \prod_{i=1}^n C(x_i - 1, r - 1) \theta^{rn} (1 - \theta)^{\sum_{i=1}^n x_i - rn}. \quad (59)$$

Consequently, the posterior is

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_0^1 f(\mathbf{x}|\theta)\pi(\theta)d\theta} \\
&= \frac{\theta^{rn}(1-\theta)^{\sum_{i=1}^n x_i - rn} \theta^{\alpha-1} (1-\theta)^{\beta-1}}{\int_0^1 \theta^{rn}(1-\theta)^{\sum_{i=1}^n x_i - rn} \theta^{\alpha-1} (1-\theta)^{\beta-1} d\theta} \\
&= \frac{\Gamma(\sum_{i=1}^n x_i + \alpha + \beta)}{\Gamma(\alpha + rn)\Gamma(\sum_{i=1}^n x_i - nr + \beta)} \theta^{\alpha+rn-1} (1-\theta)^{\sum_{i=1}^n x_i - nr + \beta - 1},
\end{aligned} \tag{60}$$

which is exactly $Be(\alpha + rn, \sum_{i=1}^n x_i - nr + \beta)$. \square

Ex. 3.10. Suppose that $\mathbf{X} = (X_1, \dots, X_p)$ follows a multivariate Gaussian $N_p(\theta, \Sigma)$, and the prior of θ is $N_P(\mu, A)$. Show that the posterior of θ given \mathbf{x} is another p -Gaussian with the mean

$$x - \Sigma(\Sigma + A)^{-1}(x - \mu)$$

and the variance

$$(\Sigma^{-1} + A^{-1})^{-1}.$$

The prior is

$$\pi(\theta) = \frac{1}{(2\pi)^{p/2}|A|^{1/2}} \exp \left\{ -\frac{1}{2}(\theta - \mu)^T A^{-1}(\theta - \mu) \right\},$$

and the likelihood is

$$f(x|\theta) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(x - \theta)^T \Sigma^{-1}(x - \theta) \right\}.$$

Therefore, the posterior is

$$\begin{aligned}
\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}} f(x|\theta)\pi(\theta)d\theta} \\
&\propto \exp \left\{ -\frac{1}{2}(\theta - \mu)^T A^{-1}(\theta - \mu) - \frac{1}{2}(x - \theta)^T \Sigma^{-1}(x - \theta) \right\}.
\end{aligned} \tag{61}$$

Let

$$\begin{aligned}
N^* &= (\theta - \mu)^T A^{-1}(\theta - \mu) + (x - \theta)^T \Sigma^{-1}(x - \theta) \\
&= \theta^T (A^{-1} + \Sigma^{-1})\theta - 2(\mu^T A^{-1} + x^T \Sigma^{-1})\theta + Const_1 \\
&= [\theta^T - (\mu^T A^{-1} + x^T \Sigma^{-1})B^{-1}] B [\theta - B^{-1}(A^{-1}\mu + \Sigma^{-1}x)] + Const_2,
\end{aligned} \tag{62}$$

where $B = A^{-1} + \Sigma^{-1}$, and $Const_1, Const_2$ are not functions of θ . Plug N^* in $\pi(\theta|x)$ and normalize it, we can conclude that the variance of the posterior is (by Sherman-Morrison-Woodbury)

$$B^{-1} = (A^{-1} + \Sigma^{-1})^{-1} = \Sigma - \Sigma(A + \Sigma)^{-1}\Sigma,$$

and the mean is

$$\begin{aligned}
B^{-1}(A^{-1}\mu + \Sigma^{-1}x) &= (\Sigma - \Sigma(A + \Sigma)^{-1}\Sigma)(A^{-1}\mu + \Sigma^{-1}x) \\
&= \Sigma A^{-1}\mu + x - \Sigma(A + \Sigma)^{-1}\Sigma A^{-1}\mu - \Sigma(A + \Sigma)^{-1}x \\
&= \Sigma(\Sigma + A)^{-1}\mu + x - \Sigma(A + \Sigma)^{-1}x \\
&= x + \Sigma(\Sigma + A)^{-1}(\mu - x).
\end{aligned} \tag{63}$$

The second to last equation holds because

$$A^{-1} - (A + \Sigma)^{-1} \Sigma A^{-1} = (\Sigma + A)^{-1}.$$

□

Ex. 3.11. Suppose that

$$\mathbf{X} = (X_1, \dots, X_n) \sim^{\text{i.i.d.}} N(\theta, \sigma^2),$$

and the joint prior of (θ, σ^2) is

$$\pi(\theta, \sigma^2) = \pi_1(\theta|\sigma^2)\pi_2(\sigma^2),$$

where $\pi_1(\theta|\sigma^2) = N(\mu, \tau\sigma^2)$, and $\pi_2(\sigma^2) = \Gamma^{-1}(\alpha, \beta)$. Show that

(1) the joint posterior of (θ, σ^2) given \mathbf{x} is

$$\pi(\theta, \sigma^2|\mathbf{x}) = \pi_1(\theta|\sigma^2, \mathbf{x})\pi_2(\sigma^2|\mathbf{x}),$$

where $\pi_1(\theta|\sigma^2, \mathbf{x}) = N(\mu(\mathbf{x}), \eta^2)$,

$$\mu(\mathbf{x}) = \frac{\mu + n\tau\bar{x}}{n\tau + 1}, \quad \eta^2 = \frac{\sigma^2}{n + \tau^{-1}}, \quad \bar{x} = \sum_{i=1}^n x_i,$$

and $\pi_2(\sigma^2|\mathbf{x})$ is $\Gamma^{-1}(\alpha + n/2, \tilde{\beta})$,

$$\tilde{\beta} = \left[\beta + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n(\bar{x} - \mu)^2}{2(1 + n\tau)} \right].$$

(2) The marginal posterior for σ^2 given \mathbf{x} is $\Gamma^{-1}(\alpha + n/2, \tilde{\beta})$.

(3) The marginal posterior for θ given \mathbf{x} is

$$\mathcal{T} \left(2\alpha + n, \mu(\mathbf{x}), \frac{\tilde{\beta}}{(\tau^{-1} + n)(\alpha + n/2)} \right).$$

(1) Since (\bar{X}, S^2) is the sufficient statistic of (θ_1, σ^2) , where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\theta_1, \frac{\sigma^2}{n}), \quad S^2 = \frac{1}{\nu} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \nu = n - 1, \quad \frac{\nu S^2}{\sigma^2} \sim \chi_\nu^2,$$

the likelihood is

$$\begin{aligned} f(t|\theta, \sigma^2) &= f_1(\bar{x}|\theta, \sigma^2)f_2(s^2|\sigma^2) \\ &= \sqrt{\frac{n}{2\pi\sigma^2}} \exp \left\{ -\frac{n}{2\sigma^2} (\bar{x} - \theta)^2 \right\} \frac{\left(\frac{\nu}{2\sigma^2}\right)^{\nu/2}}{\Gamma(\nu/2)} (s^2)^{\nu/2-1} e^{-\nu s^2/2\sigma^2} \\ &\propto \sigma^{-\nu-1} \exp \left\{ -\frac{1}{2\sigma^2} [\nu s^2 + n(\bar{x} - \theta)^2] \right\}. \end{aligned} \tag{64}$$

Therefore, the posterior is

$$\begin{aligned} \pi(\theta, \sigma^2|\mathbf{x}) &= \frac{f(t|\theta, \sigma^2)\pi_1(\theta|\sigma^2)\pi_2(\sigma^2)}{\int_{\mathbb{R}^+} \int_{\mathbb{R}} f(t|\theta, \sigma^2)\pi_1(\theta|\sigma^2)\pi_2(\sigma^2)d\theta d\sigma^2} \\ &\propto f(t|\theta, \sigma^2)\sigma^{-1} \exp \left\{ -\frac{(\theta - \mu)^2}{2\tau\sigma^2} \right\} (\sigma^2)^{-\alpha-1} \exp \left\{ -\frac{\beta}{\sigma^2} \right\} \\ &\propto (\sigma^2)^{-\nu/2-\alpha-2} \exp \left\{ -\frac{1}{2\sigma^2} \left[n(\bar{x} - \theta)^2 + \frac{(\theta - \mu)^2}{\tau} \right] \right\} \exp \left\{ -\frac{1}{\sigma^2} \left(\frac{\nu s^2}{2} + \beta \right) \right\}. \end{aligned} \tag{65}$$

In order to normalize the posterior, consider

$$\begin{aligned}
K &= \int_{\mathbb{R}^+} \int_{\mathbb{R}} (\sigma^2)^{-\nu/2-\alpha-2} \exp \left\{ -\frac{1}{2\sigma^2} \left[n(\bar{x} - \theta)^2 + \frac{(\theta - \mu)^2}{\tau} \right] \right\} \exp \left\{ -\frac{1}{\sigma^2} \left(\frac{\nu s^2}{2} + \beta \right) \right\} d\theta d\sigma^2 \\
&= \sqrt{\frac{2\pi}{n + \tau^{-1}}} \int_{\mathbb{R}^+} (\sigma^2)^{-\nu/2-\alpha-3/2} \exp \left\{ -\frac{D^*}{\sigma^2} \right\} d\sigma^2 \\
&= \sqrt{\frac{2\pi}{n + \tau^{-1}}} \frac{\Gamma(\nu/2 + \alpha + 1/2)}{D^{*(\nu/2+\alpha+1/2)}},
\end{aligned} \tag{66}$$

where

$$D^* = \frac{\nu s^2}{2} + \beta + \frac{1}{2} \left[n\bar{x}^2 + \mu^2 \tau^{-1} - \frac{(n\bar{x} + \mu \tau^{-1})^2}{n + \tau^{-1}} \right] = \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n(\bar{x} - \mu)^2}{2(1 + n\tau)}.$$

So the posterior is

$$\begin{aligned}
\pi(\theta, \sigma^2 | \mathbf{x}) &= K^{-1} (\sigma^2)^{-\nu/2-\alpha-2} \exp \left\{ -\frac{1}{2\sigma^2} \left[n(\bar{x} - \theta)^2 + \frac{(\theta - \mu)^2}{\tau} \right] \right\} \exp \left\{ -\frac{1}{\sigma^2} \left(\frac{\nu s^2}{2} + \beta \right) \right\} \\
&= K^{-1} (\sigma^2)^{-\nu/2-\alpha-2} \exp \left\{ -\frac{n + \tau^{-1}}{2\sigma^2} \left[\theta - \frac{n\bar{x} + \mu \tau^{-1}}{n + \tau^{-1}} \right]^2 \right\} \exp \left\{ -\frac{D^*}{\sigma^2} \right\}.
\end{aligned} \tag{67}$$

Of course the joint posterior can be written as

$$\pi(\theta, \sigma^2 | \mathbf{x}) = \pi_1(\theta | \sigma^2, \mathbf{x}) \pi_2(\sigma^2 | \mathbf{x}).$$

$\pi_1(\theta | \sigma^2, \mathbf{x})$ can be obtained by

$$\begin{aligned}
\pi_1(\theta | \sigma^2, \mathbf{x}) &= \frac{f(\bar{x} | \theta, \sigma^2) \pi_1(\theta | \sigma^2)}{\int_{\Theta} f(\bar{x} | \theta, \sigma^2) \pi_1(\theta | \sigma^2) d\theta} \\
&\propto \exp \left\{ -\frac{n(\bar{x} - \theta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{(\theta - \mu)^2}{2\tau\sigma^2} \right\} \\
&\propto \exp \left\{ -\frac{n + \tau^{-1}}{2\sigma^2} \left[\theta - \frac{n\bar{x} + \mu \tau^{-1}}{n + \tau^{-1}} \right]^2 \right\}.
\end{aligned} \tag{68}$$

As a result, $\pi_1(\theta | \sigma^2, \mathbf{x}) = N(\mu(\mathbf{x}), \eta^2)$,

$$\mu(\mathbf{x}) = \frac{\mu + n\tau\bar{x}}{n\tau + 1}, \quad \eta^2 = \frac{\sigma^2}{n + \tau^{-1}}, \quad \bar{x} = \sum_{i=1}^n x_i.$$

$\pi_2(\sigma^2 | \mathbf{x})$ can be obtained by

$$\begin{aligned}
\pi_2(\sigma^2 | \mathbf{x}) &= \frac{\pi(\theta, \sigma^2 | \mathbf{x})}{\pi_1(\theta | \sigma^2, \mathbf{x})} \\
&= \frac{K^{-1} (\sigma^2)^{-\nu/2-\alpha-2} \exp \left\{ -\frac{n + \tau^{-1}}{2\sigma^2} \left[\theta - \frac{n\bar{x} + \mu \tau^{-1}}{n + \tau^{-1}} \right]^2 \right\} \exp \left\{ -\frac{D^*}{\sigma^2} \right\}}{\sqrt{\frac{n + \tau^{-1}}{2\pi\sigma^2}} \exp \left\{ -\frac{n + \tau^{-1}}{2\sigma^2} \left[\theta - \frac{n\bar{x} + \mu \tau^{-1}}{n + \tau^{-1}} \right]^2 \right\}} \\
&= \frac{D^{*(n/2+\alpha)}}{\Gamma(n/2 + \alpha)} (\sigma^2)^{-n/2-\alpha-1} \exp \left\{ -\frac{D^*}{\sigma^2} \right\},
\end{aligned} \tag{69}$$

which is exactly $\pi_2(\sigma^2|\mathbf{x})$ is $\Gamma^{-1}(\alpha + n/2, \tilde{\beta})$,

$$\tilde{\beta} = \left[\beta + \frac{1}{2} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n(\bar{x} - \mu)^2}{2(1 + n\tau)} \right].$$

(2) The marginal posterior of σ^2 given \mathbf{x} is

$$\begin{aligned} \pi(\sigma^2|\mathbf{x}) &= \int_{\mathbb{R}} \pi(\theta, \sigma^2|\mathbf{x}) d\theta \\ &= K^{-1}(\sigma^2)^{-\nu/2-\alpha-2} \exp\left\{-\frac{D^*}{\sigma^2}\right\} \int_{\mathbb{R}} \exp\left\{-\frac{n + \tau^{-1}}{2\sigma^2} \left[\theta - \frac{n\bar{x} + \mu\tau^{-1}}{n + \tau^{-1}}\right]^2\right\} d\theta \\ &= \frac{D^{*(n/2+\alpha)}}{\Gamma(n/2 + \alpha)} (\sigma^2)^{-n/2-\alpha-1} \exp\left\{-\frac{D^*}{\sigma^2}\right\}, \quad (\tilde{\beta} = D^*) \end{aligned} \quad (70)$$

which is exactly $\Gamma^{-1}(\alpha + n/2, \tilde{\beta})$.

