ISyE 6412A **Theoretical Statistics** **Fall 2023**

HW #3 (due at Canvas midnight on Thursday, September 14 ET)

(There are 5 questions. The hints to Problem 1 and Problem 4(c) can be found on page #3 of this pdf file)

Problem 1. In Problem 1 of HW#1, we assume that $Y_1, \ldots, Y_n$ are iid normal $N(\theta, 1)$, and we want to guess θ when the loss function is given by $L(\theta, d) = (\theta - d)^2/(1+\theta^2)$. One of the proposed procedures is:

$$\delta_{2,n}(Y_1, \ldots, Y_n) = \frac{\bar{Y}_n + n^{-1}}{1 + n^{-1}}.$$

The purpose of this exercise is to show that this procedure is actually Bayes. Show that the procedure $\delta_{2,n}$ is Bayes relative to the prior density $\pi_a(\theta) = C_1(1+\theta^2)\phi(\theta - 1)$, where $\phi(y) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})$ is the standard normal $N(0,1)$ density and C_1 is a suitable constant. Your need *not* determine the C_1; it suffices to verify that the given function of θ have finite integrals, so that one knows such C_1 exist.

Answer: It is a simple calculus excise to show that C_1 exists. Indeed, it is not difficult to show that

$$1/C_1 = \int_{-\infty}^{\infty} (1+\theta^2)\phi(\theta - 1)d\theta = \mathbf{E}(1 + W^2) = 1 + (\mathbf{E}(W))^2 + Var(W)$$

where $W \sim N(1, 1)$. Hence $1/C_1 = 1 + 1^2 + 1 = 3$ and $C_1 = 1/3$ (this is not required).
Under our setting, for each $\mathbf{y} = (y_1, \ldots, y_n)$, a Bayes procedure shall minimize

$$\begin{aligned} h^*_\pi(\mathbf{y}, d) &= \int_{-\infty}^{\infty} L(\theta, d) f_\theta(y_1, \cdots, y_n)\pi_a(\theta)d\theta \\ &= \int_{-\infty}^{\infty} \frac{(\theta - d)^2}{1+\theta^2} f_\theta(y_1, \cdots, y_n) C_1(1+\theta^2)\phi(\theta - 1)d\theta \\ &= C_1 \int_{-\infty}^{\infty} (\theta - d)^2 f_\theta(y_1, \cdots, y_n)\phi(\theta - 1)d\theta, \end{aligned}$$

where the integral is just the Bayes risk when the loss function function is squared error loss, θ has a prior distribution $\pi(\theta) =$ pdf of $N(b = 1, \tau^2 = 1)$, and the Y_i's are iid $N(\theta, \sigma^2 = 1)$. We have shown in class that the integral is minimized at

$$\widehat{d} = \frac{\frac{\sum_i Y_i}{\sigma^2} + \frac{b}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} = \frac{\frac{n\bar{Y}}{1^2} + \frac{1}{1^2}}{\frac{n}{1^2} + \frac{1}{1^2}} = \frac{\bar{Y} + n^{-1}}{1 + n^{-1}},$$

which is exactly $\delta_{2,n}$. Thus, $\delta_{2,n}$ minimizes $h^*_\pi(\mathbf{y}, d)$ over D for all $\mathbf{y} \in S$. Hence, it is Bayes relative to the prior density $\pi_a(\theta)$. $\square$

Problem 2. Suppose $Y_1, Y_2, \ldots, Y_n$ ($n \geq 2$) are independent and identically distributed (iid) with a Uniform$[0, \theta]$, and consider a Parato prior distribution, $\theta \sim PA(\alpha, \beta)$, i.e., θ has a prior density

$$\pi(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}}, \qquad \text{for } \theta \geq \beta,$$

and mean $\mathbf{E}(\theta) = \alpha\beta/(\alpha - 1)$, for some known $\alpha > 1$ and $\beta > 0$.

(a) Show that the posterior distribution of θ has a Parato $PA(\alpha^*, \beta^*)$ distribution with $\alpha^* = \alpha + n$ and $\beta^* = \max\{\beta, Y_{(n)}\}$, where $Y_{(n)} = \max(Y_1, \ldots, Y_n)$.

(b) Find the Bayes procedure $\delta^*_B(Y)$ under the "squared error loss" function $L(\theta, d) = (\theta - d)^2$.

(c) Find the Bayes procedure $\delta^*_C(Y)$ under the "absolute error loss" function $L(\theta, d) = |\theta - d|$.

[*Hints for (a):* the domain of θ also includes $Y_1 \leq \theta, \ldots, Y_n \leq \theta$, which is equivalent to $\theta \geq Y_{(n)}$.]

