Winter 22 – STAT206B Homework 2 Solution

1. Let $X_1, \ldots, X_n$ be an i.i.d. sample such that $X_i \mid \theta, \sigma^2 \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$, where σ^2 is known and θ is unknown. Also, let your prior for θ be a mixture of conjugate priors, i.e.,

$$\pi(\theta) = \sum_{\ell=1}^{K} w_\ell \phi(\theta \mid \mu_\ell, \tau^2)$$

where $\phi(\theta \mid \mu_\ell, \tau^2)$ denotes the Gaussian density with mean μ_ℓ and variance τ^2 and mixture weights $0 < w_\ell < 1$ for all $\ell = 1, \ldots, K$ with $\sum_{\ell=1}^{K} w_\ell = 1$.

Note: This questions is challenging. Use the results from the class example with $X_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$ and $\theta \sim \mathrm{N}(\mu, \tau^2)$.

will be included in HW3.

2. Let $X_1, \ldots, X_n$ be an i.i.dṡample such that $X_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, 1)$. Suppose that you know that $\theta > 0$, and you want your prior to reflect that fact. Hence, you decide to set $\pi(\theta)$ to be a normal distribution with mean μ and variance τ^2, truncated to be positive, i.e.,

$$\pi(\theta) = \frac{1}{\sqrt{2\pi\tau^2}\Phi(\mu/\tau)} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\} \mathrm{I}_{[0,\infty)}(\theta),$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution and $\mathrm{I}_{[0,\infty)}(\cdot)$ the indicator function.

(a) Find the posterior distribution for θ based on this prior.

Using the Sufficiency Principle, we know $\pi(\theta \mid \bar{x}) = \pi(\theta \mid \boldsymbol{x})$, with $\bar{x} \mid \theta \sim \mathrm{N}(\theta, 1/n)$.

$$\begin{aligned}
\pi(\theta \mid \bar{x}) &\propto f(\bar{x} \mid \theta)\pi(\theta) \\
&\propto \exp\left\{-\frac{(\bar{x}-\theta)^2}{2/n}\right\} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\} \mathbb{1}_{[0,\infty)}(\theta) \\
&\propto \exp\left\{-\frac{1}{2}\left(n + \frac{1}{\tau^2}\right)\left(\theta - \left(n + \frac{1}{\tau^2}\right)^{-1}\left(n\bar{x} + \frac{\mu}{\tau^2}\right)\right)^2\right\} 1_{[0,\infty)}(\theta).
\end{aligned}$$

This is the kernel of the normal with $\mu_1 = \tau_1^2\left(n\bar{x} + \frac{\mu}{\tau^2}\right)$ and $\tau_1^2 = \left(n + \frac{1}{\tau^2}\right)^{-1}$ truncated at 0. Observe that the truncated normal is a conjugate prior since the prior and the posterior both have the same functional form.

(b) Find the prior predictive distribution.

With $\bar{x} = \sum_{i=1}^{n} x_i/n$, and $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$, we write

$$f(\boldsymbol{x} \mid \theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right).$$

Then we have

$$\begin{aligned} m(\boldsymbol{x}) &= \int_0^\infty f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}\sqrt{2\pi\tau^2}\Phi(\mu/\tau)} \exp\left(-\frac{s^2}{2\sigma^2}\right) \\ &\quad \times \int_0^\infty \underbrace{\exp\left\{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu)^2}{2\tau^2}\right\}}_{\text{see a kernel for } \mathrm{N}(\mu_1, \tau_1)} d\theta \\ &= \frac{\sqrt{\tau_1^2}\Phi(\mu_1/\tau_1)}{(2\pi\sigma^2)^{n/2}\sqrt{\tau^2}\Phi(\mu/\tau)} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{(\bar{x}-\mu)^2}{2(\tau^2+\sigma^2/n)}\right\}. \end{aligned}$$

where $\mu_1 = \tau_1^2\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}\right)$ and $\tau_1^2 = \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right)^{-1}$. For this problem, $\sigma^2 = 1$.

3. Let $X_1, \ldots, X_n$ be an i.i.d. sample such that each X_i comes from a truncated normal with unknown mean θ and variance 1,

$$f(X_i \mid \theta) = \frac{1}{\sqrt{2\pi}\Phi(\theta)} \exp\left\{-\frac{(X_i-\theta)^2}{2}\right\} \mathrm{I}_{[0,\infty)}(X_i).$$

If $\theta \sim \mathrm{N}(\mu, \tau^2)$, find the posterior for θ.