(3) The marginal posterior of θ given \mathbf{x} is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \int_{\mathbb{R}^+} \pi(\theta, \sigma^2|\mathbf{x}) \\ &= K^{-1} \int_{\mathbb{R}^+} (\sigma^2)^{-\nu/2-\alpha-2} \exp\left\{-\frac{D^*}{\sigma^2}\right\} \exp\left\{-\frac{n + \tau^{-1}}{2\sigma^2} \left[\theta - \frac{n\bar{x} + \mu\tau^{-1}}{n + \tau^{-1}}\right]^2\right\} d\sigma^2 \\ &= \frac{\Gamma(\alpha + n/2 + 1/2)}{\Gamma(\alpha + n/2) \sqrt{(2\alpha + n)\pi}} \sqrt{\frac{(n + \tau^{-1})(\alpha + n/2)}{D^*}} \\ &\quad \left\{ 1 + \frac{1}{2\alpha + n} \frac{(n + \tau^{-1})(\alpha + n/2)}{D^*} \left[\theta - \frac{n\bar{x} + \mu\tau^{-1}}{n + \tau^{-1}}\right]^2 \right\}^{-(n/2+1/2+\alpha)}, \end{aligned} \quad (71)$$

which is exactly

$$\mathcal{T}\left(2\alpha + n, \mu(\mathbf{x}), \frac{\tilde{\beta}}{(\tau^{-1} + n)(\alpha + n/2)}\right).$$

□

Ex. 3.12. *Cont'd Ex. 10 and Ex. 11(3). Find the posterior expectation estimator and the posterior mean square error.*

(1) **Ex. 10**

The posterior expectation estimator is

$$\hat{\theta}_E = x - \Sigma(\Sigma + A)^{-1}(x - \mu),$$

and the posterior MSE is

$$PMSE(\hat{\theta}_E) = V^\pi(x) = (\Sigma^{-1} + A^{-1})^{-1}.$$

(2) **Ex. 11(3)**

The posterior expectation estimator is

$$\hat{\theta}_E = \mu(\mathbf{x}) = \frac{\mu + n\tau\bar{x}}{n\tau + 1},$$

and the posterior MSE is

$$PMSE(\hat{\theta}_E) = V^\pi(\mathbf{x}) = \frac{2\alpha + n}{2\alpha + n - 2} \frac{\tilde{\beta}}{(\tau^{-1} + n)(\alpha + n/2)}.$$

□

Ex. 3.13. *Cont'd Ex. 2 and Ex. 6. Find the posterior mode estimator and the posterior mean square error.*

(1) **Ex. 2**

The posterior distribution is

$$\pi(\theta|\mathbf{x}) = \frac{n^{\bar{x}}}{\Gamma(n\bar{x})} \theta^{n\bar{x}-1} e^{-n\theta}.$$

As a result, according to the formula, the mode estimator is

$$\hat{\theta}_{MD} = \bar{x} - \frac{1}{n},$$

and the posterior MSE is

$$\begin{aligned} PMSE(\hat{\theta}_{MD}) &= V^\pi(\mathbf{x}) + [\mu^\pi(\mathbf{x}) - \hat{\theta}_{MD}]^2 \\ &= \frac{\bar{x}}{n} + \frac{1}{n^2}. \end{aligned} \tag{72}$$

(2) **Ex. 6**

The posterior distribution is

$$\pi(\theta|\mathbf{x}) = \frac{(\sum_{i=1}^n x_i + \beta)^{\alpha+n}}{\Gamma(\alpha+n)} \theta^{-(\alpha+n+1)} \exp \left\{ \theta^{-1} \left(\sum_{i=1}^n x_i + \beta \right) \right\}.$$

As a result, according to the formula, the mode estimator is

$$\hat{\theta}_{MD} = \frac{\sum_{i=1}^n x_i + \beta}{\alpha + n + 1},$$

and the posterior MSE is

$$\begin{aligned} PMSE(\hat{\theta}_{MD}) &= V^\pi(\mathbf{x}) + [\mu^\pi(\mathbf{x}) - \hat{\theta}_{MD}]^2 \\ &= \frac{(\sum_{i=1}^n x_i + \beta)^2}{(\alpha + n - 1)^2} \left[\frac{1}{\alpha + n - 2} - \frac{4}{(\alpha + n + 1)^2} \right]. \end{aligned} \tag{73}$$

□

Ex. 3.14. *Cont'd Ex. 1(2). Find the posterior median estimator and the posterior mean square error.*

Assume that $n = x = 1$. The posterior distribution is

$$\pi(\theta|x) = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^x(1-\theta)^{n-x}.$$

As a result, we have to find the median of this distribution. Using numerical methods (the qbeta function in R) we know

$$\hat{\theta}_{ME} = 0.7071068.$$

According to the formula, the posterior MSE is

$$\begin{aligned} PMSE(\hat{\theta}_{ME}) &= V^\pi(\mathbf{x}) + [\mu^\pi(\mathbf{x}) - \hat{\theta}_{ME}]^2 \\ &= \frac{2}{3^2 \cdot 4} + [2/3 - 0.7071068]^2 \\ &= 0.05719. \end{aligned} \tag{74}$$

□

Ex. 3.15. Suppose that X follows the distribution

$$P(X = x) = \theta(1 - \theta)^{x-1} \quad (x = 1, 2, \dots),$$

and the prior for θ is $U(0, 1)$. Show that

- (1) If only one X (3) is observed, find the posterior expectation estimator of θ .
- (2) If three X 's (3, 2, 5) are observed, find the posterior expectation estimator of θ .

The posterior is

$$\begin{aligned} \pi(\theta) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_0^1 \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta} \\ &= \frac{\theta^n(1-\theta)^{n\bar{x}-n}}{\int_0^1 \theta^n(1-\theta)^{n\bar{x}-n}d\theta} \\ &= \frac{\Gamma(n\bar{x}+2)}{\Gamma(n+1)\Gamma(n\bar{x}-n+1)}\theta^n(1-\theta)^{n\bar{x}-n}. \end{aligned} \tag{75}$$

As a result, the posterior expectation estimator is

$$\hat{\theta}_E = \frac{n+1}{n\bar{x}+2}.$$

- (1) With $n = 1$, $\bar{x} = 3$, the posterior expectation estimator is $2/5$.
- (2) With $n = 3$, $\bar{x} = 10/3$, the posterior expectation estimator is $1/3$. □

Ex. 3.16. Suppose that X follows the distribution $\text{Exp}(\lambda)$, where λ follows a Gamma distribution with the mean of 0.2 and the standard deviation of 1.0. For 20 observations the average $\bar{x} = 3.8$. Find the posterior expectation estimator for λ and λ^{-1} .

- (1) For the parameter λ , the prior is

$$\lambda \sim \Gamma(1/25, 1/5).$$

The posterior is

$$\begin{aligned}
\pi(\lambda|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\lambda)\pi(\lambda)}{\int_{\mathbb{R}^+} \prod_{i=1}^n f(x_i|\lambda)\pi(\lambda)d\lambda} \\
&= \frac{\lambda^n e^{-\lambda n\bar{x}} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_{\mathbb{R}^+} \lambda^n e^{-\lambda n\bar{x}} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda} \\
&= \frac{(n\bar{x} + \beta)^{\alpha+n}}{\Gamma(\alpha + n)} \lambda^{n+\alpha-1} e^{-(n\bar{x}+\beta)\lambda}.
\end{aligned} \tag{76}$$

As a result, the posterior expectation estimator is

$$\hat{\lambda}_E = \frac{n + \alpha}{n\bar{x} + \beta} = \frac{20 + 1/25}{20 \cdot 3.8 + 1/5} = 0.263.$$

(2) For the parameter $t = \lambda^{-1}$, the prior is

$$t \sim \Gamma^{-1}(1/25, 1/5).$$

The posterior is

$$\begin{aligned}
\pi(t|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|t)\pi(t)}{\int_{\mathbb{R}^+} \prod_{i=1}^n f(x_i|t)\pi(t)dt} \\
&= \frac{t^{-n} e^{-n\bar{x}/t} t^{-\alpha-1} e^{-\beta/t}}{\int_{\mathbb{R}^+} t^{-n} e^{-n\bar{x}/t} t^{-\alpha-1} e^{-\beta/t} dt} \\
&= \frac{(n\bar{x} + \beta)^{n+\alpha}}{\Gamma(n + \alpha)} t^{-(n+\alpha+1)} e^{-(n\bar{x}+\beta)/t}.
\end{aligned} \tag{77}$$

As a result, the posterior expectation estimator is

$$\hat{t}_E = \frac{n\bar{x} + \beta}{n + \alpha - 1} = \frac{20 \cdot 3.8 + 1/5}{20 + 1/25 - 1} = 4.002.$$

□

Ex. 3.17. Suppose that X follows the distribution $P(\lambda)$, where $\lambda \sim \Gamma(3, 1)$. Three X 's are observed: 2, 0, 6. Find the posterior expectation estimator and the posterior variance for λ .

The posterior distribution is

$$\begin{aligned}
\pi(\lambda|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\lambda)\pi(\lambda)}{\int_{\mathbb{R}^+} \prod_{i=1}^n f(x_i|\lambda)\pi(\lambda)d\lambda} \\
&= \frac{e^{-n\lambda} \lambda^{n\bar{x}} \lambda^{\alpha-1} e^{-\beta\lambda}}{\int_{\mathbb{R}^+} e^{-n\lambda} \lambda^{n\bar{x}} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda} \\
&= \frac{(n + \beta)^{n\bar{x}+\alpha}}{\Gamma(n\bar{x} + \alpha)} \lambda^{n\bar{x}+\alpha-1} e^{-(n+\beta)\lambda}.
\end{aligned} \tag{78}$$

Therefore, the posterior expectation estimator is

$$\hat{\lambda}_E = \frac{n\bar{x} + \alpha}{n + \beta} = \frac{8 + 3}{3 + 1} = \frac{11}{4},$$

and the posterior variance is

$$V^\pi(\mathbf{x}) = \frac{n\bar{x} + \alpha}{(n + \beta)^2} = \frac{11}{16}.$$

□

Ex. 3.18. Suppose that the prior for the defective rate θ is $Be(5, 10)$. Find the posterior mode estimator and the posterior expectation estimator, given the following observations in order:

- (1) The first batch: 3 out of 20 products are defective.
- (2) The second batch: 0 out of 20 products are defective.