Answer: (a) First, the joint distribution of θ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is

$$\begin{aligned} f(\theta, y_1, \cdots, y_n) &= \pi(\theta) f_\theta(y_1) \cdots f_\theta(y_n) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} 1\{\theta \geq \beta\} (\frac{1}{\theta} 1\{y_1 \leq \theta\}) \cdots (\frac{1}{\theta} 1\{y_n \leq \theta\}) \\ &= \frac{\alpha\beta^\alpha}{\theta^{\alpha+1+n}} 1\{\theta \geq \max(\beta, y_{(n)})\} \\ &= \frac{\alpha\beta^\alpha}{\theta^{\alpha^*+1}} 1\{\theta \geq \beta^*\}, \end{aligned}$$

where $\alpha^* = \alpha + n$ and $\beta^* = \max\{\beta, Y_{(n)}\}$.
Second, the marginal distribution of $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is

$$\begin{aligned} m(y_1, \cdots, y_n) &= \int_\Omega f(\theta, y_1, \cdots, y_n) d\theta \\ &= \int_0^\infty \frac{\alpha\beta^\alpha}{\theta^{\alpha^*+1}} 1\{\theta \geq \beta^*\} d\theta \\ &= \alpha\beta^\alpha \int_{\beta^*}^\infty \frac{1}{\theta^{\alpha^*+1}} d\theta \\ &= \alpha\beta^\alpha \Big(-\frac{1}{\alpha^*} \frac{1}{\theta^{\alpha^*}} \Big|_{\theta=\beta^*}^{\theta=\infty} \Big) = \alpha\beta^\alpha \frac{1}{\alpha^*(\beta^*)^{\alpha^*}}. \end{aligned}$$

Third, the posterior distribution of θ given $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is

$$\begin{aligned} \pi(\theta | y_1, \cdots, y_n) &= \frac{f(\theta, y_1, \cdots, y_n)}{m(y_1, \cdots, y_n)} \\ &= \frac{\frac{\alpha\beta^\alpha}{\theta^{\alpha^*+1}} 1\{\theta \geq \beta^*\}}{\alpha\beta^\alpha \frac{1}{\alpha^*(\beta^*)^{\alpha^*}}} \\ &= \frac{\alpha^*(\beta^*)^{\alpha^*}}{\theta^{\alpha^*+1}} 1\{\theta \geq \beta^*\}, \end{aligned}$$

which is exactly the pdf of the Parato $PA(\alpha^*, \beta^*)$ distribution.
(b) Under the squared error loss function, the Bayes procedure is

$$\begin{aligned} \delta_B^*(\mathbf{Y}) &= \text{ mean of the posterior distribution } \pi(\theta | y_1, \cdots, y_n) \\ &= \text{ mean of Parato } PA(\alpha^*, \beta^*) \\ &= \alpha^* \beta^* / (\alpha^* - 1) \\ &= \frac{\alpha + n}{\alpha + n - 1} \max\{\beta, Y_{(n)}\}. \end{aligned}$$

(c) Under the absolute error loss function, the Bayes procedure is

$$\begin{aligned} \delta_C^*(\mathbf{Y}) &= \text{ median of the posterior distribution } \pi(\theta | y_1, \cdots, y_n) \\ &= \text{ median of Parato } PA(\alpha^*, \beta^*) \end{aligned}$$

For a Parato prior distribution, $\theta \sim PA(\alpha, \beta)$, what is its median m? This requires us to find m such that

$$\begin{aligned} \frac{1}{2} &= \int_{-\infty}^m \pi(\theta) d\theta \\ &= \int_\beta^m \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} d\theta \qquad \text{(when } m > \beta) \\ &= -\frac{\beta^\alpha}{\theta^\alpha} \Big|_{\theta=\beta}^{\theta=m} \\ &= 1 - \Big(\frac{\beta}{m}\Big)^\alpha, \end{aligned}$$

and thus $\left(\frac{\beta}{m}\right)^\alpha = \frac{1}{2}$ and $m = 2^{1/\alpha}\beta$. Likewise, a Parato prior distribution $PA(\alpha^*, \beta^*)$ has median $2^{1/\alpha^*}\beta^*$. Hence, under the absolute error loss function, the Bayes procedure is

$$
\begin{aligned}
\delta_C^*(\mathbf{Y}) &= 2^{1/\alpha^*}\beta^* \\
&= 2^{1/(\alpha+n)} \max\{\beta, Y_{(n)}\}.
\end{aligned}
$$

[For the purpose of easy understanding, you may rewrite $2^{1/(\alpha+n)} = \exp\big(\log 2/(\alpha+n)\big) = \exp\big(\frac{\log 2}{\alpha+n}\big)$.] □

Problem 3. (**LINEX Loss. Problem 7.65 on page 367 of our text**). Suppose Ω can be labeled according to the values of a real parameter θ, and the decisions are $D = \{d : -\infty < d < \infty\}$, representing guesses as to the true value of θ. The loss function is the LINEX (LINear-EXponential) loss given by

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

where c is a positive constant. The LINEX loss was investigated by Zellner (1986), and can handle asymmetries in a smooth way: as the constant c varies, the loss function varies from very asymmetric to almost symmetric. We also assume that the prior density on $\Omega = \{\theta : -\infty < \theta < \infty\}$ is $\pi(\theta)$.