Since $\pi(\theta)$ is proper, $\pi(\theta \mid \boldsymbol{x})$ is proper. We use the proportionality argument to find the posterior,

$$\begin{aligned} \pi(\theta \mid \bar{x}) &\propto f(\bar{x} \mid \theta)\pi(\theta) \\ &\propto \frac{1}{\Phi(\theta)^n} \exp\left\{-\frac{1}{2}\left(n + \frac{1}{\tau^2}\right)\left(\theta - \left(n + \frac{1}{\tau^2}\right)^{-1}\left(n\bar{x} + \frac{\mu}{\tau^2}\right)\right)^2\right\}. \end{aligned}$$

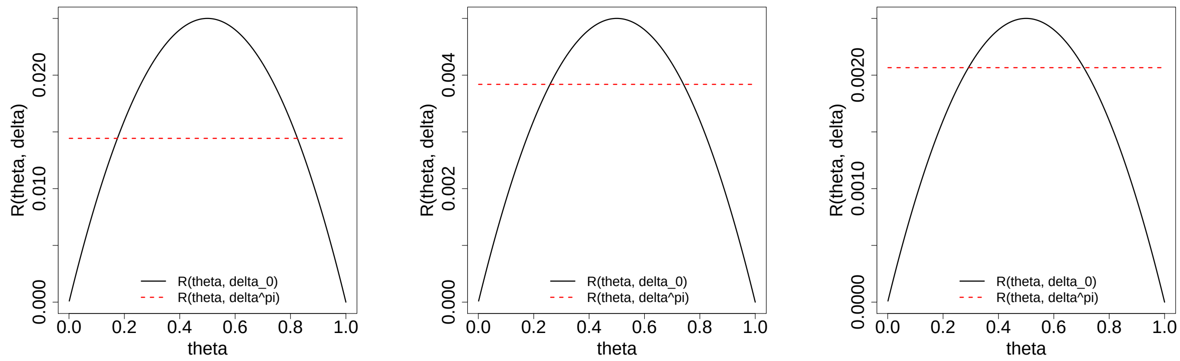
Due to the factor, $1/\Phi(\theta)^n$, this is not a normal distribution. Thus, it is not a conjugate prior.

4. (Robert 2.28 - slightly reworded) Consider $x \mid \theta \sim \text{Binomial}(n, \theta)$ with n known.

   (a) If the prior is $\theta \sim \mathrm{Be}(\sqrt{n}/2, \sqrt{n}/2)$, give the associate posterior.

   From class, we have $\theta \mid x \sim \mathrm{Be}(\sqrt{n}/2 + x, \sqrt{n}/2 + n - x)$.

   (b) What is the estimator that minimizes the posterior expected loss if the loss function is $L(\delta, \theta) = (\theta - \delta)^2$? Call such estimator $\delta^\pi(x)$ and show that its associated risk $R(\theta, \delta^\pi(x))$ is constant.



(a) $n = 10$ (b) $n = 50$ (c) $n = 100$

Figure 1: Comparison of $R(\theta, \delta_0(x))$ and $R(\theta, \delta^\pi(x))$ with $n = 10, 50$ and 100

Under the squared error loss function, $\delta^\pi(x) = \mathrm{E}(\theta \mid x) = \frac{\sqrt{n}/2+x}{\sqrt{n}+n}$. We have $\mathrm{E}(\delta^\pi(x)) = \frac{\sqrt{n}/2+n\theta}{\sqrt{n}+n}$ and we find

$$\begin{aligned} R(\theta, \delta^\pi(x)) &= \mathrm{E}_\theta\left((\theta - \delta^\pi(x))^2\right) \\ &= \mathrm{E}_\theta\left((\theta \pm \mathrm{E}_\theta(\delta^\pi(x)) - \delta^\pi(x))^2\right) \\ &= \mathrm{bias}^2 + \mathrm{Var}(\delta^\pi(x)) \\ &= \left(\theta - \frac{\sqrt{n}/2 + n\theta}{\sqrt{n}+n}\right)^2 + \mathrm{Var}\left(\frac{\sqrt{n}/2 + x}{\sqrt{n}+n}\right) \\ &= \left(\frac{\theta - 1/2}{1+\sqrt{n}}\right)^2 + \frac{\mathrm{Var}\,(x)}{(\sqrt{n}+n)^2} \\ &= \frac{1}{4(1+\sqrt{n})^2}. \end{aligned}$$

i.e., $R(\theta, \delta^\pi(x)) = \frac{1}{4(1+\sqrt{n})^2}$ is constant over θ.