(1) The posterior given the first batch is

$$\begin{aligned} \pi_1(\theta|\mathbf{x}^{(1)}) &= \frac{\prod_{i=1}^{20} f(x_i|\theta)\pi_0(\theta)}{\int_0^1 \prod_{i=1}^{20} f(x_i|\theta)\pi_0(\theta)d\theta} \\ &= \frac{\theta^3(1-\theta)^{17}\theta^4(1-\theta)^9}{\int_0^1 \theta^3(1-\theta)^{17}\theta^4(1-\theta)^9d\theta} \\ &= \frac{\Gamma(35)}{\Gamma(8)\Gamma(27)}\theta^7(1-\theta)^{26}. \end{aligned} \tag{79}$$

As a result,

$$\begin{aligned} \hat{\theta}_{MD} &= \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{7}{33}, \\ \hat{\theta}_E &= \frac{\alpha}{\alpha + \beta} = \frac{8}{35}. \end{aligned}$$

(2) The posterior given the second batch is

$$\begin{aligned} \pi_2(\theta|\mathbf{x}^{(2)}) &= \frac{\prod_{i=1}^{20} f(x_i|\theta)\pi_1(\theta|\mathbf{x}^{(1)})}{\int_0^1 \prod_{i=1}^{20} f(x_i|\theta)\pi_1(\theta|\mathbf{x}^{(1)})d\theta} \\ &= \frac{\theta^0(1-\theta)^{20}\theta^7(1-\theta)^{26}}{\int_0^1 \theta^0(1-\theta)^{20}\theta^7(1-\theta)^{26}d\theta} \\ &= \frac{\Gamma(55)}{\Gamma(8)\Gamma(47)}\theta^7(1-\theta)^{46}. \end{aligned} \tag{80}$$

As a result,

$$\begin{aligned} \hat{\theta}_{MD} &= \frac{\alpha - 1}{\alpha + \beta - 2} = \frac{7}{53}, \\ \hat{\theta}_E &= \frac{\alpha}{\alpha + \beta} = \frac{8}{55}. \end{aligned}$$

□

Ex. 3.19. Suppose that θ is the rate of people that approve the policy A in a city. There are two candidates for the prior for θ :

$$\begin{aligned} A : \pi_A(\theta) &= 2\theta \quad (0 < \theta < 1); \\ B : \pi_B(\theta) &= 4\theta^3 \quad (0 < \theta < 1). \end{aligned}$$

Assume that 1000 people are randomly sampled and 710 among them approve the policy A.

(1) Find the posterior for θ , with both priors respectively.

(2) Find the posterior expectation estimator for θ , with both priors respectively.

(3) Show that: with the sample size of 1000, the difference between the two posterior expectation estimators will be within 0.002, regardless of the number of people who approve the policy A.

(1) **For prior A:** the posterior is

$$\begin{aligned}\pi_A(\theta|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi_A(\theta)}{\int_0^1 \prod_{i=1}^n f(x_i|\theta)\pi_A(\theta)d\theta} \\ &= \frac{\theta^{710}(1-\theta)^{290}2\theta}{\int_0^1 \theta^{710}(1-\theta)^{290}2\theta d\theta} \\ &= \frac{\Gamma(1003)}{\Gamma(712)\Gamma(291)}\theta^{711}(1-\theta)^{290}.\end{aligned}\tag{81}$$

For prior B: the posterior is

$$\begin{aligned}\pi_B(\theta|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi_B(\theta)}{\int_0^1 \prod_{i=1}^n f(x_i|\theta)\pi_B(\theta)d\theta} \\ &= \frac{\theta^{710}(1-\theta)^{290}4\theta^3}{\int_0^1 \theta^{710}(1-\theta)^{290}4\theta^3 d\theta} \\ &= \frac{\Gamma(1005)}{\Gamma(714)\Gamma(291)}\theta^{713}(1-\theta)^{290}.\end{aligned}\tag{82}$$

(2) **For prior A:**

$$\hat{\theta}_E = \frac{\alpha}{\alpha + \beta} = \frac{712}{1003} = 0.70987.$$

For prior B:

$$\hat{\theta}_E = \frac{\alpha}{\alpha + \beta} = \frac{714}{1005} = 0.71045.$$

(3) With $n = 1000$, the posteriors are

$$\pi_A(\theta|\mathbf{x}) = \frac{\Gamma(1003)}{\Gamma(x+2)\Gamma(1001-x)}\theta^{x+1}(1-\theta)^{1000-x}$$

and

$$\pi_B(\theta|\mathbf{x}) = \frac{\Gamma(1005)}{\Gamma(x+4)\Gamma(1001-x)}\theta^{x+3}(1-\theta)^{1000-x}.$$

Therefore, the posterior expectation estimators are

$$\hat{\theta}_A = \frac{\alpha}{\alpha + \beta} = \frac{x+2}{1003},$$

and

$$\hat{\theta}_B = \frac{\alpha}{\alpha + \beta} = \frac{x+4}{1005}.$$

Since $x \in \{0, 1, \dots, 1000\}$, the difference is

$$|\hat{\theta}_A - \hat{\theta}_B| = \frac{2002 - 2x}{1008015} \leq \frac{2002}{1008015} < 0.002.$$

□

Ex. 3.20. An insurance company wants to set up a new program for car accident. Suppose that the number of car accidents per thousand people per year, X , follows a distribution $P(\lambda)$, where the prior for λ is $\Gamma(35, 1)$. According to a survey, 85 car accidents occurred among 2000 people last year. If the insurance company pays \$1000 for one accident on average, the advertisement cost is \$500,000 a year, the selling price is \$50 per person, and 100,000 people a year buy the insurance. Find the profit the company can make in a year.

Since $x = 42.5$, the posterior distribution is

$$\begin{aligned}\pi(\lambda|x) &= \frac{f(x|\lambda)\pi(\lambda)}{\int_{\mathbb{R}^+} f(x|\lambda)\pi(\lambda)d\lambda} \\ &= \frac{e^{-\lambda}\lambda^{42.5}\lambda^{34}e^{-\lambda}}{\int_{\mathbb{R}^+} e^{-\lambda}\lambda^{42.5}\lambda^{34}e^{-\lambda}d\lambda} \\ &= \frac{2^{77.5}}{\Gamma(77.5)}\lambda^{76.5}e^{-2\lambda}.\end{aligned}\tag{83}$$

Then the estimated number of car accidents is

$$\hat{\lambda}_E = \frac{\alpha}{\beta} = 38.75.$$

As a result, the profit is

$$100,000 \cdot \$50 - \$500,000 - 38.75 \cdot \$1000 = \$4,461,250.$$

□

Ex. 3.21. Suppose that

$$(X_1, \dots, X_5) \sim^{\text{i.i.d.}} \text{Exp}(1/\theta),$$

and $\theta \sim \Gamma^{-1}(10, 100)$. If the 5 observations are 5, 12, 14, 10, 12, find the posterior mode estimator and the posterior expectation estimator.

The posterior is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta} \\ &= \frac{\theta^{-5}e^{-53/\theta}\theta^{-11}e^{-100/\theta}}{\int_{\mathbb{R}^+} \theta^{-5}e^{-53/\theta}\theta^{-11}e^{-100/\theta}d\theta} \\ &= \frac{153^{15}}{\Gamma(15)}\theta^{-16}e^{-153/\theta}.\end{aligned}\tag{84}$$

As a result,

$$\hat{\theta}_{MD} = \frac{\beta}{\alpha + 1} = \frac{153}{16} = 9.5625,$$

and

$$\hat{\theta}_E = \frac{\beta}{\alpha - 1} = \frac{153}{14} = 10.9286.$$

□

Ex. 3.22. Suppose that

$$X \sim \Gamma\left(\frac{n}{2}, \frac{1}{2\theta}\right),$$

and the prior for θ is $\Gamma^{-1}(\alpha, \beta)$.

(1) Find the posterior expectation estimator and the posterior variance.

(2) If $\mathbf{X} = (X_1, \dots, X_n)$ are randomly sampled, find the posterior mode estimator and the posterior expectation estimator.

(1) The posterior is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{\theta^{-n/2}e^{-x/(2\theta)}\theta^{-\alpha-1}e^{-\beta/\theta}}{\int_{\mathbb{R}^+} \theta^{-n/2}e^{-x/(2\theta)}\theta^{-\alpha-1}e^{-\beta/\theta}d\theta} \\ &= \frac{(x/2 + \beta)^{n/2+\alpha}}{\Gamma(n/2 + \alpha)}\theta^{-n/2-\alpha-1}e^{-(x/2+\beta)/\theta}. \end{aligned} \tag{85}$$

Therefore,

$$\hat{\theta}_E = \frac{x/2 + \beta}{n/2 + \alpha - 1},$$

and

$$V^\pi(\theta) = \frac{(x/2 + \beta)^2}{(n/2 + \alpha - 1)^2(n/2 + \alpha - 2)}.$$

(2) The posterior is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} f(\mathbf{x}|\theta)\pi(\theta)d\theta} \\ &= \frac{\theta^{-n^2/2}e^{-n\bar{x}/(2\theta)}\theta^{-\alpha-1}e^{-\beta/\theta}}{\int_{\mathbb{R}^+} \theta^{-n^2/2}e^{-n\bar{x}/(2\theta)}\theta^{-\alpha-1}e^{-\beta/\theta}d\theta} \\ &= \frac{(n\bar{x}/2 + \beta)^{n^2/2+\alpha}}{\Gamma(n^2/2 + \alpha)}\theta^{-n^2/2-\alpha-1}e^{-(n\bar{x}/2+\beta)/\theta}. \end{aligned} \tag{86}$$

Therefore,

$$\hat{\theta}_{MD} = \frac{n\bar{x}/2 + \beta}{n^2/2 + \alpha + 1},$$

and

$$\hat{\theta}_E = \frac{n\bar{x}/2 + \beta}{n^2/2 + \alpha - 1}.$$

□

Ex. 3.23. Suppose that

$$\mathbf{X} = (X_1, \dots, X_r) \sim^{\text{i.i.d.}} M(r, \theta).$$

The prior for $\theta = (\theta_1, \dots, \theta_r)$ is $D(\alpha_1, \dots, \alpha_r)$. Find the posterior mode estimator and the posterior expectation estimator.

The posterior is

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\theta_1} \dots \int_{\theta_r} f(\mathbf{x}|\theta)\pi(\theta)d\theta_1\dots d\theta_r} \\
&= \frac{\theta_1^{x_1} \dots \theta_r^{x_r} \theta_1^{\alpha_1-1} \dots \theta_r^{\alpha_r-1}}{\int_{\theta_1} \dots \int_{\theta_r} \theta_1^{x_1} \dots \theta_r^{x_r} \theta_1^{\alpha_1-1} \dots \theta_r^{\alpha_r-1} d\theta_1 \dots d\theta_r} \\
&= \frac{\Gamma(\sum_{i=1}^r (x_i + \alpha_i))}{\prod_{i=1}^r \Gamma(x_i + \alpha_i)} \theta_1^{x_1 + \alpha_1 - 1} \dots \theta_r^{x_r + \alpha_r - 1}.
\end{aligned} \tag{87}$$

Therefore,

$$\hat{\theta}_{iMD} = \frac{x_i + \alpha_i - 1}{\sum_{j=1}^r (x_j + \alpha_j) - r}$$

and

$$\hat{\theta}_{iE} = \frac{x_i + \alpha_i}{\sum_{j=1}^r (x_j + \alpha_j)}.$$

□

Ex. 3.24. Three(3) observations (2, 4, 3) are obtained from the distribution $N(\theta, 1)$. The prior for θ is $N(3, 1)$. Find the 0.95 Bayes credible interval.

It is known that the posterior is

$$\pi(\theta|\mathbf{x}) \sim N(\mu_n, \eta_n^2)$$

where

$$\mu_n = \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x} = 3, \quad \eta_n^2 = \frac{\sigma^2 \tau^2}{n\tau^2 + \sigma^2} = 0.25.$$

Therefore, the 0.95 Bayes credible interval is

$$[3 - 1.96 \cdot 0.5, 3 + 1.96 \cdot 0.5] = [2.02, 3.98].$$

□

Ex. 3.25. Five(5) products are sampled from a batch. Suppose that the number of defective products, X , follows $B(5, \theta)$. The prior for θ is $Beta(1, 9)$. If $x = 0$, find the 95% HPD credible interval.

The posterior is

$$\begin{aligned}
\pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_0^1 f(x|\theta)\pi(\theta)d\theta} \\
&= \frac{\theta^0(1-\theta)^5 \theta^0(1-\theta)^8}{\int_0^1 \theta^0(1-\theta)^5 \theta^0(1-\theta)^8 d\theta} \\
&= \frac{\Gamma(15)}{\Gamma(1)\Gamma(14)} (1-\theta)^{13}.
\end{aligned} \tag{88}$$

Since the posterior density is not symmetric, we can use the `hpd` function in the TeachingDemos R package to find the HPD interval:

$$[0.0453, 0.1926].$$

□

Ex. 3.26. Suppose the random variable X follows $P(\theta)$, and the prior for θ is the non-informative $\pi(\theta) = \theta^{-1}I(\theta > 0)$. Find the 90% HPD credible interval using normal approximation.

The posterior is

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta} \\ &= \frac{e^{-n\theta}\theta^{n\bar{x}}\theta^{-1}}{\int_{\mathbb{R}^+} e^{-n\theta}\theta^{n\bar{x}}\theta^{-1}d\theta} \\ &= \frac{(n\theta)^{n\bar{x}}}{\Gamma(n\bar{x})}\theta^{n\bar{x}-1}e^{-n\theta}.\end{aligned}\tag{89}$$

Therefore

$$\mu^{\pi}(\mathbf{x}) = \frac{\bar{x}}{\theta}, \quad V^{\pi}(\mathbf{x}) = \frac{\bar{x}}{n\theta^2}.$$

The normal approximation of the 90% HPD credible interval is

$$\left[\frac{\bar{x}}{\theta} - 1.64 \cdot \sqrt{\frac{\bar{x}}{n\theta^2}}, \frac{\bar{x}}{\theta} + 1.64 \cdot \sqrt{\frac{\bar{x}}{n\theta^2}} \right].$$

□

Ex. 3.27. Five(5) observations (1.2, 1.6, 1.3, 1.4, 1.4) are obtained from the distribution $N(\theta, \sigma^2)$. The prior for (θ, σ^2) is the non-informative prior $\pi(\theta, \sigma^2) = \sigma^{-1}I(\sigma^2 > 0)$. Find the 90% HPD credible interval for θ .