(a) For $c = 0.2, 0.5, 1$, plot $L(\theta, d)$ as a function of $d - \theta$. Feel free to use any computer software.

Answer: Plot $L(\theta, d)$ by yourself. □

(b) *"No data problem."* Suppose we had to make a guess of θ based on no observations, but merely on the given prior law $\pi(\theta)$. Show that you minimize your expected loss $\mathbf{E}(L(\theta, d)) = \int_{-\infty}^{\infty} L(\theta, d)\pi(\theta)d\theta$ by guessing θ as $\delta_b = -\frac{1}{c}\log\int_{-\infty}^{\infty} e^{-c\theta}\pi(\theta)d\theta = -\frac{1}{c}\log \mathbf{E}(e^{-c\theta})$.

Answer: We want to make a decision that minimize the expected loss

$$
\begin{aligned}
\bar{h}_\pi(d) &= \mathbf{E}(L(\theta, d)) = \int_{-\infty}^{\infty} L(\theta, d)\pi(\theta)d\theta \\
&= \int_{-\infty}^{\infty} \Big(e^{c(d-\theta)} - c(d-\theta) - 1\Big)\pi(\theta)d\theta \\
&= e^{cd}\int_{-\infty}^{\infty} e^{-c\theta)}\pi(\theta)d\theta - cd + c\int_{-\infty}^{\infty} \theta\pi(\theta)d\theta - 1 \\
&= e^{cd}\mathbf{E}(e^{-c\theta}) - cd + c\mathbf{E}(\theta) - 1 \qquad (1)
\end{aligned}
$$

Differentiating this with respect to d and setting it equal to 0, we have

$$0 = \frac{\partial \bar{h}_\pi(d)}{\partial d} = ce^{cd}\mathbf{E}(e^{-c\theta}) - c,$$

and thus

$$\widehat{d} = -\frac{1}{c}\log \mathbf{E}(e^{-c\theta}) = -\frac{1}{c}\log\int_{-\infty}^{\infty} e^{-c\theta}\pi(\theta)d\theta.$$

In addition, taking the second order derivatives yields that

$$\frac{\partial^2 \bar{h}_\pi(d)}{\partial d^2} = c^2 e^{cd}\mathbf{E}(e^{-c\theta}) \geq 0,$$

since $e^{-c\theta} \geq 0$. Hence, $\bar{h}_\pi(d)$ is a convex function of d and thus $\widehat{d} = -\frac{1}{c}\log \mathbf{E}(e^{-c\theta})$ is the unique solution to minimize (1). Thus $\delta_b = -\frac{1}{c}\log \mathbf{E}(e^{-c\theta})$. □

(c) (2 pts). *"Problem with data."* Now suppose that we also observe data $\mathbf{Y}$ which has distribution function $F(\mathbf{y}|\theta)$. Show that the Bayes procedure is given by $\delta_c(\mathbf{Y}) = -\frac{1}{c}\log \mathbf{E}(e^{-c\theta}|\mathbf{Y})$.

Answer: This is a simple application of "No data principal." If you want to work out the details, note that the Bayes procedure can be obtained by minimizing the posterior risk

$$h_\pi(\mathbf{y}, d) = \mathbf{E}(L(\theta, d)|\mathbf{Y} = \mathbf{y}) = \int_{-\infty}^{\infty} L(\theta, d)\pi(\theta|\mathbf{y})d\theta,$$

and similar arguments in part (b) will work in part (c) except replacing the prior distribution $\pi(\theta)$ by the posterior distribution $\pi(\theta|\mathbf{y})$. □

(d) Let $Y_1, \ldots, Y_n$ be iid $N(\theta, \sigma^2)$, where σ^2 is known, and suppose that θ has the so-called "noninformative" prior $\pi(\theta) \equiv 1$. Show that the Bayes procedure with respect to LINEX loss is given by $\delta_d(\bar{Y}) = \bar{Y} - (c\sigma^2/(2n))$. See the remarks/hints at the end of this question.