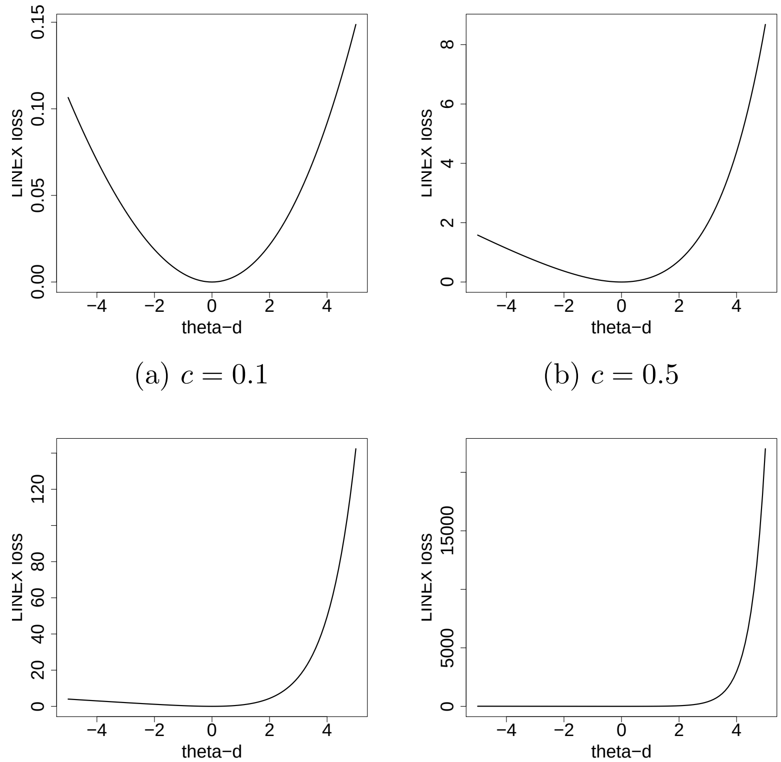
(c) Let $\delta_0(x) = x/n$. Find the risk for this estimator, i.e., find $R(\theta, \delta_0(x))$. Compare the risks for $\delta^\pi(x)$ and $\delta_0(x)$ for $n = 10, 50,$ and 100. Conclude about the appeal of $\delta^\pi(x)$.

First, $R(\theta, \delta_0(x)) = \mathrm{E}\left\{(\theta - x/n)^2\right\} = \theta(1-\theta)/n$. Fig 1 compares $R(\theta, \delta^\pi(x))$ and $R(\theta, \delta_0(x))$ for each of $n = 10, 50$ and 100. $\delta^\pi(x)$ is appealing under the minimax criterion since $R(\theta, \delta^\pi(x)) \leq \sup_{\theta\in\Theta} R(\theta, \delta_0(x))$.

5. (Robert 2.30 - slightly reworded). Consider $x \sim \mathrm{N}(\theta, 1)$ and $\theta \sim \mathrm{N}(0, n)$. Let $\delta^\pi(x)$ be the estimator that minimizes the posterior expected loss under the square error loss. Show that the Bayes risk $r(\pi, \delta^\pi)$ is equal to $n/(n+1)$.

Under the squared error loss, we have $\delta^\pi(x) = \mathrm{E}(\theta \mid x) = \frac{n}{n+1}x$. We then have

$$r(\pi, \delta^\pi) = \mathrm{E}\left[\mathrm{E}\left\{\left(\frac{n}{n+1}x - \theta\right)^2 \mid \theta\right\}\right] = \mathrm{E}\left\{1 - \frac{2}{n+1} + \frac{1+\theta^2}{(n+1)^2}\right\} = \frac{n}{n+1}.$$



(a) $c = 0.1$ (b) $c = 0.5$

(c) $c = 1.0$ (d) $c = 2$

Figure 2: Comparison of LINEX loss function $L(\theta, d)$ with $c = 0.1, 0.5,$ 1, and 2

6. (Adapted from Robert 2.42) Consider the LINEX loss function defined by

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

(a) Show that $L(\theta, d) \geq 0$ and plot this loss as a function of $(\theta - d)$ when $c = 0.1, 0.5, 1, 2$.