It is known that the posterior for θ is

$$\mathcal{T}\left(n-1, \bar{x}, \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}\right),$$

which is

$$\mathcal{T}(4, 1.38, 0.0044).$$

Therefore,

$$\mu^{\pi}(\mathbf{x}) = 1.38, \quad V^{\pi}(\mathbf{x}) = \frac{\nu}{\nu-2}\tau^2 = 0.0088.$$

The 90% HPD credible interval is

$$\left[\mu^{\pi}(\mathbf{x}) - 2.13 \cdot \sqrt{V^{\pi}(\mathbf{x})}, \mu^{\pi}(\mathbf{x}) + 2.13 \cdot \sqrt{V^{\pi}(\mathbf{x})} \right],$$

which is

$$[1.1802, 1.5798].$$

□

Ex. 3.28. Suppose that $X \sim N(\theta, 1)$. The observation is $x = 6$.

- (1) If the prior is $N(0, 2.19)$, find the 90% HPD credible interval for θ .
- (2) If the prior is $C(0, 1)$, find the 95% HPD credible interval for θ .

(1) It is known that the posterior is

$$\pi(\theta|\mathbf{x}) \sim N(\mu(x), \eta^2)$$

where

$$\mu(x) = \frac{\sigma^2}{\sigma^2 + \tau^2}\mu + \frac{\tau^2}{\sigma^2 + \tau^2}x = 4.119, \quad \eta^2 = \frac{\sigma^2\tau^2}{\tau^2 + \sigma^2} = 0.68652.$$

Therefore, the 90% HPD credible interval is

$$[4.119 - 1.64 \cdot 0.829, 4.119 + 1.64 \cdot 0.829] = [2.759, 5.479].$$

(2) The posterior distribution is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}} f(x|\theta)\pi(\theta)d\theta} \\ &\propto \exp\left\{-\frac{1}{2}(x-\theta)^2\right\} \frac{1}{1+\theta^2}. \end{aligned} \quad (90)$$

Since this is not easy to estimate analytically, we should resort to computational packages and Monte-Carlo methods. Details can be found here with reproducible codes.

The 95% HPD credible interval is [3.63, 7.68]. \square

Ex. 3.29. *Ex. 21 Cont'd. Find the 95% HPD credible set using normal approximation.*

Since the posterior is

$$\pi(\theta|\mathbf{x}) = \frac{153^{15}}{\Gamma(15)}\theta^{-16}e^{-153/\theta},$$

we known

$$\mu^\pi(\mathbf{x}) = \frac{153}{15-1} = 10.929, \quad V^\pi(\mathbf{x}) = \frac{153^2}{(15-1)^2(15-2)} = 9.1872.$$

The normal approximation of the 95% HPD credible interval is

$$[10.929 - 1.96 \cdot 3.031, 10.929 + 1.96 \cdot 3.031] = [4.988, 16.870].$$

\square

Ex. 3.30. *Suppose that $Q^2/\sigma^2 \sim \chi^2(n)$, and the prior for σ is $\pi(\sigma) = \sigma^{-1}$. Show that:*

- (1) *The non-informative prior for σ^2 is $\pi^*(\sigma^2) = \sigma^{-2}$;*
- (2) *If $n = 2$, $Q^2 = 2$, the 95% HPD credible intervals for σ and σ^2 are different.*

(1) Since

$$\frac{Q^2}{\sigma^2} \sim \chi^2(n) = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right),$$

we know

$$Q^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right).$$

Therefore, the distribution of Q^2 can be written as $(2\sigma^2)^{-1}\phi(\frac{t}{2\sigma^2})$, where

$$\phi(t) = \frac{1}{\Gamma(n/2)}t^{n/2-1}e^{-t},$$

indicating that the non-informative prior for σ^2 is $\pi^*(\sigma^2) = \sigma^{-2}$.

(2) The posterior for σ is

$$\begin{aligned}
\pi(\sigma|t) &= \frac{f(t|\sigma)\pi(\sigma)}{\int_{\mathbb{R}^+} f(t|\sigma)\pi(\sigma)d\sigma} \\
&= \frac{\sigma^{-n}e^{-t/(2\sigma^2)}\sigma^{-1}}{\int_{\mathbb{R}^+} \sigma^{-n}e^{-t/(2\sigma^2)}\sigma^{-1}d\sigma} \\
&= \frac{t^{n/2}}{2^{n/2-1}\Gamma(n/2)}\sigma^{-n-1}e^{-\frac{t}{2\sigma^2}} \\
&= 2\sigma^{-3}e^{-\frac{1}{\sigma^2}}.
\end{aligned} \tag{91}$$

While the posterior for σ^2 is

$$\begin{aligned}
\pi(\sigma^2|t) &= \frac{f(t|\sigma^2)\pi(\sigma^2)}{\int_{\mathbb{R}^+} f(t|\sigma^2)\pi(\sigma^2)d\sigma^2} \\
&= \frac{(\sigma^2)^{-n/2}e^{-t/(2\sigma^2)}(\sigma^2)^{-1}}{\int_{\mathbb{R}^+} (\sigma^2)^{-n/2}e^{-t/(2\sigma^2)}(\sigma^2)^{-1}d\sigma^2} \\
&= \frac{(t/2)^{n/2}}{\Gamma(n/2)}(\sigma^2)^{-n/2-1}e^{-\frac{t}{2\sigma^2}} \\
&= \sigma^{-4}e^{-1/\sigma^2}.
\end{aligned} \tag{92}$$

Since the HPD intervals are not easy to find analytically, we have to resort to statistical software packages and Monte-Carlo methods. Details can be found here with reproducible codes.

Ex. 3.31. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} N(0, \sigma^2),$$

and the prior for σ^2 is $\Gamma^{-1}(\alpha, \lambda)$. Find the 0.90 Bayes upper credible limit for σ^2 .

It is known that the posterior is

$$\pi(\sigma^2|t) = \frac{(t/2 + \lambda)^{n/2+\alpha}}{\Gamma(n/2 + \alpha)}(\sigma^2)^{-n/2-\alpha-1} \exp\left\{-\frac{t/2 + \lambda}{\sigma^2}\right\},$$

where

$$t = \sum_{i=1}^n x_i^2.$$

Therefore, the 0.90 Bayes upper credible limit, $\hat{\sigma}_U^2(t)$, is the 90% quantile of the distribution $\Gamma^{-1}(n/2 + \alpha, t/2 + \lambda)$. \square

Ex. 3.32. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} U(0, \theta),$$

and the prior for θ is $Pa(\theta_0, \alpha)$. Find the $1 - \alpha$ Bayes upper credible limit for θ .

It is known that the posterior is

$$\pi(\theta|\mathbf{x}) = \begin{cases} (n + \alpha)K^{n+\alpha}/\theta^{n+\alpha+1}, & \theta \geq K, \\ 0, & \text{otherwise,} \end{cases}$$

where $K = \max\{\theta_0, x_1, \dots, x_n\}$. Therefore, the $1 - \alpha$ credible limit for θ is the $1 - \alpha$ quantile of the distribution $Pa(\max\{\theta_0, x_1, \dots, x_n\}, n + \alpha)$. \square

Ex. 3.33. *Cont'd Ex. 25. Consider the hypothesis test*

$$H_0 : \theta \leq 0.1 \leftrightarrow H_1 : \theta > 0.1.$$

Find the posterior probabilities of the hypotheses, the posterior odds ratio, and the Bayes factor.

The prior is

$$\pi(\theta) = \frac{\Gamma(10)}{\Gamma(1)\Gamma(9)}(1-\theta)^8$$

and the posterior is

$$\pi(\theta|x) = \frac{\Gamma(15)}{\Gamma(1)\Gamma(14)}(1-\theta)^{13}.$$

Therefore,

$$\alpha_0 = \Pr(\theta \leq 0.1|x) = \int_0^{0.1} \pi(\theta|x)d\theta = 0.7712321,$$

$$\alpha_1 = \Pr(\theta > 0.1|x) = \int_{0.1}^1 \pi(\theta|x)d\theta = 0.2287679,$$

$$\frac{\alpha_0}{\alpha_1} = 3.371243,$$

$$\frac{\pi_0}{\pi_1} = \frac{\Pr(\theta \leq 0.1)}{\Pr(\theta > 0.1)} = \frac{0.6125795}{0.3874205} = 1.581175,$$

$$B^\pi(x) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = 2.132113.$$

□

Ex. 3.34. *Cont'd Ex. 27. Consider the hypothesis test*

$$H_0 : \theta \leq 0.1 \leftrightarrow H_1 : \theta > 0.1.$$

Find the posterior probabilities of the hypotheses, and the posterior odds ratio.

The posterior is

$$\mathcal{T}\left(n-1, \bar{x}, \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}\right) = \mathcal{T}(4, 1.38, 0.0044)$$

Therefore,

$$\alpha_0 = \Pr(\theta \leq 0.1|\mathbf{x}) = \int_{-\infty}^{0.1} \pi(\theta|\mathbf{x})d\theta \approx 0,$$

$$\alpha_1 = \Pr(\theta > 0.1|\mathbf{x}) = \int_{0.1}^{+\infty} \pi(\theta|\mathbf{x})d\theta \approx 1,$$

$$\frac{\alpha_0}{\alpha_1} = 0.$$

□

Ex. 3.35. Suppose that

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta, \\ 0, & x \leq \theta, \end{cases}$$

and

$$\pi(\theta) = \begin{cases} e^{-\theta}, & \theta > 0, \\ 0, & \theta \leq 0. \end{cases}$$

Do the hypothesis test:

$$H_0 : \theta \leq 1 \leftrightarrow H_1 : \theta > 1.$$

The posterior is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{e^{-(x-\theta)}e^{-\theta}I(0 < \theta < x)}{\int_0^x e^{-(x-\theta)}e^{-\theta}d\theta} \\ &= \frac{1}{x}I(0 < \theta < x). \end{aligned} \tag{93}$$

Therefore, if $\Pr(\theta \leq 1|x) \leq \Pr(\theta > 1|x)$, then accept the null hypothesis; otherwise reject it. \square

Ex. 3.36. Suppose that

$$X_1, \dots, X_m \sim^{\text{i.i.d.}} N(a, 1), \quad Y_1, \dots, Y_n \sim^{\text{i.i.d.}} N(b, 1),$$

and X 's and Y 's independent. Assume that a and b are independent, and that $a \sim N(\mu_1, \tau_1^2)$, $b \sim N(\mu_2, \tau_2^2)$. Do the hypothesis test:

$$H_0 : a - b \leq 0 \leftrightarrow H_1 : a - b > 0.$$

It is known that the posterior for a and b are

$$a|X_1, \dots, X_m \sim N(\mu_a, \eta_a^2), \quad \mu_a = \frac{1/m}{1/m + \tau_1^2}\mu_1 + \frac{\tau_1^2}{1/m + \tau_1^2}\bar{x}_m, \quad \eta_a^2 = \frac{\tau_1^2}{m\tau_1^2 + 1}$$

and

$$b|Y_1, \dots, Y_n \sim N(\mu_b, \eta_b^2), \quad \mu_b = \frac{1/n}{1/n + \tau_2^2}\mu_2 + \frac{\tau_2^2}{1/n + \tau_2^2}\bar{y}_n, \quad \eta_b^2 = \frac{\tau_2^2}{n\tau_2^2 + 1}.$$

Therefore, let $s = a - b$, then the difference between a and b has the distribution

$$\pi(s|\mathbf{x}) = N(\mu_a - \mu_b, \eta_a^2 + \eta_b^2).$$

Consider

$$\Pr(H_0|\mathbf{x}) = \Pr(s \leq 0|\mathbf{x}) = \Pr\left(\frac{s - (\mu_a - \mu_b)}{\sqrt{\eta_a^2 + \eta_b^2}} \leq \frac{\mu_b - \mu_a}{\sqrt{\eta_a^2 + \eta_b^2}}\right) = \Phi\left(\frac{\mu_b - \mu_a}{\sqrt{\eta_a^2 + \eta_b^2}}\right).$$

As a result, if $\mu_b - \mu_a > 0$, accept the null hypothesis; otherwise, reject it. \square

Ex. 3.37. Suppose that $X \sim N(\theta, 1)$, and five(5) observations are 4.9, 5.6, 5.1, 4.6, 3.6. The prior is

$$\pi(\theta) = \begin{cases} 0.5, & \theta = 4.01, \\ 0.5N(4.01, 1), & \text{otherwise.} \end{cases}$$

Do the hypothesis test:

$$H_0 : \theta = 4.01 \leftrightarrow H_1 : \theta \neq 4.01.$$

The likelihood is

$$f(\mathbf{x}|\theta) = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2 \right\} = \frac{1}{(2\pi)^{5/2}} \exp \left\{ -\frac{1}{2} (5\theta^2 - 47.6\theta + 115.5) \right\},$$

and

$$\begin{aligned} m_1(\mathbf{x}) &= \int_{\theta \neq \theta_0} f(\mathbf{x}|\theta) g_1(\theta) d\theta \\ &= \int_{\theta \neq 4.01} \frac{1}{(2\pi)^{5/2}} \exp \left\{ -\frac{1}{2} (5\theta^2 - 47.6\theta + 115.5) \right\} \frac{1}{(2\pi)^{1/2}} \exp \left\{ -\frac{1}{2} (\theta - 4.01)^2 \right\} d\theta \\ &= \frac{1}{(2\pi)^3} \exp \{-1.340375\} \int_{\theta \neq 4.01} \exp \left\{ -\frac{6}{2} (\theta - 4.635)^2 \right\} \\ &= \frac{1}{(2\pi)^3} \exp \{-1.340375\} \sqrt{\frac{3}{\pi}} \\ &= 0.001031166. \end{aligned} \tag{94}$$

Therefore,

$$B^\pi(\mathbf{x}) = \frac{f(\mathbf{x}|\theta_0)}{m_1(\mathbf{x})} = \frac{0.0008193963}{0.001031166} = 0.7946308 < 1.$$

H_0 should be rejected. □

Ex. 3.38. Suppose that $X \sim U(0, \theta)$, and the prior for θ is $Pa(5, 3)$. The five(5) observations are 10, 3, 2, 5, 14. Do the hypothesis test

$$H_0 : 0 \leq \theta \leq 15 \leftrightarrow H_1 : \theta > 15.$$

Find the posterior probabilities, the posterior odds ratio, and the Bayes factor.