Answer: Note that $\bar{Y}_n \sim N(\theta, \frac{\sigma^2}{n})$ and $f_\theta(\bar{y}) = \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}}\exp(-\frac{(\bar{y}-\theta)^2}{2\sigma^2/n})$. When $\pi(\theta) \equiv 1$, the marginal distribution

$$m(\bar{y}) = \int_{-\infty}^{\infty} f_\theta(\bar{y})\pi(\theta)d\theta = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}}\exp(-\frac{(\bar{y}-\theta)^2}{2\sigma^2/n})d\theta \equiv 1!$$

Here some students might have difficulty to see that $f_\theta(\bar{y})$ is also a normal density as a function of θ. If we treat $m(\bar{y})$ as a posterior density (rigorously speaking, this is not a well-defined density over $(-\infty, \infty)$), then the posterior distribution

$$\pi(\theta|\bar{y}) = f_\theta(\bar{y})\pi(\theta)/m(\bar{y}) = \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}}\exp(-\frac{(\bar{y}-\theta)^2}{2\sigma^2/n}).$$

In other words, the posterior distribution of θ conditional on $\bar{Y}$ is $N(\bar{Y}, \frac{\sigma^2}{n})$. By part (c), the Bayes procedure is

$$\begin{aligned}
\delta_d(\bar{Y}) &= -\frac{1}{c}\log \mathbf{E}(e^{-c\theta}|Y) \\
&= -\frac{1}{c}\log \int_{-\infty}^{\infty} e^{-c\theta}\frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}}\exp(-\frac{(\bar{Y}-\theta)^2}{2\sigma^2/n})d\theta \\
&= -\frac{1}{c}\log\left\{e^{\frac{c^2\sigma^2}{2n} - c\bar{Y}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}}\exp(-\frac{[\theta-(\bar{Y}-c\sigma^2/n)]^2}{2\sigma^2/n})d\theta\right\} \\
&= -\frac{1}{c}\log\left\{e^{\frac{c^2\sigma^2}{2n} - c\bar{Y}}\right\} \\
&= \bar{Y} - \frac{c\sigma^2}{2n}.
\end{aligned}$$

□

(e) Under the assumption of part (d), calculate the posterior expected loss for both $\delta_d(\bar{Y})$ and $\delta_e(\bar{Y}) = \bar{Y}$ using LINEX loss. Recall that for a given procedure $\delta(\mathbf{Y})$, the posterior expected loss is defined as $h_\pi(\mathbf{y}, \delta(\mathbf{y}))$ where $h_\pi(\mathbf{y}, d) = \int_{-\infty}^{\infty} L(\theta, d)\pi(\theta|\mathbf{y})d\theta$.

Answer: Since $\pi(\theta|\bar{Y})$ is just the density function of $N(\bar{Y}, \frac{\sigma^2}{n})$, the posterior risk of $\delta_d = \delta_d(\bar{Y}) = \bar{Y} - \frac{c\sigma^2}{2n}$ is

$$\begin{aligned}
h_\pi(\mathbf{y}, \delta_d) &= \mathbf{E}(e^{c(\delta_d-\theta)} - c(\delta_d - \theta) - 1|\mathbf{y}) \\
&= \int_{-\infty}^{\infty}\left\{e^{c(\bar{Y}-\frac{c\sigma^2}{2n}-\theta)} - c(\bar{Y} - \frac{c\sigma^2}{2n} - \theta) - 1\right\}\frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}}\exp(-\frac{(\bar{Y}-\theta)^2}{2\sigma^2/n})d\theta
\end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \left\{ e^{c\frac{\sigma}{\sqrt{n}}z - \frac{c^2\sigma^2}{2n}} - c\frac{\sigma}{\sqrt{n}}z + \frac{c^2\sigma^2}{2n} - 1 \right\} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \qquad (\text{letting } \ z = \frac{\bar{Y}-\theta}{\sigma/\sqrt{n}}) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-c\sigma/\sqrt{n})^2}{2}} dz - c\frac{\sigma}{\sqrt{n}} \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \frac{c^2\sigma^2}{2n} - 1 \\
&= 1 - 0 + \frac{c^2\sigma^2}{2n} - 1 \\
&= \frac{c^2\sigma^2}{2n}.
\end{aligned}$$