Let $z = \theta - d$ and $f(z) = e^{cz} - cz - 1$. We can check that $f(z)$ obtains its minimum at $z = 0$, i.e., $f(z) \geq f(0) = 0$ for any $z \in \mathbb{R}$. As the constant c varies, the loss function varies from very asymmetric to almost symmetric. Fig 2 illustrates how $L(\theta, d)$ changes with different values of c.

(b) Give the expression of a Bayes estimator $\delta^\pi(x)$ under this loss, i.e., find the estimator that minimizes the posterior posterior loss.

To obtain $\delta^\pi(x)$, we find $d \in \mathbb{R}$ for every x that minimizes $\rho(\pi, d \mid x) = \mathrm{E}^\pi(L(\theta, d) \mid x) = \mathrm{E}^\pi(e^{c(\theta - d)} - c(\theta - d) - 1 \mid x)$.

(c) Find $\delta^\pi(x)$ when $x_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, 1)$, $i = 1, \ldots, n$ and $\theta \sim \mathrm{N}(\mu, \tau^2)$.

From class, we know that $\theta \mid \boldsymbol{x} \sim \mathrm{N}(\mu_1, \tau_1^2)$, where $\mu_1 = \{1/\tau^2 + 1/(\sigma^2/n)\}^{-1}\{\mu/\tau^2 +$

$\bar{x}/(\sigma^2/n)\}$ and $\tau_1^2 = \{1/\tau^2 + 1/(\sigma^2/n)\}^{-1}$.

$$\begin{aligned}
\rho(\pi, d \mid x) &= \mathrm{E}^\pi \left(e^{c(\theta-d)} - c(\theta - d) - 1 \mid x \right) \\
&= e^{-cd}\mathrm{E}^\pi(e^{c\theta} \mid x) - c(\mu_1 - d) - 1 \\
&= \exp\left(-cd + c\mu_1 + \frac{c^2\tau_1^2}{2} \right) - c(\mu_1 - d) - 1.
\end{aligned}$$

We take derivatives of $\rho(\pi, d \mid x)$,

$$\begin{aligned}
\frac{d\rho(\pi, d \mid x)}{dd} &= -c\exp\left(-cd + c\mu_1 + \frac{c^2\tau_1^2}{2} \right) + c. \\
\frac{d^2\rho(\pi, d \mid x)}{dd^2} &= c^2\exp\left(-cd + c\mu_1 + \frac{c^2\tau_1^2}{2} \right) > 0.
\end{aligned}$$

We set $-c\exp\left(-cd + c\mu_1 + \frac{c^2\tau_1^2}{2} \right) + c = 0$ and solve for d. We find $d = \mu_1 + \frac{c\tau_1^2}{2}$. Thus, $\delta^\pi(x) = \mu_1 + \frac{c\tau_1^2}{2}$.

7. Let $L(\theta, d) = w(\theta)(\theta - d)^2$, with $w(\theta)$ a non-negative function, be the weighted quadratic loss (See CR Corollary 2.5.2). Show that $\delta^\pi(x)$, the estimator that minimizes the posterior expected loss $\rho(\pi, d \mid x)$ has the form $\delta^\pi(x) = \mathrm{E}(w(\theta)\theta \mid x)/\mathrm{E}(w(\theta) \mid x)$.

   Hint: Show that any other estimator has a larger posterior expected loss.