It is known that the posterior is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \begin{cases} (n + \alpha)K^{n+\alpha}/\theta^{n+\alpha+1}, & \theta \geq \max\{\theta_0, x_1, \dots, x_n\}, \\ 0, & \text{otherwise,} \end{cases} \\ &= \begin{cases} 8 \cdot 14^8/\theta^9, & \theta \geq 14, \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{95}$$

Therefore,

$$\alpha_0 = \Pr(\Theta_0|\mathbf{x}) = \Pr(0 \leq \theta \leq 15) = 8 \cdot 14^8 \int_{14}^{15} \frac{1}{\theta^9} d\theta = 0.4241701,$$

$$\alpha_1 = \Pr(\Theta_1|\mathbf{x}) = \Pr(\theta > 15) = 8 \cdot 14^8 \int_{15}^{+\infty} \frac{1}{\theta^9} d\theta = 0.5758299,$$

$$\frac{\alpha_0}{\alpha_1} = 0.736624,$$

$$\pi_0 = 3 \cdot 5^3 \int_5^{15} \frac{1}{\theta^4} d\theta = 0.962963,$$

$$\pi_1 = 3 \cdot 5^3 \int_{15}^{+\infty} \frac{1}{\theta^4} d\theta = 0.037037,$$

$$B^\pi(\mathbf{x}) = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1} = 0.02833166.$$

The Bayes factor strongly supports the alternative hypothesis, so we should accept H_1 . \square

4 Bayesian Decision Theory

Ex. 4.1. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} N(\theta, 1)$$

where the prior for θ is $N(0, 1)$. Find the Bayes estimator under the square error loss.

The posterior $\pi(\theta|\mathbf{x})$ is $N(\mu(\mathbf{x}), \eta_n^2)$ where

$$\mu(\mathbf{x}) = \frac{1}{1/n + 1} \bar{x} = \frac{n}{1 + n} \bar{x}, \quad \eta_n^2 = \frac{1/n}{1/n + 1} = \frac{1}{1 + n}.$$

Therefore, the Bayes estimator is

$$\hat{\theta}_E = \frac{n}{1 + n} \bar{x}.$$

□

Ex. 4.2. Suppose that X follows the distribution

$$P(X = x) = \theta(1 - \theta)^{x-1} \quad (x = 1, 2, \dots),$$

and the prior for θ is $Be(\alpha, \beta)$. Find the Bayes estimator under the square error loss.

The posterior is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_0^1 f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{\theta(1 - \theta)^{x-1}\theta^{\alpha-1}(1 - \theta)^{\beta-1}}{\int_0^1 \theta(1 - \theta)^{x-1}\theta^{\alpha-1}(1 - \theta)^{\beta-1}d\theta} \\ &= \frac{\Gamma(\alpha + \beta + x)}{\Gamma(\alpha + 1)\Gamma(\beta + x)} \theta^\alpha (1 - \theta)^{x+\beta-2}. \end{aligned} \tag{96}$$

Therefore, the Bayes estimator is

$$\hat{\theta}_E = \frac{\alpha + 1}{\alpha + \beta + x}.$$

□

Ex. 4.3. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} P(\theta)$$

where the prior for θ is $Exp(\lambda)$. Find the Bayes estimator under the square error loss.

The posterior is

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta} \\ &= \frac{e^{-n\theta}\theta^{n\bar{x}}\lambda e^{-\lambda\theta}}{\int_{\mathbb{R}^+} e^{-n\theta}\theta^{n\bar{x}}\lambda e^{-\lambda\theta}d\theta} \\ &= \frac{(n + \lambda)^{n\bar{x}+1}}{\Gamma(n\bar{x} + 1)} \theta^{n\bar{x}} e^{-(n+\lambda)\theta}. \end{aligned} \tag{97}$$

Therefore, the Bayes estimator is

$$\hat{\theta}_E = \frac{n\bar{x} + 1}{n + \lambda}.$$

□

Ex. 4.4. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} \text{Exp}(\theta),$$

where the prior for θ is $\Gamma(\alpha, \beta)$. Find the Bayes estimator for θ and $1/\theta$ under the square error loss.

The posterior is

$$\begin{aligned} \pi_\theta(\theta|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta} \\ &= \frac{\theta^n e^{-n\theta\bar{x}} \theta^{\alpha-1} e^{-\beta\theta}}{\int_{\mathbb{R}^+} \theta^n e^{-n\theta\bar{x}} \theta^{\alpha-1} e^{-\beta\theta} d\theta} \\ &= \frac{(n\bar{x} + \beta)^{n+\alpha}}{\Gamma(n+\alpha)} \theta^{n+\alpha-1} e^{-(n\bar{x}+\beta)\theta}. \end{aligned} \quad (98)$$

Therefore, the Bayes estimator for θ is

$$\hat{\theta}_E = \frac{n + \alpha}{n\bar{x} + \beta}.$$

As for the Bayes estimator for $1/\theta$, we have to find the posterior for $t = 1/\theta$:

$$\pi_t(t|\mathbf{x}) = \pi_\theta(1/t|\mathbf{x})/t^2 = \frac{(n\bar{x} + \beta)^{n+\alpha}}{\Gamma(n+\alpha)} t^{-(n+\alpha+1)} e^{-\frac{(n\bar{x}+\beta)}{t}}.$$

Therefore, the Bayes estimator is

$$\left(\frac{1}{\theta}\right)_E = \frac{n\bar{x} + \beta}{n + \alpha - 1}.$$

□

Ex. 4.5. Suppose that

$$X_1, \dots, X_n \sim^{\text{i.i.d.}} U(0, \theta),$$

where the prior for θ is $U(0, a)$, $a > 0$. Find the Bayes estimator for θ under the square error loss.

The posterior is

$$\begin{aligned} \pi_\theta(\theta|\mathbf{x}) &= \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta} \\ &= \frac{1/\theta^n I(0 < x_{\max} < \theta) 1/a I(0 < \theta < a)}{\int_{\Theta} 1/\theta^n I(0 < x_{\max} < \theta) 1/a I(0 < \theta < a) d\theta}. \end{aligned} \quad (99)$$

When $n = 1$,

$$\pi_\theta(\theta|\mathbf{x}) = \begin{cases} \log \frac{x_{\max}}{a} \cdot \frac{1}{\theta}, & x_{\max} < \theta < a, \\ 0, & \text{otherwise.} \end{cases} \quad (100)$$

When $n > 1$

$$\pi_\theta(\theta|\mathbf{x}) = \begin{cases} \frac{n-1}{x_{\max}^{1-n} - a^{1-n}} \cdot \frac{1}{\theta^n}, & x_{\max} < \theta < a, \\ 0, & \text{otherwise.} \end{cases} \quad (101)$$

Therefore, the Bayes estimator is

$$\hat{\theta}_E = \begin{cases} (a - x_{\max}) \log \frac{x_{\max}}{a}, & n = 1, \\ \frac{n-1}{n-2} \cdot \frac{x_{\max}^{2-n} - a^{2-n}}{x_{\max} - a^{1-n}}, & n > 1. \end{cases}$$

□

Ex. 4.6. Suppose that

$$X_1, \dots, X_n \sim \text{i.i.d. } N(0, \tau),$$

where the prior for τ is $\Gamma^{-1}(\alpha, \beta)$. Find the Bayes estimator for τ under the weighted square error loss $L(\tau, \hat{\tau}) = (\tau - \hat{\tau})/\tau^2$.

It is known that the posterior, $\pi(\tau|\mathbf{x})$, is $\Gamma^{-1}(n/2 + \alpha, t + \beta)$ where

$$t = \frac{1}{2} \sum_{i=1}^n x_i^2.$$

Therefore, the Bayes estimator is

$$\hat{\tau}_B = \frac{E(\tau w(\tau)|\mathbf{x})}{E(w(\tau)|\mathbf{x})}, \quad (102)$$

where $w(\tau) = \tau^{-2}$, and

$$\begin{aligned} E(\tau w(\tau)|\mathbf{x}) &= \int_{\mathbb{R}^+} \tau^{-1} \pi(\tau|\mathbf{x}) d\tau = \frac{n/2 + \alpha}{t + \beta}, \\ E(w(\tau)|\mathbf{x}) &= \int_{\mathbb{R}^+} \tau^{-2} \pi(\tau|\mathbf{x}) d\tau = \frac{(n/2 + \alpha + 1)(n/2 + \alpha)}{(t + \beta)^2}. \end{aligned}$$

As a result, the Bayes estimator is

$$\hat{\tau}_B = \frac{t + \beta}{n/2 + \alpha + 1}.$$

□

Ex. 4.7. Suppose that $X \sim B(n, \theta)$, and $\theta \sim Be(\alpha, \beta)$. Find the Bayes estimator for θ under the weighted square error loss $L(a, \theta) = (a - \theta)^2/[\theta(1 - \theta)]$.

It is known that the posterior is $Be(\alpha + x, n - x + \beta)$. Therefore, the Bayes estimator is

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)},$$

where $w(\theta) = 1/[\theta(1 - \theta)]$, and

$$\begin{aligned} E(\theta w(\theta)|x) &= \int_0^1 (1 - \theta)^{-1} \pi(\theta|x) d\theta = \frac{n + \alpha + \beta - 1}{n - x + \beta - 1}, \\ E(w(\theta)|x) &= \int_0^1 \theta^{-1} (1 - \theta)^{-1} \pi(\theta|x) d\theta = \frac{(n + \alpha + \beta - 1)(n + \alpha + \beta - 2)}{(x + \alpha - 1)(n - x + \beta - 1)}. \end{aligned}$$

As a result, the Bayes estimator is

$$\hat{\theta}_B = \frac{x + \alpha - 1}{n + \alpha + \beta - 2}.$$

□

Ex. 4.8. Suppose that $X \sim \Gamma(n/2, (2\theta)^{-1})$, and $\theta \sim \Gamma^{-1}(\alpha, \beta/2)$. Find the Bayes estimator for θ under the weighted square error loss $L(a, \theta) = (a - \theta)^2/\theta^2$.

It is known that the posterior, $\pi(\theta|x)$, is $\Gamma^{-1}(n/2 + \alpha, (x + \beta)/2)$. Therefore, the Bayes estimator is

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)},$$

where $w(\theta) = \theta^{-2}$, and

$$\begin{aligned} E(\theta w(\theta)|x) &= \int_{\mathbb{R}^+} \theta^{-1} \pi(\theta|x) d\theta = \frac{n + 2\alpha}{x + \beta}, \\ E(w(\theta)|x) &= \int_{\mathbb{R}^+} \theta^{-2} \pi(\theta|x) d\theta = \frac{(n + 2\alpha + 2)(n + 2\alpha)}{(x + \beta)^2}. \end{aligned}$$

As a result, the Bayes estimator is

$$\hat{\theta}_B = \frac{x + \beta}{n + 2\alpha + 2}.$$

□

Ex. 4.9. Suppose that θ, X and a are all real numbers, and the posterior $\pi(\theta|x)$ is symmetric and uni-modal. The loss function $L(\theta - a)$ is the increasing function of $|\theta - a|$. Show that the Bayes estimator is the mode of $\pi(\theta|x)$.

First, notice that for a symmetric and uni-modal distribution, the mode is exactly the median. Let the mode and median of $\pi(\theta|x)$ be M .

