

Homework 3

STA-360-602

Due at 5:00 PM EDT on Friday January 21.

There is no reproducibility component to this homework, so you only need to upload this assignment to Gradescope. You do not need to submit your solution to the lab exercise since it's not worth any points.

General instructions for homeworks: Please follow the uploading file instructions according to the syllabus. You will give the commands to answer each question in its own code block, which will also produce plots that will be automatically embedded in the output file. Each answer must be supported by written statements as well as any code used. Your code must be completely reproducible and must compile.

Advice: Start early on the homeworks and it is advised that you not wait until the day of. While the professor and the TA's check emails, they will be answered in the order they are received and last minute help will not be given unless we happen to be free.

Commenting code Code should be commented. See the Google style guide for questions regarding commenting or how to write code <https://google.github.io/styleguide/Rguide.xml>. No late homework's will be accepted.

1. *Lab component* (0 points total) Please refer to module 2 and lab 3 and complete tasks 3—5. **This will not be graded as the entire solution is already posted. You will still be responsible for this material on the exam.**

- (a) (0) Task 3
- (b) (0) Task 4
- (c) (0) Task 5

Total points: Q1 (15) + Q2 (15) = 30 points total

2. (15 points total) *The Uniform-Pareto*

The goal of this problem is to continue getting more practice calculating the posterior distribution.

Suppose $a < x < b$. Consider the notation $I_{(a,b)}(x)$, where I denotes the indicator function. We define $I_{(a,b)}(x)$ to be the following:

$$I_{(a,b)}(x) = \begin{cases} 1 & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Let X be a random variable and let x be an observed value. Let

$$\begin{aligned} X = x \mid \theta &\sim \text{Uniform}(0, \theta) \\ \theta &\sim \text{Pareto}(\alpha, \beta), \end{aligned}$$

where $p(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}}I_{(\beta,\infty)}(\theta)$. Write out the likelihood $p(X = x \mid \theta)$. Then calculate the posterior distribution of $\theta \mid X = x$.

Q2 Ans:

- Since $X = x \mid \theta \sim \text{Uniform}(0, \theta)$, the likelihood function is

$$p(X = x \mid \theta) = \frac{1}{\theta}I_{(0,\theta)}(x)$$

- The posterior distribution is

$$\begin{aligned} p(\theta \mid X = x) &= p(X = x \mid \theta)p(\theta) \\ &= \frac{1}{\theta}I_{(0,\theta)}(x) \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}}I_{(\beta,\infty)}(\theta) \\ &= \frac{\alpha\beta^\alpha}{\theta^{\alpha+2}}I_{(\beta,\infty)}(\theta)I_{(0,\theta)}(x) \\ &= \frac{\alpha\beta^\alpha}{\theta^{\alpha+2}}I(\theta > \max(x, \beta)) \\ &\propto \frac{1}{\theta^{\alpha+2}}I(\theta > \max(x, \beta)) \end{aligned}$$

Therefore, $\theta \mid X = x \sim \text{Pareto}(\alpha + 1, \max(x, \beta))$.

3. (15 points total) *The Bayes estimator or Bayes rule*

The goal of this problem is to practice a similar problem that we considered in Module 2, where we derived the Bayes rule under squared error loss and found the result was the posterior mean.

- (a) (5 pts) Find the Bayes estimator (or Bayes rule) when the loss function is $L(\theta, \delta(x)) = c(\theta - \delta(x))^2$, where $c > 0$ is a constant.

Q3a Ans:

- The posterior risk is

$$\begin{aligned} p(\delta(x), X) &= \mathbb{E}(L(\theta, \delta(x)) \mid X) \\ &= \mathbb{E}(c(\theta - \delta(x))^2 \mid X) \\ &= c\mathbb{E}(\theta^2 - 2\theta\delta(x) + \delta(x)^2 \mid X) \\ &= c\mathbb{E}(\theta^2 \mid X) - 2c\delta(x)\mathbb{E}(\theta \mid X) + c\delta(x)^2 \end{aligned}$$

- Because $c > 0$ and is a constant, the posterior risk is a convex function of $\delta(x)$, which has a unique global minimum.
- Now, suppose that the posterior risk achieves global minimum when $\delta(x) = \mathbb{E}(\theta \mid X) + k$. Then,

$$\begin{aligned} p(\delta(x), X) &= c\mathbb{E}(\theta^2 \mid X) - 2c(\mathbb{E}(\theta \mid X) + k)\mathbb{E}(\theta \mid X) + c(\mathbb{E}(\theta \mid X) + k)^2 \\ &= c\mathbb{E}(\theta^2 \mid X) - 2c(\mathbb{E}(\theta \mid X))^2 - 2ck\mathbb{E}(\theta \mid X) + c(\mathbb{E}(\theta \mid X))^2 + 2ck\mathbb{E}(\theta \mid X) + ck^2 \\ &= c\mathbb{E}(\theta^2 \mid X) - c(\mathbb{E}(\theta \mid X))^2 + ck^2 \\ &\geq c\mathbb{E}(\theta^2 \mid X) - c(\mathbb{E}(\theta \mid X))^2 \quad \text{when } k = 0 \end{aligned}$$

- Hence, the Bayes rule is the posterior mean, i.e., $\delta(x) = \mathbb{E}(\theta \mid X)$.

- (b) (10 pts) Derive the Bayes estimator (or Bayes rule) when $L(\theta, \delta(x)) = w(\theta)(g(\theta) - \delta(x))^2$. Do so without writing any integrals. Note that you can write $\rho(\pi, \delta(x)) = E[L(\theta, \delta(x)) | X]$. You may assume that $w(\theta) > 0$. Don't forget to prove or state why the Bayes rule(s) are unique.

Q3b Ans:

- The posterior risk is

$$\begin{aligned} p(\delta(x), X) &= \mathbb{E}(L(\theta, \delta(x)) | X) \\ &= \mathbb{E}(w(\theta)(g(\theta) - \delta(x))^2 | X) \\ &= \mathbb{E}(w(\theta)(g(\theta)^2 - 2g(\theta)\delta(x) + \delta(x)^2) | X) \\ &= \mathbb{E}(w(\theta)g(\theta)^2 | X) - 2\delta(x)\mathbb{E}(w(\theta)g(\theta) | X) + \delta(x)^2\mathbb{E}(w(\theta) | X) \end{aligned}$$

because $\delta(x)$ is fixed when conditioning on $X = x$.

- As before, this posterior risk is a convex function of $\delta(x)$, which has a unique global minimum.
- Taking the partial derivative with respect to $\delta(x)$,

$$\begin{aligned} \frac{\partial p(\delta(x), X)}{\partial \delta(x)} &= \frac{\partial (\mathbb{E}(w(\theta)g(\theta)^2 | X) - 2\delta(x)\mathbb{E}(w(\theta)g(\theta) | X) + \delta(x)^2\mathbb{E}(w(\theta) | X))}{\partial \delta(x)} \\ &= 0 - 2\mathbb{E}(w(\theta)g(\theta) | X) + 2\delta(x)\mathbb{E}(w(\theta) | X) \end{aligned}$$

Setting the first-order derivative to zero, we have

$$\delta(x) = \frac{\mathbb{E}(w(\theta)g(\theta) | X)}{\mathbb{E}(w(\theta) | X)}$$

The convexity of the $p(\delta(x), X)$ as a function of $\delta(x)$ guarantees that this estimator uniquely minimizes the posterior risk, so this is the Bayes rule.