   Let $\delta^\pi(x) = \mathrm{E}(w(\theta)\theta \mid x)/\mathrm{E}(w(\theta) \mid x)$. For any decision rule, $\delta(x)$, we have

$$\begin{aligned}
\rho(\pi, \delta \mid x) &= \mathrm{E}\{w(\theta)(\theta - \delta)^2 \mid x\} \\
&= \mathrm{E}\{w(\theta)(\theta \pm \delta^\pi - \delta)^2 \mid x\} \\
&= \mathrm{E}\{w(\theta)(\theta - \delta^\pi)^2 \mid x\} + \mathrm{E}\{w(\theta)(\delta^\pi - \delta)^2 \mid x\} + \mathrm{E}\{w(\theta)(\theta - \delta^\pi)(\delta^\pi - \delta) \mid x\} \\
&= \mathrm{E}\{w(\theta)(\theta - \delta^\pi)^2 \mid x\} + \mathrm{E}\{w(\theta)(\delta^\pi - \delta)^2 \mid x\} \\
&\qquad + (\delta^\pi - \delta)\underbrace{\left[\mathrm{E}\{w(\theta)\theta \mid x\} - \mathrm{E}\{w(\theta) \mid x\}\delta^\pi\right]}_{\text{is zero after plugging in } \delta^\pi} \\
&= \underbrace{\mathrm{E}\{w(\theta)(\theta - \delta^\pi)^2 \mid x\}}_{=\rho(\pi, \delta^\pi \mid x)} + \mathrm{E}\{w(\theta)(\delta^\pi - \delta)^2 \mid x\}.
\end{aligned}$$

   Both terms are positive for any $\delta(x)$, which means any $\delta(x)$ has larger posterior expected loss than $\delta^\pi(x)$. That is, $\delta^\pi(x) = \mathrm{E}(w(\theta)\theta \mid x)/\mathrm{E}(w(\theta) \mid x)$ is the optimal.

8. Let $X \mid \theta \sim \text{Binomial}(n, \theta)$ with $\theta \sim \text{Be}(\alpha, \beta)$. Let $L(\theta, d) = (\theta - d)^2/\{\theta(1 - \theta)\}$. Find the estimator that minimizes the posterior expected loss $\rho(\pi, \delta \mid x)$ under this loss function.

   From Q7, we know that with $w(\theta) = 1/\{\theta(1 - \theta)\}$,

$$\delta^\pi(x) = \frac{\mathrm{E}(w(\theta)\theta \mid x)}{\mathrm{E}(w(\theta) \mid x)} = \frac{\mathrm{E}(\theta/\{(1 - \theta)\theta\} \mid x)}{\mathrm{E}(1/\{\theta(1 - \theta)\} \mid x)}.$$

From class, we also have $\theta \mid x \sim \mathrm{Be}(\alpha + x, \beta + n - x)$. We thus have

$$\delta^\pi(x) = \frac{\mathrm{E}(1/(1-\theta) \mid x)}{\mathrm{E}(1/\{\theta(1-\theta)\} \mid x)} = \frac{B(\alpha + x, \beta + n - x - 1)}{B(\alpha + x - 1, \beta + n - x - 1)} = \frac{\alpha + x - 1}{\alpha + \beta + n - 2}.$$

9. (Adapted from Robert 2.43). Consider $x \mid \theta \sim \mathrm{N}(\theta, 1)$, $\theta \sim \mathrm{N}(0, 1)$ and the loss $L(\theta, d) = e^{3\theta^2/4}(\theta - d)^2$.

   (a) Show that the estimator that minimizes the Bayesian expected posterior loss in this case is $\delta^\pi(x) = 2x$. *Hint:* use results from #7.

   We have $\theta \mid X \sim \mathrm{N}\left(\frac{x}{2}, \frac{1}{2}\right)$. Also, taking $w(\theta) = e^{3\theta^2/4}$, we have the Bayes rule, $\delta^B(x) = \frac{E(\omega(\theta)\theta|x)}{E(\omega(\theta)|x)}$.
   We next find

$$\begin{aligned}\mathrm{E}(w(\theta)\theta \mid x) &= \int_{\mathbb{R}} \theta \exp\left(\frac{3\theta^2}{4}\right) \frac{1}{\sqrt{\pi}} \exp\left\{-\left(\theta - \frac{x}{2}\right)^2\right\} d\theta \\ &= 4x \exp\left(\frac{3x^2}{4}\right),\end{aligned}$$

   and

$$\mathrm{E}(w(\theta) \mid x) = 2 \exp\left(\frac{3x^2}{4}\right).$$

   Thus, $\delta^\pi(x) = 2x$.

   (b) Show that $\delta_0(x) = x$ dominates $\delta^\pi(x)$.