Next, let the loss function be

$$L(\theta - a) = g(|\theta - a|),$$

where $g(\cdot)$ is an differentiable increasing function.

As a result, the posterior risk can be expressed as

$$R(a|x) = \int_{\mathbb{R}^+} L(\theta, a) \pi(\theta|x) d\theta = \int_{-\infty}^a g(a - \theta) \pi(\theta|x) d\theta + \int_a^{+\infty} g(\theta - a) \pi(\theta|x) d\theta.$$

Take the first derivative of $R(a|x)$ w.r.t. a ,

$$\frac{\partial R(a|x)}{\partial a} = \int_{-\infty}^a g'(a - \theta) \pi(\theta|x) d\theta + g(0) \pi(a|x) - \int_{-\infty}^a g'(a - \theta) \pi(\theta|x) d\theta - g(0) \pi(a|x) = 0,$$

we have

$$\int_{-\infty}^a g'(a - \theta) \pi(\theta|x) d\theta = \int_a^{+\infty} g'(\theta - a) \pi(\theta|x) d\theta.$$

Now we want to find the a that satisfies the equation above. Without losing of generality, let $a = M$. Since $\forall \theta \in (-\infty, M)$,

$$g'(M - \theta) = g'((2M - \theta) - M)$$

and

$$\pi(\theta|x) = \pi(2M - \theta|x)$$

where $2M - \theta \in (M, +\infty)$, and the mapping from $(-\infty, M)$ to $(M, +\infty)$

$$\theta \mapsto 2M - \theta$$

is bijective, we can conclude that

$$\int_{-\infty}^M g'(M - \theta) \pi(\theta|x) d\theta = \int_M^{+\infty} g'(\theta - M) \pi(\theta|x) d\theta.$$

Therefore, M is the Bayes estimator that minimizes the posterior risk. \square

Ex. 4.10. Suppose that the random variable X follows the distribution

$$P(X = x|p) = C(x + r - 1, r - 1) p^x (1 - p)^r, \quad x = 0, 1, 2, \dots; \quad 0 < p < 1.$$

Let $\theta = p/(1 - p)$, and the loss function be $L(d, \theta) = (d - \theta)^2 / [\theta(1 + \theta)]$:

(1) If $\pi(\theta) \equiv 1$, find the Bayes estimator using the loss function above, and compare it with the MLE.

(2) If the prior for θ is

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 + \theta)^{-(\alpha+\beta)},$$

find the posterior for θ , and the Bayes estimator.

The likelihood is

$$f(x|\theta) = C(x + r - 1, r - 1) \theta^x (1 + \theta)^{-(r+x)}.$$

(1) The posterior is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{\theta^x (1 + \theta)^{-(r+x)}}{\int_{\mathbb{R}^+} \theta^x (1 + \theta)^{-(r+x)} d\theta} \\ &= \frac{\Gamma(r + x)}{\Gamma(x + 1)\Gamma(r - 1)} \theta^x (1 + \theta)^{-(r+x)}. \end{aligned} \tag{103}$$

Since the loss function is a weighted square error loss, with $w(\theta) = [\theta(1 + \theta)]^{-1}$, we have

$$\begin{aligned} E(\theta w(\theta)|x) &= \int_{\mathbb{R}^+} (1 + \theta)^{-1} \pi(\theta|x) d\theta = \frac{r - 1}{r + x}, \\ E(w(\theta)|x) &= \int_{\mathbb{R}^+} [\theta(1 + \theta)]^{-1} \pi(\theta|x) d\theta = \frac{r(r - 1)}{(r + x)x}, \end{aligned}$$

and as a result,

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)} = \frac{x}{r}.$$

On the other hand, consider the log-likelihood

$$\log f(x|\theta) = \log C(x + r - 1, r - 1) + x \log \theta - (r + x) \log(1 + \theta).$$

Take the first derivative of the log-likelihood w.r.t. θ , and set it to 0:

$$\frac{x}{\theta} = \frac{r+x}{1+\theta},$$

we have

$$\hat{\theta}_{MLE} = \frac{x}{r},$$

which is the same as the Bayes estimator.

(2) The posterior is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{\theta^x(1+\theta)^{-(r+x)}\theta^{\alpha-1}(1+\theta)^{-(\alpha+\beta)}}{\int_{\mathbb{R}^+} \theta^x(1+\theta)^{-(r+x)}\theta^{\alpha-1}(1+\theta)^{-(\alpha+\beta)}d\theta} \\ &= \frac{\Gamma(r+x+\alpha+\beta)}{\Gamma(x+\alpha)\Gamma(r+\beta)}\theta^{x+\alpha-1}(1+\theta)^{-(r+x+\alpha+\beta)}. \end{aligned} \tag{104}$$

With the weighted square error loss and $w(\theta) = [\theta(1+\theta)]^{-1}$, we have

$$\begin{aligned} E(\theta w(\theta)|x) &= \int_{\mathbb{R}^+} (1+\theta)^{-1}\pi(\theta|x)d\theta = \frac{r+\beta}{r+x+\alpha+\beta}, \\ E(w(\theta)|x) &= \int_{\mathbb{R}^+} [\theta(1+\theta)]^{-1}\pi(\theta|x)d\theta = \frac{(r+\beta+1)(r+\beta)}{(r+x+\alpha+\beta)(x+\alpha-1)}, \end{aligned}$$

and as a result,

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)} = \frac{x+\alpha-1}{r+\beta+1}.$$

□

Ex. 4.11. Suppose that $X \sim N(\theta, 100)$, and the prior for θ is $N(100, 225)$. Find the Bayes estimator using the linear error loss function

$$L(d, \theta) = \begin{cases} 3(\theta - d), & d < \theta, \\ d - \theta, & d \geq \theta. \end{cases}$$

It is known that the posterior is

$$\pi(\theta|\mathbf{x}) \sim N(\mu(x), \eta^2)$$

where

$$\mu(x) = \frac{\sigma^2}{\sigma^2 + \tau^2}\mu + \frac{\tau^2}{\sigma^2 + \tau^2}x = \frac{9x + 400}{13}, \quad \eta^2 = \frac{\sigma^2\tau^2}{\tau^2 + \sigma^2} = 67.164.$$

Therefore, the Bayes estimator is 0.75 quantile of $N(\mu(x), \eta^2)$, which is

$$\mu(x) + Z_{0.75} \cdot \eta = 0.6923077x + 36.29692.$$

□

Ex. 4.12. Suppose that

$$\mathbf{X} = (X_1, \dots, X_n) \sim (\theta, \Sigma),$$

and denote the estimator of $\theta = (\theta_1, \dots, \theta_p)^T$ by $a = (a_1, \dots, a_p)^T$. If the prior of θ is $\pi(\theta)$, and the loss function is

$$L(a, \theta) = (a - \theta)^T D (a - \theta),$$

where D is a $p \times p$ positive definite matrix. Show that the Bayes estimator is

$$\hat{\theta}_B = E(\theta|\mathbf{x}).$$

The posterior risk is

$$R(\theta|\mathbf{x}) = \int_{\Theta} (a - \theta)^T D (a - \theta) \pi(\theta|\mathbf{x}) d\theta.$$

In order to find the a that minimizes $R(\theta|\mathbf{x})$, take the first derivative of $R(\theta|\mathbf{x})$ w.r.t. a , and set it as 0:

$$\nabla_a R(\theta|\mathbf{x}) = \int_{\Theta} (2Da - 2D\theta) \pi(\theta|x) d\theta = 0.$$

So we have

$$\hat{\theta}_B = \hat{a} = \int_{\Theta} \theta \pi(\theta|\mathbf{x}) d\theta = E(\theta|x).$$

□

Ex. 4.13. Suppose that there are N products in a batch, among which there are M (unknown) defects. If we randomly select n products from this batch, we will get x defects. Assume that the prior for M is the uniform distribution:

$$P(M = k) = \frac{1}{N+1} \quad (k = 0, 1, \dots, N).$$

Find the Bayes estimator of the defective rate $p = M/N$ using the square error loss.

The likelihood follows a hyper-geometric distribution:

$$f(x|M) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}.$$

As a result, the posterior is

$$\begin{aligned} \pi(M|x) &= \frac{f(x|M)\pi(M)}{\sum_{M=0}^N f(x|M)\pi(M)} \\ &= \frac{\binom{M}{x} \binom{N-M}{n-x}}{\sum_{M=0}^N \binom{M}{x} \binom{N-M}{n-x}} \\ &= Z \binom{(M-x)+x}{x} \binom{(n-x)+(N-n)-(M-x)}{n-x}, \end{aligned} \tag{105}$$

where Z is the normalization factor. Note that if let

$$y = M - x, \quad a = x, \quad b = n - x, \quad c = N - n,$$

Table 1: Parameter transformation		
Original	ABC distribution	Beta-binomial
x	a	$\alpha - 1$
$n - x$	b	$\beta - 1$
$N - n$	c	n

then we have

$$\sum_{y=0}^c \binom{a+y}{a} \binom{b+c-y}{b} = \binom{a+b+c+1}{a+b+1},$$

indicating that $Z^{-1} = \binom{N+1}{n+1}$. Furthermore, such a distribution is also a beta-binomial distribution with the transformations shown in table 1.

Therefore,

$$E(y|x) = \frac{n\alpha}{\alpha + \beta} = \frac{(N-n)(x+1)}{n+2},$$

and

$$E(p|x) = E(M/N|x) = (E(y|x) + x)/N = \frac{Nx + N - n + 2x}{(n+2)N},$$

which is exactly the Bayes estimator. \square

Ex. 4.14. Suppose that $X \sim B(5, \theta)$, and the prior for θ is $Be(1, 9)$. If $x = 1$, find the Bayes estimator using the following loss functions:

- (1) $L(a, \theta) = (\theta - a)^2$;
- (2) $L(a, \theta) = |\theta - a|$;
- (3) $L(a, \theta) = (\theta - a)^2 / [\theta(1 - \theta)]$;
- (4) $L(a, \theta) = \begin{cases} \theta - a, & \theta > a, \\ 2(a - \theta), & \theta \leq a. \end{cases}$

The posterior is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_0^1 f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{\Gamma(15)}{\Gamma(x+1)\Gamma(14-x)} \theta^x (1-\theta)^{13-x} \\ &= \frac{\Gamma(15)}{\Gamma(2)\Gamma(13)} \theta(1-\theta)^{12}. \end{aligned} \tag{106}$$

(1) With the square error loss,

$$\hat{\theta}_B = E(\theta|x) = \frac{x+1}{15} = \frac{2}{15}.$$

(2) With the absolute error loss,

$$\hat{\theta}_B = \text{Median}(\theta|x) = 0.1170221.$$

(3) With the weighted square error loss, and the weight $w(\theta) = [\theta(1-\theta)]^{-1}$, we have

$$\begin{aligned} E(\theta w(\theta)|x) &= \int_0^1 (1-\theta)^{-1} \pi(\theta|x) d\theta = \frac{14}{12}, \\ E(w(\theta)|x) &= \int_0^1 [\theta(1-\theta)]^{-1} \pi(\theta|x) d\theta = \frac{13 \cdot 14}{12}, \end{aligned}$$

and as a result,

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)} = \frac{1}{13}.$$

(4) With the linear error loss function,

$$\hat{\theta}_B = \frac{1}{3} \text{Quantile}(\theta|x) = 0.08437303.$$

Ex. 4.15. Consider the Linex loss:

$$L(d, \theta) = e^{c(\theta-d)} - c(\theta-d) - 1.$$

Show that:

- (1) $L(d, \theta) > 0$;
- (2) Draw the plot of Linex loss as function of $\theta - d$ when $c = 0.1, 0.5, 1.2$, respectively;
- (3) Find the Bayes estimator based on the Linex loss;
- (4) Find the Bayes estimator when

$$X_1, \dots, X_n \sim \text{i.i.d. } N(\theta, 1)$$

and $\pi(\theta) \equiv 1$.