Similarly, for $\delta_e(\bar{Y}) = \bar{Y}$, we have

$$\begin{aligned}
h_\pi(\mathbf{y}, \delta_e) &= \mathbf{E}(e^{c(\delta_e - \theta)} - c(\delta_e - \theta) - 1 | Y) \\
&= \int_{-\infty}^{\infty} \left\{ e^{c(\bar{Y}-\theta)} - c(\bar{Y}-\theta) - 1 \right\} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} \exp(-\frac{(\bar{Y}-\theta)^2}{2\sigma^2/n}) d\theta \\
&= \int_{-\infty}^{\infty} \left\{ e^{c\frac{\sigma}{\sqrt{n}}z} - c\frac{\sigma}{\sqrt{n}}z - 1 \right\} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \qquad (\text{letting } \ z = \frac{\bar{Y}-\theta}{\sigma/\sqrt{n}}) \\
&= e^{\frac{c^2\sigma^2}{2n}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-c\sigma/\sqrt{n})^2}{2}} dz - c\frac{\sigma}{\sqrt{n}} \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - 1 \\
&= e^{\frac{c^2\sigma^2}{2n}} - 1.
\end{aligned}$$

It is interesting that both $h_\pi(\mathbf{y}, \delta_d)$ and $h_\pi(\mathbf{y}, \delta_e)$ do not depend on $\mathbf{Y}$. Since $u \le e^u - 1$, we can see that $h_\pi(\mathbf{y}, \delta_d) \le h_\pi(\mathbf{y}, \delta_e)$ for every $\mathbf{Y} = \mathbf{y}$. This is consistent with the fact that a Bayesian procedure $\delta_\pi(\mathbf{Y})$ will minimize the posterior expected loss $h_\pi(\mathbf{y}, \delta_\pi(Y))$ for every $\mathbf{y} \in S$. □

(f) Repeat part (e) using squared error loss, i.e., calculate the posterior expected loss for both $\delta_d(\bar{Y})$ and $\delta_e(\bar{Y})$ when the loss function is $L(\theta, d) = (\theta - d)^2$.

[*Remarks/Hints for (d):* Rigorously speaking, $\pi(\theta) \equiv 1$ is not a probability density on $\Omega = (-\infty, \infty)$, but for our purpose, here we pretend that it can be thought of as a valid prior distribution. To simplify your computation, you can assume that you just observe "one new observation" $\bar{Y}_n \sim N(\theta, \frac{\sigma^2}{n})$ (this simplification is reasonable because $\bar{Y}_n$ is a sufficient statistic which will be discussed later). For a data $\bar{Y}$, what is the marginal distribution $m(\bar{y}) = \int_{-\infty}^{\infty} f_\theta(\bar{y})\pi(\theta)d\theta$ when $\pi(\theta) \equiv 1$? What is the posterior distribution $\pi(\theta|\bar{y}) = f_\theta(\bar{y})\pi(\theta)/m(\bar{y})$? Is it well-defined?]

Answer: For the squared error loss , we have

$$\begin{aligned}
h_\pi(\mathbf{y}, \delta_d) &= \mathbf{E}((\delta_d - \theta)^2 | \mathbf{Y} = \mathbf{y}) \\
&= \int_{-\infty}^{\infty} (\bar{Y} - \frac{c\sigma^2}{2n} - \theta)^2 \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} \exp(-\frac{(\bar{Y}-\theta)^2}{2\sigma^2/n}) d\theta \\
&= \int_{-\infty}^{\infty} (\frac{\sigma}{\sqrt{n}}z - \frac{c\sigma^2}{2n})^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \qquad (\text{letting } \ z = \frac{\bar{Y}-\theta}{\sigma/\sqrt{n}}) \\
&= \mathbf{E}(\frac{\sigma}{\sqrt{n}}Z - \frac{c\sigma^2}{2n})^2 \qquad (\text{letting } \ Z \sim N(0,1)) \\
&= \frac{\sigma^2}{n} + \frac{c^2\sigma^4}{4n^2} \qquad (\text{using the facts } \ \mathbf{E}(Z) = 0, \mathbf{E}(Z^2) = Var(Z) = 1).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
h_\pi(\mathbf{y}, \delta_e) &= \mathbf{E}((\delta_e - \theta)^2 | \mathbf{Y} = \mathbf{y}) \\
&= \int_{-\infty}^{\infty} (\bar{Y} - \theta)^2 \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} \exp(-\frac{(\bar{Y}-\theta)^2}{2\sigma^2/n}) d\theta \\
&= \int_{-\infty}^{\infty} (\frac{\sigma}{\sqrt{n}}z)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \qquad (\text{letting } \ z = \frac{\bar{Y}-\theta}{\sigma/\sqrt{n}})
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sigma^2}{n}\mathbf{E}(Z^2) \qquad \text{(letting } \ Z \sim N(0,1))\\
&= \frac{\sigma^2}{n} \qquad \text{(using the facts } \ \mathbf{E}(Z) = 0, \mathbf{E}(Z^2) = Var(Z) = 1).
\end{aligned}
$$

□

Problem 4. Suppose that a random variable Y has a Gamma$(q, 1/\theta)$ distribution with density $f_\theta(y) = \frac{\theta^q y^{q-1} e^{-\theta y}}{\Gamma(q)}$, for $y \geq 0$, where $\Gamma(q) = \int_0^\infty y^{q-1}e^{-y}dy = (q-1)\Gamma(q-1)$ and q is positive and known. Here $S = \{y : y \geq 0\}, \Omega = \{\theta : \theta \geq 0\}, D = \{d : d \geq 0\}$. Under a Bayesian setting, suppose that the prior distribution θ is Gamma(α, β), i.e., θ has a prior density $\pi(\theta) = \frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha}$ for $\theta \geq 0$.