   We let $w(\theta) = e^{3\theta^2/4} > 0$ for $\theta \in \mathbb{R}$ and find

$$\begin{aligned}R(\delta^\pi(x), \theta) &= \mathrm{E}(w(\theta)(\theta - 2x)^2) \\ &= w(\theta)\mathrm{E}(\theta^2 - 4\theta x + 4x^2) \\ &= w(\theta)\{\theta^2 - 4\theta^2 + 4(1 + \theta^2)\} = w(\theta)(\theta^2 + 4),\end{aligned}$$

   and

$$\begin{aligned}R(\delta_0, \theta) &= \mathrm{E}(w(\theta)(\theta - x)^2) \\ &= w(\theta)\mathrm{E}(\theta^2 - 2\theta x + x^2) \\ &= w(\theta)\{\theta^2 - 2\theta^2 + (1 + \theta^2)\} = w(\theta).\end{aligned}$$

   Thus, $R(\delta_0(x), \theta) \leq R(\delta^\pi(x), \theta)$ for any $\theta \in \mathbb{R}$ and $R(\delta_0(x), \theta) < R(\delta^\pi(x), \theta)$ for any $\theta \neq 0$, which implies that δ_0 dominates δ^π.
   It can be shown that $r(\pi, \delta^\pi) = \infty$. Formal Bayes rules need not be admissible if their Bayes risks are infinite. The reaction of many Bayesians to such inadmissibility is "*So what? After all, admissibility is a frequentist criterion, and is hence suspect.*" See Section 4.8 Adimissibility of Bayes Rules and Long Run Evaluations of JB book for more.

10. Assume you have to guess a secret number θ. You know that θ is an integer. You can perform an experiment that would yield either the number before it or the number after it, with equal probability. You perform the experiment twice. More formally, let x_1 and x_2 be independent observations from $f(x = \theta - 1 \mid \theta) = f(x = \theta + 1 \mid \theta) = 1/2$. Consider the 0-1 loss function, i.e.,

$$L(\theta, d) = \begin{cases} 0 & \text{if } \theta = d, \\ 1 & \text{if } \theta \neq d. \end{cases}$$

(a) Find the risks $R(\theta, \delta)$ for the estimators $\delta_0(x_1, x_2) = (x_1 + x_2)/2$ and $\delta_1(x_1, x_2) = x_1 + 1$.

By the definition,

$$R(\delta_0, \theta) = \mathrm{E}\left[L\left(\frac{X_1 + X_2}{2}, \theta\right)\right] = 0P(X_1 \neq X_2) + 1P(X_1 = X_2) = \frac{1}{2},$$

since X_1 and X_2 are independent. Similarly,

$$R(\delta_1, \theta) = \mathrm{E}[L(X_1 + 1, \theta)] = 0P(X_1 = \theta - 1) + 1P(X_1 = \theta + 1) = \frac{1}{2}.$$

(b) Find the estimator $\delta^\pi(x_1, x_2)$ that minimizes the posterior expected loss.

We first find the posterior distribution of θ. We first consider the following two cases;

i. Suppose $x_1 \neq x_2$.

$$\pi(\theta \mid x_1 \neq x_2) = \begin{cases} 1, & \text{if } \theta = \frac{x_1 + x_2}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

In words, the posterior distribution has probability 1 at $\theta = (x_1 + x_2)/2$. Thus, in this case, $\delta^\pi(x_1, x_2) = \frac{x_1 + x_2}{2}$ since its posterior expected loss $\mathrm{E}(L(\delta^\pi, \theta) \mid x_1 \neq x_2) = 0$.

ii. Suppose $x_1 = x_2 = x$.

$$\begin{aligned}
&\pi(\theta = x + 1 \mid x_1 = x_2 = x) \\
&\quad = \frac{p(x_1 = x_2 = x \mid \theta = x + 1)\pi(\theta = x + 1)}{p(x_1 = x_2 = x \mid \theta = x - 1)\pi(\theta = x - 1) + p(x_1 = x_2 = x \mid \theta = x + 1)\pi(\theta = x + 1)} \\
&\quad = \frac{\frac{1}{4}\pi(\theta = x + 1)}{\frac{1}{4}\pi(\theta = x - 1) + \frac{1}{4}\pi(\theta = x + 1)} \\
&\quad = \frac{\pi(\theta = x + 1)}{\pi(\theta = x - 1) + \pi(\theta = x + 1)}.
\end{aligned}$$