(1) According to the Taylor expansion,

$$e^x = \sum_{i=1}^{\infty} \frac{x^i}{i!} > 1 + x,$$

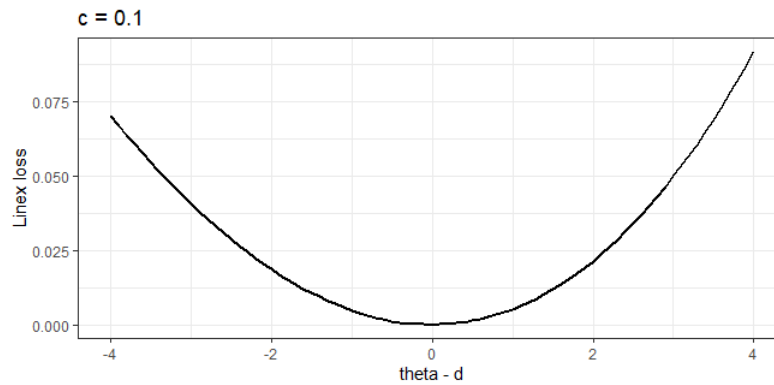
so

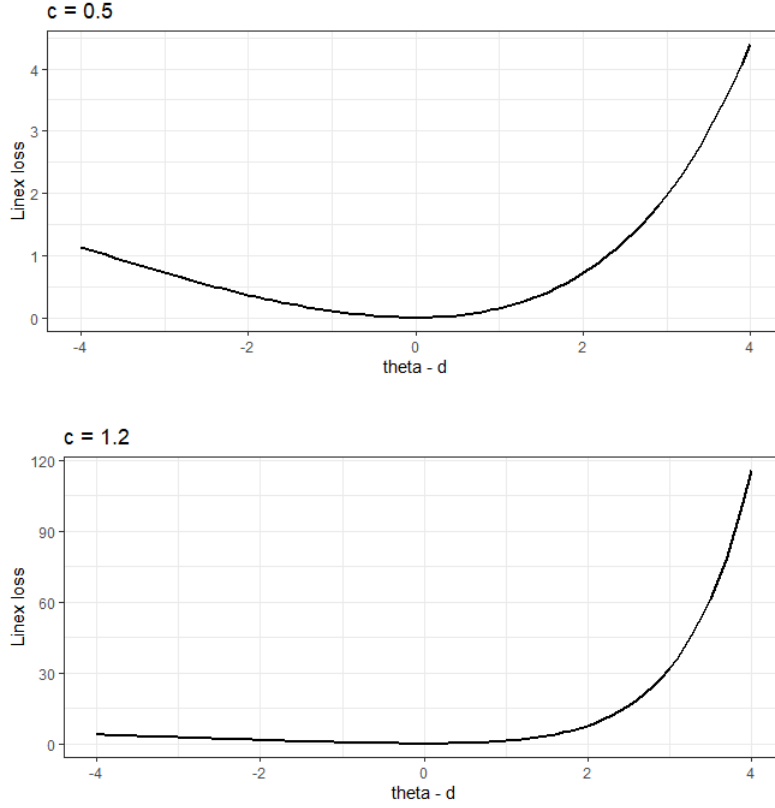
$$e^{c(\theta-d)} > 1 + c(\theta-d),$$

which is

$$L(d, \theta) = e^{c(\theta-d)} - c(\theta-d) - 1 > 0.$$

(2)





(3) The posterior risk can be expressed as

$$\begin{aligned}
 R(d|x) &= \int_{\mathbb{R}} L(d, \theta) \pi(\theta|x) d\theta \\
 &= \int_{\mathbb{R}} \left[e^{c(\theta-d)} - c(\theta-d) - 1 \right] \pi(\theta|x) d\theta.
 \end{aligned} \tag{107}$$

Take the first derivative of $R(d|x)$ w.r.t. d ,

$$\begin{aligned}
 \frac{\partial R(d|x)}{\partial d} &= -c \int_{\mathbb{R}} e^{c\theta} e^{-cd} \pi(\theta|x) d\theta + c \int_{\mathbb{R}} \pi(\theta|x) d\theta \\
 &= -ce^{-cd} E(e^{c\theta}|x) + c \\
 &= 0,
 \end{aligned} \tag{108}$$

we have

$$\hat{\theta}_B = d^* = \frac{1}{c} \log E(e^{c\theta}|x).$$

(4) It is known that the posterior on θ is $N(\bar{x}, 1/n)$. As a result,

$$\begin{aligned}
 E(e^{c\theta}|x) &= \int_{\mathbb{R}} \sqrt{\frac{n}{2\pi}} \exp \left\{ -\frac{n}{2} (\theta - \bar{x})^2 \right\} \exp \{ c\theta \} d\theta \\
 &= \sqrt{\frac{n}{2\pi}} \int_{\mathbb{R}} \exp \left\{ -\frac{n}{2} \left(\theta - \bar{x} - \frac{c}{n} \right)^2 \right\} \exp \left\{ c\bar{x} + \frac{c^2}{2n} \right\} d\theta \\
 &= \exp \left\{ c\bar{x} + \frac{c^2}{2n} \right\}.
 \end{aligned} \tag{109}$$

Thus,

$$\hat{\theta}_B = \frac{1}{c} \log E(e^{c\theta}|x) = \bar{x} + \frac{c}{2n}.$$

□

Ex. 4.16. Suppose that $X \sim \Gamma(\alpha, 1/\theta)$. The prior on θ is $\pi(\theta) = 1/\theta \cdot I(0 < \theta < +\infty)$. Find the Bayes estimator for θ using the weighted square error loss $L(d, \theta) = (d - \theta)^2/\theta^2$.

The posterior is

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\mathbb{R}^+} f(x|\theta)\pi(\theta)d\theta} = \frac{x^\alpha}{\Gamma(\alpha)} \theta^{-(\alpha+1)} e^{-x/\theta}.$$

With the weight $w(\theta) = \theta^{-2}$, we have

$$\begin{aligned} E(\theta w(\theta)|x) &= \int_{\mathbb{R}^+} \theta^{-1} \pi(\theta|x) d\theta = \frac{\alpha}{x}, \\ E(w(\theta)|x) &= \int_{\mathbb{R}^+} \theta^{-2} \pi(\theta|x) d\theta = \frac{(\alpha+1)\alpha}{x^2}, \end{aligned}$$

and as a result,

$$\hat{\theta}_B = \frac{E(\theta w(\theta)|x)}{E(w(\theta)|x)} = \frac{x}{\alpha+1}.$$

□

Ex. 4.17. When we lack information about loss functions, we can use the distance between $p(x|\theta)$ and $p(x|a)$ as a measure of loss. Consider

- (1) $L_e(a, \theta) = E^{X|\theta} \left[\log \frac{p(x|\theta)}{p(x|a)} \right];$
 - (2) $L_H(a, \theta) = \frac{1}{2} E^{X|\theta} \left[\sqrt{\frac{p(x|a)}{p(x|\theta)}} - 1 \right]^2.$
- Suppose that $X \sim N(0, \theta)$. Show that

$$\begin{aligned} L_e(a, \theta) &= \frac{1}{2}(a - \theta)^2, \\ L_H(a, \theta) &= 1 - \exp \left\{ -\frac{(a - \theta)^2}{8} \right\}. \end{aligned}$$

(1) Since

$$p(x|\theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x - \theta)^2}{2} \right\},$$

we have

$$\log \frac{p(x|\theta)}{p(x|a)} = \frac{a^2 - \theta^2}{2} - (a - \theta)x.$$

Therefore,

$$L_e(a, \theta) = \frac{a^2 - \theta^2}{2} - (a - \theta)E^{X|\theta}[X] = \frac{a^2 - \theta^2}{2} - (a - \theta)\theta = \frac{1}{2}(a - \theta)^2.$$

(2) Since

$$\begin{aligned}
\left[\sqrt{\frac{p(x|a)}{p(x|\theta)}} - 1 \right]^2 &= \left[\sqrt{\exp\left\{\frac{\theta^2 - a^2}{2}\right\}} \sqrt{\exp\{(a - \theta)x\}} - 1 \right]^2 \\
&= 1 + \exp\left\{\frac{\theta^2 - a^2}{2}\right\} \exp\{(a - \theta)x\} - 2\sqrt{\exp\left\{\frac{\theta^2 - a^2}{2}\right\}} \sqrt{\exp\{(a - \theta)x\}},
\end{aligned} \tag{110}$$

consider

$$\begin{aligned}
E^{X|\theta}[\exp\{(a - \theta)x\}] &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\{(a - \theta)x\} \exp\left\{-\frac{(x - \theta)^2}{2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left\{-\frac{1}{2}(x - a)^2\right\} \exp\left\{\frac{1}{2}a^2 - \frac{1}{2}\theta^2\right\} dx \\
&= \exp\left\{\frac{1}{2}a^2 - \frac{1}{2}\theta^2\right\},
\end{aligned} \tag{111}$$

and

$$\begin{aligned}
E^{X|\theta}[\sqrt{\exp\{(a - \theta)x\}}] &= E^{X|\theta}\left[\exp\left\{\frac{(a - \theta)x}{2}\right\}\right] \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{(a - \theta)x}{2}\right\} \exp\left\{-\frac{(x - \theta)^2}{2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left\{-\frac{1}{2}\left(x - \frac{\theta + a}{2}\right)^2\right\} \exp\left\{\frac{1}{8}a^2 - \frac{3}{8}\theta^2 + \frac{1}{4}\theta a\right\} dx \\
&= \exp\left\{\frac{1}{8}a^2 - \frac{3}{8}\theta^2 + \frac{1}{4}\theta a\right\}.
\end{aligned} \tag{112}$$

Therefore,

$$\begin{aligned}
L_H(a, \theta) &= \frac{1}{2} E^{X|\theta} \left[\sqrt{\frac{p(x|a)}{p(x|\theta)}} - 1 \right]^2 \\
&= \frac{1}{2} \left[1 + \exp\left\{\frac{\theta^2 - a^2}{2}\right\} \exp\left\{\frac{1}{2}a^2 - \frac{1}{2}\theta^2\right\} - 2\sqrt{\exp\left\{\frac{\theta^2 - a^2}{2}\right\}} \exp\left\{\frac{1}{8}a^2 - \frac{3}{8}\theta^2 + \frac{1}{4}\theta a\right\} \right] \\
&= 1 - \exp\left\{-\frac{(a - \theta)^2}{8}\right\}.
\end{aligned} \tag{113}$$

□

Ex. 4.18. *Ex. 17 Cont'd. Suppose that $X \sim N(0, \theta)$. Find L_e and L_H .*

(1) Similar to Ex. 4.17(1),

$$L_e(a, \theta) = \frac{a^2 - \theta^2}{2} - (a - \theta)E^{X|\theta}[X] = \frac{a^2 - \theta^2}{2} - (a - \theta) \cdot 0 = \frac{a^2 - \theta^2}{2}.$$

(2) Similar to Ex. 4.17(2), we have to consider

$$\begin{aligned}
E^{X|\theta} [\exp\{(a - \theta)x\}] &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\theta}} \exp\{(a - \theta)x\} \exp\left\{-\frac{x^2}{2\theta}\right\} dx \\
&= \frac{1}{\sqrt{2\pi\theta}} \int_{\mathbb{R}} \exp\left\{-\frac{1}{2\theta}[x - \theta(a - \theta)]^2\right\} \exp\left\{\frac{\theta}{2}(a - \theta)^2\right\} dx \quad (114) \\
&= \exp\left\{\frac{1}{2}\theta(a - \theta)^2\right\},
\end{aligned}$$

and

$$\begin{aligned}
E^{X|\theta} \left[\sqrt{\exp\{(a - \theta)x\}} \right] &= E^{X|\theta} \left[\exp\left\{\frac{(a - \theta)x}{2}\right\} \right] \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\theta}} \exp\left\{\frac{(a - \theta)x}{2}\right\} \exp\left\{-\frac{x^2}{2\theta}\right\} dx \\
&= \frac{1}{\sqrt{2\pi\theta}} \int_{\mathbb{R}} \exp\left\{-\frac{1}{2\theta}\left[x - \frac{\theta(a - \theta)}{2}\right]^2\right\} \exp\left\{\frac{1}{8}\theta(a - \theta)^2\right\} dx \\
&= \exp\left\{\frac{1}{8}\theta(a - \theta)^2\right\}. \quad (115)
\end{aligned}$$

Therefore,

$$\begin{aligned}
L_H(a, \theta) &= \frac{1}{2} E^{X|\theta} \left[\sqrt{\frac{p(x|a)}{p(x|\theta)}} - 1 \right]^2 \\
&= \frac{1}{2} \left[1 + \exp\left\{\frac{\theta^2 - a^2}{2}\right\} \exp\left\{\frac{1}{2}\theta(a - \theta)^2\right\} - 2\sqrt{\exp\left\{\frac{\theta^2 - a^2}{2}\right\} \exp\left\{\frac{1}{8}\theta(a - \theta)^2\right\}} \right]. \quad (116)
\end{aligned}$$

□

Ex. 4.19. *Ex. 17 Cont'd. Suppose that $X \sim N(\theta, 1)$, and the posterior on θ is $N(\mu(x), \eta^2)$. Show that the Bayes estimators using $L_e(a, \theta)$ and $L_H(a, \theta)$ are both $\mu(x)$.*

(1) When $X \sim N(\theta, 1)$, $L_e(a, \theta) = \frac{1}{2}(a - \theta)^2$, which is a multiple of the square error loss. Therefore, $\hat{\theta}_B$ for $L_e(a, \theta)$ is the same as that for the square error loss.