(a) Show that the posterior distribution of θ has a Gamma(α', β') distribution with $\alpha' = q+\alpha$ and $\beta' = \frac{\beta}{1+\beta y}$. This implies that $\pi(\theta|y) = \frac{\pi(\theta)f_\theta(y)}{\int_0^\infty \pi(\theta)f_\theta(y)d\theta}$ can be written as $\pi(\theta|y) = \frac{\theta^{\alpha'-1}e^{-\theta/\beta'}}{\Gamma(\alpha')(\beta')^{\alpha'}}$.

(b) (**Point Estimation**) Find the Bayes procedure $\delta_{bayes}(Y)$ under the loss function $L(\theta, d) = (\theta - d)^2$.

(c) (**Confidence Interval**) In the fixed-width confidence interval estimation setting, suppose one wants to estimate θ by the interval $[d-2, d+2]$ for some decision $d \geq 2$. This corresponds to estimate θ under a new binary loss function $L^*(\theta, d) = \begin{cases} 1 & \text{if } |\theta - d| > 2; \\ 0 & \text{if } |\theta - d| < 2. \end{cases}$ That is, the loss is the same for all "bad" decisions (those which mis-estimate θ by more than 2 units) and is 0 for all "correct" decisions. Show that the Bayes procedure is given by

$$
\begin{aligned}
\delta^*(Y) &= \begin{cases} 2, & \text{if } q+\alpha \leq 1; \\ \text{the solution } d \text{ of } \ \pi(d+2|Y) = \pi(d-2|Y), & \text{if } q+\alpha > 1. \end{cases}\\
&= \begin{cases} 2, & \text{if } q+\alpha \leq 1; \\ -2 + 4/[1 - \exp(-\frac{4(Y+1/\beta)}{\alpha+q-1})], & \text{if } q+\alpha > 1. \end{cases}
\end{aligned}
$$

In other words, the Bayes confidence interval with fixed-width of 4 is $[\delta^*(Y) - 2, \delta^*(Y) + 2]$.

Answer: (a) The posterior density of θ, given that $Y = y$, is

$$
\begin{aligned}
\pi(\theta|y) &= \frac{\pi(\theta)f_\theta(y)}{\int_0^\infty (\text{same})d\theta}\\
&= \frac{\frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)\beta^\alpha}\frac{\theta^q y^{q-1}e^{-\theta y}}{\Gamma(q)}}{\int_0^\infty (\text{same})d\theta}\\
&= \frac{\theta^{q+\alpha-1}e^{-(\frac{1}{\beta}+y)\theta}}{\int_0^\infty (\text{same})d\theta} \qquad \text{(since the constant and the } y \text{ terms canceled out)}\\
&= \text{Gamma}(\alpha', \beta') \qquad \text{with } \ \alpha' = q+\alpha \quad \text{and} \quad \beta' = \frac{1}{1/\beta + y} = \frac{\beta}{1+\beta y}.
\end{aligned}
$$

(b) Under the loss function $L(\theta, d) = (\theta - d)^2$, by the previous problem, the Bayes procedure is given by

$$
\begin{aligned}
\delta_{bayes}(Y) &= \text{mean of posterior law} \pi(\theta|y)\\
&= \text{mean of Gamma}(\alpha', \beta')\\
&= \alpha'\beta'\\
&= (q+\alpha)\frac{\beta}{1+\beta y}\\
&= \frac{\beta(q+\alpha)}{1+\beta y}
\end{aligned}
$$

(c) From the hints, it is straightforward to see that

$$\delta^*(Y) \quad = \quad \begin{cases} 2, & \text{if } q+\alpha \le 1; \\ \text{the solution } d \text{ of } \ \pi(d+2|y) = \pi(d-2|y), & \text{if } q+\alpha > 1. \end{cases}$$