And we have

$$\begin{aligned}
\pi(\theta = x - 1 \mid x_1 = x_2 = x) &= 1 - \pi(\theta = x + 1 \mid x_1 = x_2 = x) \\
&= \frac{\pi(\theta = x - 1)}{\pi(\theta = x - 1) + \pi(\theta = x + 1)}.
\end{aligned}$$

Thus,

$$\pi(\theta \mid x_1 = x_2 = x) = \begin{cases} \frac{\pi(\theta=x+1)}{\pi(\theta=x-1)+\pi(\theta=x+1)}, & \text{if } \theta = x+1, \\ \frac{\pi(\theta=x-1)}{\pi(\theta=x-1)+\pi(\theta=x+1)}, & \text{if } \theta = x-1, \\ 0, & \text{otherwise.} \end{cases}$$

In words, the posterior supports two values, $\{x+1, x-1\}$. Since x can be either of $\theta-1$ or $\theta+1$ with probability $1/2$, the posterior probabilities of θ being $x+1$ and $x-1$ are proportional to their prior probabilities. Under the 0-1 loss function, the posterior expected loss is; if $d = x+1$, $\rho(\pi, d \mid x_1, x_2) = \frac{\pi(\theta=x-1)}{\pi(\theta=x-1)+\pi(\theta=x+1)}$. If $d = x-1$, $\rho(\pi, d \mid x_1, x_2) = \frac{\pi(\theta=x+1)}{\pi(\theta=x-1)+\pi(\theta=x+1)}$. If $d \notin \{x-1, x+1\}$, $\rho(\pi, d \mid x_1, x_2) = 1$.

Therefore, $\delta^\pi(x_1, x_2)$ is

$$\delta^\pi(x_1, x_2) = \begin{cases} \frac{x_1+x_2}{2}, & \text{if } x_1 \neq x_2, \\ x+1, & \text{if } x_1 = x_2 = x \text{ and } \pi(\theta = x-1) < \pi(\theta = x+1), \\ x-1, & \text{if } x_1 = x_2 = x \text{ and } \pi(\theta = x-1) \geq \pi(\theta = x+1). \end{cases}$$

For each Bayes decision, you lose 0, $\pi(\theta = x-1)$ and $\pi(\theta = x+1)$, which are smaller than or equal to $1/2$.

11. Consider a point estimation problem in which you observe $x_1, \ldots, x_n$ as i.i.d. random variables of the Poisson distribution with parameter θ. Assume a squared error loss and a prior of the form $\theta \sim \text{Gamma}(\alpha, \beta)$.

    (a) Show that the Bayes estimator is $\delta^\pi(x) = a + b\bar{x}$ where $a > 0$, $b \in (0,1)$ and $\bar{x} = \sum_{i=1}^n x_i/n$. You may use the fact that the distribution of $\sum_{i=1}^n x_i$ is Poisson with parameter $n\theta$ without proof.

    From class, we know $\theta \mid \boldsymbol{x} \sim \text{Gamma}(\alpha + \sum x_i, \beta + n)$ and so $\delta^\pi(\boldsymbol{x}) = \text{E}(\theta \mid \boldsymbol{x}) = (\alpha + \sum x_i)/(\beta + n) = \frac{\alpha}{\beta+n} + \frac{n}{\beta+n}\bar{x}$ with $a = \frac{\alpha}{\beta+n}$ and $b = \frac{n}{\beta+n}$.

    (b) Find the MLE for θ (*Note:* to remind how to find MLEs, read Casella and Berger, Section 7.2.2 – see Def 7.2.4).