(2) When $X \sim N(\theta, 1)$, $L_H(a, \theta) = 1 - \exp\left\{-\frac{(a - \theta)^2}{8}\right\}$. As a result,

$$\begin{aligned}
R(a|x) &= \int_{\mathbb{R}} L_H(a, \theta) \pi(\theta|x) d\theta \\
&= 1 - \int_{\mathbb{R}} \exp\left\{-\frac{(a - \theta)^2}{8}\right\} \pi(\theta|x) d\theta. \quad (117)
\end{aligned}$$

Since,

$$\frac{\partial R(a|x)}{\partial a} = \int_{\mathbb{R}} \exp\left\{-\frac{(a - \theta)^2}{8}\right\} \frac{a - \theta}{4} \pi(\theta|x) d\theta = 0,$$

we have

$$a \int_{\mathbb{R}} \exp \left\{ -\frac{(a-\theta)^2}{8} \right\} \pi(\theta|x) d\theta = \int_{\mathbb{R}} \theta \exp \left\{ -\frac{(a-\theta)^2}{8} \right\} \pi(\theta|x) d\theta.$$

Now consider

$$\begin{aligned} \exp \left\{ -\frac{(a-\theta)^2}{8} \right\} \pi(\theta|x) &= K \exp \left\{ -\frac{(a-\theta)^2}{8} \right\} \exp \left\{ -\frac{(\theta-\mu(x))^2}{2\eta^2} \right\} \\ &= K \exp \left\{ -\frac{\eta^2+4}{8\eta^2} \left(\theta - \frac{a\eta^2+4\mu(x)}{\eta^2+4} \right)^2 \right\} \exp \left\{ \frac{(a\eta^2+4\mu(x))^2}{8\eta^2(\eta^2+4)} \right\}, \end{aligned} \quad (118)$$

thus

$$a \int_{\mathbb{R}} \exp \left\{ -\frac{\eta^2+4}{8\eta^2} \left(\theta - \frac{a\eta^2+4\mu(x)}{\eta^2+4} \right)^2 \right\} d\theta = \int_{\mathbb{R}} \theta \exp \left\{ -\frac{\eta^2+4}{8\eta^2} \left(\theta - \frac{a\eta^2+4\mu(x)}{\eta^2+4} \right)^2 \right\} d\theta.$$

Let $N = \int_{\mathbb{R}} \exp \left\{ -\frac{\eta^2+4}{8\eta^2} \left(\theta - \frac{a\eta^2+4\mu(x)}{\eta^2+4} \right)^2 \right\} d\theta$, we have

$$a = \int_{\mathbb{R}} \theta \frac{1}{N} \exp \left\{ -\frac{\eta^2+4}{8\eta^2} \left(\theta - \frac{a\eta^2+4\mu(x)}{\eta^2+4} \right)^2 \right\} d\theta = \frac{a\eta^2+4\mu(x)}{\eta^2+4},$$

indicating that

$$\hat{\theta}_B = \hat{a} = \mu(x).$$

□

Ex. 4.20. Suppose that $X \sim N(\theta, 100)$, $\theta \sim N(100, 225)$, and the loss function is

$$L(a, \theta) = (\theta - a)^2 \exp \left\{ \frac{(\theta - 100)^2}{900} \right\}.$$

Find the Bayes estimator of θ .

It is known that the posterior on θ is

$$\theta|x \sim N \left(\frac{400}{13} + \frac{9}{13}x, \frac{900}{13} \right).$$

Since

$$R(a|x) = \int_{\mathbb{R}} L(a, \theta) \pi(\theta|x) d\theta = \int_{\mathbb{R}} (\theta - a)^2 \exp \left\{ \frac{(\theta - 100)^2}{900} \right\} \pi(\theta|x) d\theta,$$

we have

$$\frac{\partial R(a|x)}{\partial a} = 2 \int_{\mathbb{R}} (a - \theta) \exp \left\{ \frac{(\theta - 100)^2}{900} \right\} \pi(\theta|x) d\theta = 0,$$

which is

$$a \int_{\mathbb{R}} \exp \left\{ \frac{(\theta - 100)^2}{900} \right\} \pi(\theta|x) d\theta = \int_{\mathbb{R}} \theta \exp \left\{ \frac{(\theta - 100)^2}{900} \right\} \pi(\theta|x) d\theta.$$

Now consider

$$\begin{aligned} \exp \left\{ \frac{(\theta - 100)^2}{900} \right\} \pi(\theta|x) &= K \exp \left\{ \frac{(\theta - 100)^2}{900} \right\} \exp \left\{ -\frac{(\theta - \frac{400-9x}{13})^2}{1800/13} \right\} \\ &= K^* \exp \left\{ -\frac{11}{1800} \left(\theta - \frac{200+9x}{11} \right)^2 \right\}, \end{aligned} \quad (119)$$

and let $N = \int_{\mathbb{R}} \exp \left\{ -\frac{11}{1800} \left(\theta - \frac{200+9x}{11} \right)^2 \right\} d\theta$, then we have

$$\hat{\theta}_B = \hat{a} = \int_{\mathbb{R}} \theta \frac{1}{N} \exp \left\{ -\frac{11}{1800} \left(\theta - \frac{200+9x}{11} \right)^2 \right\} d\theta = \frac{200+9x}{11}.$$

□

Ex. 4.21. Suppose that $\mathbf{X} = (X_1, \dots, X_k)$ follows the multinomial distribution $M(n, \theta)$, where $\theta = (\theta_1, \dots, \theta_k)$; and that the prior on θ is $D(\alpha_1, \dots, \alpha_k)$. Find the Bayes estimator and the posterior risk of θ using the loss

$$L(\mathbf{a}, \theta) = \sum_{i=1}^k (\theta_i - a_i)^2.$$

It is known that the posterior on θ is $D(\alpha_1 + x_1, \dots, \alpha_k + x_k)$. The posterior risk is

$$\begin{aligned} R(\mathbf{a}|\mathbf{x}) &= \int_{\Theta} L(\mathbf{a}, \theta) \pi(\theta|\mathbf{x}) d\theta_1 \dots d\theta_k \\ &= \int_{\Theta} \sum_{i=1}^k (\theta_i - a_i)^2 \pi(\theta|\mathbf{x}) d\theta_1 \dots d\theta_k \\ &= \sum_{i=1}^k \int_{\theta_i} (\theta_i - a_i)^2 \pi_i(\theta_i|\mathbf{x}) d\theta_i, \end{aligned} \quad (120)$$

where $\pi_i(\theta_i|\mathbf{x}) = \text{Beta}(\alpha_i, \sum_{j \neq i} \alpha_j)$, is the marginal distribution of θ_i . Since

$$\frac{\partial R(\mathbf{a}|\mathbf{x})}{\partial a_i} = 2 \int_{\theta_i} (a_i - \theta_i) \pi_i(\theta_i|\mathbf{x}) d\theta_i = 0, \quad (121)$$

we know

$$\hat{a}_i = \int_{\theta_i} \theta_i \pi_i(\theta_i|\mathbf{x}) = \frac{\alpha_i}{\sum_{j=1}^k \alpha_j}.$$

In other words,

$$\hat{\theta}_B = \left(\frac{\alpha_1}{\sum_{j=1}^k \alpha_j}, \dots, \frac{\alpha_k}{\sum_{j=1}^k \alpha_j} \right).$$

In this case,

$$\begin{aligned} R(\mathbf{a}|\mathbf{x}) &= \sum_{i=1}^k \text{Var}(\theta_i|\mathbf{x}) \\ &= \sum_{i=1}^k \frac{\alpha_i \sum_{j \neq i} \alpha_j}{(\sum_{j=1}^k \alpha_j)^2 (\sum_{j=1}^k \alpha_j + 1)^2}. \end{aligned} \quad (122)$$

□

Ex. 4.22. Suppose that $X \sim B(5, \theta)$, and the prior for θ is $Be(1, 9)$. If $x = 0$, do the hypothesis test

$$H_0 : 0 \leq \theta \leq 0.15 \leftrightarrow H_1 : \theta > 0.15$$

using the following loss functions (where a_i indicates the acceptance of H_i):

(1)

$$L(a_0, \theta) = \begin{cases} 0, & \theta \leq 0.15, \\ 1, & \theta > 0.15, \end{cases} \quad L(a_1, \theta) = \begin{cases} 2, & \theta \leq 0.15, \\ 0, & \theta > 0.15; \end{cases}$$

(2)

$$L(a_0, \theta) = \begin{cases} 0, & \theta \leq 0.15, \\ 1, & \theta > 0.15, \end{cases} \quad L(a_1, \theta) = \begin{cases} 2, & \theta \leq 0.15, \\ 0.15 - \theta, & \theta > 0.15; \end{cases}$$

It is known that the posterior is

$$\pi(\theta|x=0) = \frac{\Gamma(15)}{\Gamma(x+1)\Gamma(14-x)}\theta^x(1-\theta)^{13-x} = \frac{\Gamma(15)}{\Gamma(1)\Gamma(14)}(1-\theta)^{13}.$$

(1) Since

$$P(\Theta_1|x=0) = \int_{0.15}^1 \pi(\theta|x=0)d\theta = 0.1027697,$$

we know the sample is out of the rejection region

$$D = \left\{ X : P(\Theta_1|x=0) \geq \frac{2}{1+2} \right\}.$$

As a result, we should accept the null hypothesis.

(2) Since

$$R(a_0|x) = P(\Theta_1|x=0) = 0.1027697,$$

and

$$\begin{aligned} R(a_1|x) &= \int_{\Theta_0} 2\pi(\theta|x)d\theta + \int_{\Theta_1} (0.15 - \theta)\pi(\theta|x)d\theta \\ &= 2P(\Theta_0|x=0) + 0.15P(\Theta_1|x=0) - \int_{0.15}^1 \theta\pi(\theta|x=0)d\theta \\ &= 2 \times 0.8972303 + 0.15 \times 0.1027697 - 0.02123907 \\ &= 1.788637. \end{aligned} \tag{123}$$

Since $R(a_1|x) > R(a_0|x)$, we should accept the null hypothesis. \square

Ex. 4.23. Suppose that $X \sim B(5, \theta)$, the prior for θ is $Be(1, 9)$, and $x = 1$. If we have 2 options, and the loss of the first option is 10 θ , while the loss of the second one is 1. Which option should we take?

It is known that the posterior is

$$\pi(\theta|x=1) = \frac{\Gamma(15)}{\Gamma(x+1)\Gamma(14-x)}\theta^x(1-\theta)^{13-x} = \frac{\Gamma(15)}{\Gamma(2)\Gamma(14)}\theta(1-\theta)^{12}.$$

Then

$$R(O_1|x=1) = \int_0^1 10\theta\pi(\theta|x=1)d\theta = \frac{2}{3},$$

and

$$R(O_2|x=1) = \int_0^1 \pi(\theta|x=1)d\theta = 1.$$

Since $R(O_1|x=1) < R(O_2|x=1)$, we should take the first option. \square

Ex. 4.24. Suppose that

$$f(x|\theta) = e^{-(x-\theta)}I(x > \theta),$$

and the prior for θ is

$$\pi(\theta) = e^{-\theta}I(\theta > 0).$$

If $x = 4$, do the hypothesis test

$$H_1 : 0 < \theta \leq 1,$$

$$H_2 : 1 < \theta \leq 2,$$

$$H_3 : 2 < \theta \leq 3,$$

$$H_4 : \theta > 3.$$

using the following loss functions (where a_i indicates the acceptance of H_i):

$$\begin{aligned} L(\theta, a_1) &= \begin{cases} 0, & 0 < \theta \leq 1, \\ 1, & 1 < \theta \leq 2, \\ 1, & 2 < \theta \leq 3, \\ 3, & \theta > 3. \end{cases} & L(\theta, a_2) &= \begin{cases} 1, & 0 < \theta \leq 1, \\ 0, & 1 < \theta \leq 2, \\ 2, & 2 < \theta \leq 3, \\ 3, & \theta > 3. \end{cases} \\ L(\theta, a_3) &= \begin{cases} 1, & 0 < \theta \leq 1, \\ 2, & 1 < \theta \leq 2, \\ 0, & 2 < \theta \leq 3, \\ 3, & \theta > 3. \end{cases} & L(\theta, a_4) &= \begin{cases} 2, & 0 < \theta \leq 1, \\ 2, & 1 < \theta \leq 2, \\ 2, & 2 < \theta \leq 3, \\ 0, & \theta > 3. \end{cases} \end{aligned}$$

The posterior is

$$\begin{aligned} \pi(\theta|x) &= \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta} \\ &= \frac{e^{-(x-\theta)}e^{-\theta}I(x > \theta > 0)}{\int_0^x e^{-(x-\theta)}e^{-\theta}I(x > \theta > 0)d\theta} \\ &= \frac{1}{x}I(0 < \theta < x) \\ &= \frac{1}{4}I(0 < \theta < 4). \end{aligned} \tag{124}$$

As a result,

$$\begin{aligned} R(a_1|x=4) &= P(\Theta_2|x=4) + P(\Theta_3|x=4) + 3P(\Theta_4|x=4) = 5, \\ R(a_2|x=4) &= P(\Theta_1|x=4) + 2P(\Theta_3|x=4) + 3P(\Theta_4|x=4) = 6, \\ R(a_3|x=4) &= P(\Theta_1|x=4) + 2P(\Theta_2|x=4) + 3P(\Theta_4|x=4) = 6, \\ R(a_4|x=4) &= 2P(\Theta_1|x=4) + 2P(\Theta_2|x=4) + 2P(\Theta_3|x=4) = 6. \end{aligned}$$

We should accept the hypothesis H_1 . \square