The equation of the second line happens to be solvable in this case. It becomes, for $q + \alpha > 1$,

$$\begin{aligned} & (d+2)^{\alpha'-1} e^{-(d+2)/\beta'} = (d-2)^{\alpha'-1} e^{-(d-2)/\beta'} \\ \Longleftrightarrow \quad & (d+2)^{q+\alpha-1} e^{-(\frac{1}{\beta}+y)(d+2)} = (d-2)^{q+\alpha-1} e^{-(\frac{1}{\beta}+y)(d-2)} \\ \Longleftrightarrow \quad & e^{-4(\frac{1}{\beta}+y)} = \Big(\frac{d-2}{d+2}\Big)^{q+\alpha-1} = \Big(1 - \frac{4}{d+2}\Big)^{q+\alpha-1} \\ \Longleftrightarrow \quad & e^{-\frac{4(\frac{1}{\beta}+y)}{q+\alpha-1}} = 1 - \frac{4}{d+2} \\ \Longleftrightarrow \quad & d = -2 + 4/[1 - \exp(-\frac{4(y+\frac{1}{\beta})}{\alpha+q-1})]. \end{aligned}$$

$\square$

Problem 5 (Bayes with constant risk). Assume that $Y_1, \ldots, Y_n$ are iid Bernoulli(θ) with $0 \le \theta \le 1$, i.e., $\mathbf{P}_\theta(Y_i = y) = \theta^y(1-\theta)^{1-y}$ for $y = 0$ or 1, and suppose that we want to estimate θ under the square error loss function $L(\theta, d) = (\theta - d)^2$. We have shown in class when θ has a prior Beta(α, β) distribution, i.e.,

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad 0 \le \theta \le 1$$

for two pre-specified constants $\alpha > 0, \beta > 0$, the Bayes procedure is given by

$$\delta_{\alpha,\beta,n}(Y_1, \ldots, Y_n) = \frac{\alpha + \sum_{i=1}^n Y_i}{\alpha+\beta+n} = \frac{\alpha+\beta}{\alpha+\beta+n} \times \frac{\alpha}{\alpha+\beta} + \frac{n}{\alpha+\beta+n} \bar{Y}_n,$$

which is the weighted average of the prior mean $\alpha/(\alpha+\beta)$ and the sample mean $\bar{Y}_n = (Y_1 + \ldots + Y_n)/n$.

(a) Compute the risk function of the Bayes procedure $\delta_{\alpha,\beta,n}(Y_1, \ldots, Y_n)$.

(b) An interesting scenario is to choose a specific pair of $\alpha^* > 0$ and $\beta^* > 0$ in such a way that the corresponding Bayes procedure has constant risks when $\alpha = \alpha^*$ and $\beta = \beta^*$, i.e., the risk function in part (a) does not depend on θ. Find such a pair (α^*, β^*). [A check: $\alpha^* = \sqrt{n}/2$.]

(c) Calculate the Bayes risk of the Bayes procedure $\delta_{\alpha^*,\beta^*,n}$ in part (b) when θ has a prior distribution $\pi(\theta) = \text{Beta}(\alpha^*, \beta^*)$.

(d) Compute the risk function of the sample mean $\delta_0 = \delta_0(Y_1, \ldots, Y_n) = \bar{Y}_n$, and calculate the Bayes risk of δ_0 when θ has a prior distribution $\pi(\theta) = \text{Beta}(\alpha^*, \beta^*)$.

(e) Show that the Bayes procedure $\delta_{\alpha^*,\beta^*,n}$ indeed has a smaller Bayes risk than the sample mean $\delta_0 = \bar{Y}_n$ when θ has a prior distribution $\pi(\theta) = \text{Beta}(\alpha^*, \beta^*)$, where α^* and β^* are determined in part (b).

Answer: (a) For the procedure $\delta_{\alpha,\beta,n}$, its risk function is

$$\begin{aligned} R_{\delta_{\alpha,\beta,n}}(\theta) &= \mathbf{E}_\theta(\delta_{\alpha,\beta,n} - \theta)^2 \\ &= \Big[\mathbf{E}_\theta(\delta_{\alpha,\beta,n}) - \theta\Big]^2 + Var_\theta(\delta_{\alpha,\beta,n}) \\ &= \Big[\frac{\alpha+n\theta}{\alpha+\beta+n} - \theta\Big]^2 + \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} \\ &= \frac{[\alpha - (\alpha+\beta)\theta]^2 + n\theta(1-\theta)}{(\alpha+\beta+n)^2}. \end{aligned}$$

(b) Rewrite the risk function as a function of θ, we have

$$R_{\delta_{\alpha,\beta,n}}(\theta) = \frac{\alpha^2 + [n - 2\alpha(\alpha+\beta)]\theta + [(\alpha+\beta)^2 - n]\theta^2}{(\alpha+\beta+n)^2}.$$