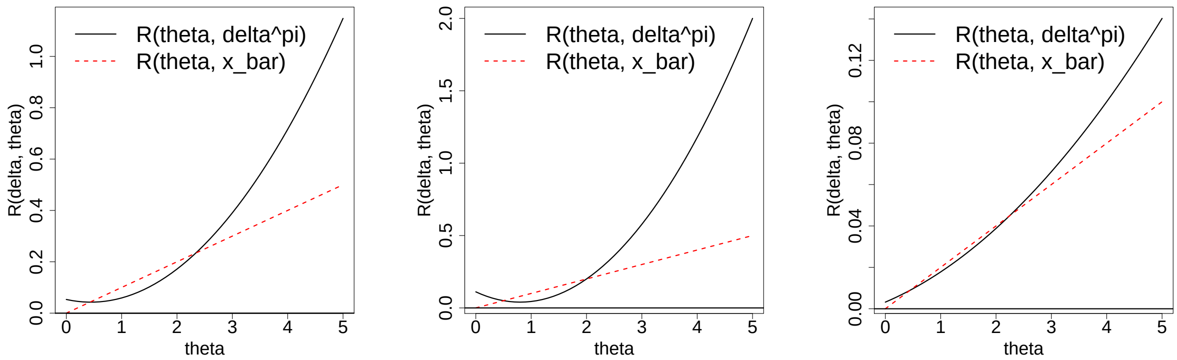
    $f(\boldsymbol{x} \mid \theta) \propto e^{-n\theta}\theta^t$ so $\log(f(\boldsymbol{x} \mid \theta)) \propto -n\theta + t\log(\theta)$. Then,

    $$\frac{\partial \log(f(\boldsymbol{x} \mid \theta))}{\partial \theta} = -n + \frac{t}{\theta} \quad \Rightarrow \hat{\theta} = t/n = \bar{x}.$$

    We also check

    $$\frac{\partial^2 \log(f(\boldsymbol{x} \mid \theta))}{\partial \theta^2} = -\frac{t}{\theta^2} < 0,$$

    which confirms that $\hat{\theta}$ achieves the unique maximum.



(a) $\alpha = \beta = 3$ and $n = 10$ (b) $\alpha = \beta = 5$ and $n = 10$ (c) $\alpha = \beta = 3$ and $n = 50$

Figure 3: Comparison of $R(\theta, \delta)$ for $\delta^\pi(x)$ and the MLE of θ

(c) Compute and graph the frequentist risks $R(\theta, \delta)$ for $\delta^\pi(x)$ and the MLE of θ.

We have $\mathrm{E}(\bar{x}) = \theta$, and $\mathrm{E}(\bar{x}^2) = \mathrm{Var}(\bar{x}) + (\mathrm{E}(\bar{x}))^2 = \frac{\theta}{n} + \theta^2$. Therefore,

$$R(\hat{\theta}, \theta) = \mathrm{E}((\bar{x} - \theta)^2) = \mathrm{E}(\bar{x}^2) - 2\theta\mathrm{E}(\bar{x}) + \theta^2 = \frac{\theta}{n},$$

while

$$\begin{aligned} R(\delta, \theta) &= \mathrm{E}((a + b\bar{x} - \theta)^2) = \mathrm{E}(a^2 + 2a(b\bar{x} - \theta) + (b\bar{x} - \theta)^2) \\ &= (b-1)^2\theta^2 + \left[2a(b-1) + \frac{b^2}{n}\right]\theta + a^2 \\ &= \frac{1}{(\beta + n)^2}\left(\beta^2\theta^2 + (-2\alpha\beta + n)\theta + \alpha^2\right). \end{aligned}$$

Figure 3 illustrates $R(\theta, \delta)$ for $\delta^\pi(x)$ and the MLE of θ with different values of α, β and n. Both are admissible.

(d) Compute the Bayes risk of $\delta^\pi(x)$.

Since $\theta \sim \mathrm{Gamma}(\alpha, \beta)$, we have $\mathrm{E}(\theta) = \alpha/\beta$ and $\mathrm{E}(\theta^2) = \alpha/\beta^2 + \alpha^2/\beta^2$.

$$\begin{aligned} r(\pi, \delta) &= \mathrm{E}\left[\frac{1}{(\beta + n)^2}\left(\beta^2\theta^2 + (-2\alpha\beta + n)\theta + \alpha^2\right)\right] \\ &= \frac{\alpha}{\beta(\beta + n)}. \end{aligned}$$

(e) Suppose that an investigator wants to collect a sample that is large enough that the Bayes risk after the experiment is half of the Bayes risk before the experiment. Find that sample size.

We find n such that

$$\frac{\alpha}{\beta(\beta + n)} < \frac{\alpha}{2\beta^2}.$$

We solve for n and get $n > \beta$.