Thus $R_{\delta_{\alpha,\beta,n}}(\theta)$ is a constant that does not depend on θ if and only if

$$2\alpha(\alpha+\beta) = n \qquad \text{and} \qquad (\alpha+\beta)^2 = n.$$

Solving these equations yields that $\alpha^* = \beta^* = \sqrt{n}/2$.
Note that by the theorem that a (unique) Bayes procedure that has constant risk is minimax, we can conclude that $\delta_{\alpha^*,\beta^*,n}$ is minimax (under the squared error loss function).
(c) From the above computations, we have

$$R_{\delta_{\alpha^*,\beta^*,n}}(\theta) = \frac{(\alpha^*)^2}{(\alpha^*+\beta^*+n)^2} = \frac{(\sqrt{n}/2)^2}{(\sqrt{n}+n)^2} = \frac{1}{4(1+\sqrt{n})^2}$$

and thus the Bayes risk of $\delta_{\alpha^*,\beta^*,n}$ is

$$r_{\delta_{\alpha^*,\beta^*,n}}(\pi) = \int_0^1 \frac{1}{4(1+\sqrt{n})^2}\pi(\theta)d\theta = \frac{1}{4(1+\sqrt{n})^2}\int_0^1 \pi(\theta)d\theta = \frac{1}{4(1+\sqrt{n})^2},$$

(d) The risk function of the sample mean is

$$R_{\delta_0}(\theta) = \mathbf{E}_\theta(\bar{Y}_n - \theta)^2 = \frac{\theta(1-\theta)}{n}.$$

and thus the Bayes risk of δ_0 is

$$\begin{aligned}
r_{\delta_0}(\pi) &= \int_0^1 \frac{\theta(1-\theta)}{n}\pi(\theta)d\theta = \int_0^1 \frac{\theta(1-\theta)}{n}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta \\
&= \frac{1}{n}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 \theta^{1+\alpha-1}(1-\theta)^{1+\beta-1}d\theta \\
&= \frac{1}{n}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} \\
&= \frac{\alpha\beta}{n(\alpha+\beta)(\alpha+\beta+1)}.
\end{aligned}$$

In particular, when $\alpha = \alpha^* = \sqrt{n}/2$ and $\beta = \beta^* = \sqrt{n}/2$, we have

$$r_{\delta_0}(\pi) = \frac{(\sqrt{n}/2)^2}{n\sqrt{n}(\sqrt{n}+1)} = \frac{1}{4\sqrt{n}(\sqrt{n}+1)}.$$

(e) It is clear that for all n,

$$r_{\delta_{\alpha^*,\beta^*,n}}(\pi) = \frac{1}{4(1+\sqrt{n})^2} < \frac{1}{4\sqrt{n}(1+\sqrt{n})} = r_{\delta_0}(\pi).$$

$\square$

Hint to Problem 1: You can show that a Bayes procedure shall minimize

$$h^*_\pi(\mathbf{y}, d) \quad = \quad \int_{-\infty}^{\infty} L(\theta, d) f_\theta(y_1, \cdots, y_n) \pi_a(\theta) d\theta = C_1 \int_{-\infty}^{\infty} (\theta - d)^2 f_\theta(y_1, \cdots, y_n) \phi(\theta - 1) d\theta,$$

where the integral is just the Bayes risk when the loss function function is squared error loss and θ has a prior distribution $N(\mu = 1, \tau^2 = 1)$. Which decision d minimizes this integral? Did we find such a minimizer d in class?

Hints to Problem 4(c): Note that a Gamma(α, β) random variable has mean $\alpha\beta$, see page 624 of our textbook for more properties of Gamma distribution. In part (c), we want to choose d to minimize

$$\int_0^{\infty} L^*(\theta, d) \pi(\theta|y) d\theta = 1 - \int_{d-2}^{d+2} \pi(\theta|y) d\theta. \tag{2}$$

Note that in this problem, $\pi(\theta|y)$ is continuous and *unimodal*; that is, there is a *mode* (possibly 0) such that $\pi(\theta|y)$ increases for $\theta < mode$ and decreases for $\theta > mode$. Then it is an easy calculus exercise to show that relation (2) is minimized by that value d^* for which $\pi(d^* - 2|y) = \pi(d^* + 2|y)$ if $\pi(0|y) < \pi(4|y)$, and by $d^* = 2$ otherwise (please feel free to use this fact if you cannot prove it or if you simply do not want to spend time to prove it! No penalty here!)

In addition, note that $\pi(0|y) = \infty$ if $\alpha' = q + \alpha < 1$ and $\pi(0|y) = 0$ if $\alpha' > 1$. Hence, $\pi(0|y) \geq \pi(4|y)$ if and only if $q + \alpha \leq 1$. What happens if $q + \alpha = 1$?