

# Homework 2

STA-360

Due at 5:00 PM EDT on Friday, 3 September

Total pts: 10 (reproducibility) + 30 (Q1) + 20 (Q2) + 40 (Q3) = 100

**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* (30 points total) Please refer to lab 2 and complete tasks 3—5.
  - (a) (10) Task 3
  - (b) (10) Task 4
  - (c) (10) Task 5
2. (20 points total) *The Exponential-Gamma Model* We write  $X \sim \text{Exp}(\theta)$  to indicate that  $X$  has the Exponential distribution, that is, its p.d.f. is

$$p(x|\theta) = \text{Exp}(x|\theta) = \theta \exp(-\theta x) \mathbf{1}(x > 0).$$

The Exponential distribution has some special properties that make it a good model for certain applications. It has been used to model the time between events (such as neuron spikes, website hits, neutrinos captured in a detector), extreme values such as maximum daily rainfall over a period of one year, or the amount of time until a product fails (lightbulbs are a standard example).

Suppose you have data  $x_1, \dots, x_n$  which you are modeling as i.i.d. observations from an Exponential distribution, and suppose that your prior is  $\theta \sim \text{Gamma}(a, b)$ , that is,

$$p(\theta) = \text{Gamma}(\theta|a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \mathbb{1}(\theta > 0).$$

- (a) (5) Derive the formula for the posterior density,  $p(\theta|x_{1:n})$ . Give the form of the posterior in terms of one of the most common distributions (Bernoulli, Beta, Exponential, or Gamma).
- (b) (5) Why is the posterior distribution a *proper* density or probability distribution function?
- (c) (5) Now, suppose you are measuring the number of seconds between lightning strikes during a storm, your prior is  $\text{Gamma}(0.1, 1.0)$ , and your data is

$$(x_1, \dots, x_8) = (20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0).$$

Plot the prior and posterior p.d.f.s. (Be sure to make your plots on a scale that allows you to clearly see the important features.)

- (d) (5) Give a specific example of an application where an Exponential model would be reasonable. Give an example where an Exponential model would NOT be appropriate, and explain why.
3. (40 points total) *Priors, Posteriors, Predictive Distributions (Hoff, 3.9)*  
 An unknown quantity  $Y | \theta$  has a Galenshore( $a, \theta$ ) distribution if its density is given by

$$p(y | \theta) = \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2}$$

for  $y > 0, \theta > 0, a > 0$ . Assume for now that  $a$  is known and  $\theta$  is unknown and a random variable. For this density,

$$E[Y] = \frac{\Gamma(a + 1/2)}{\theta \Gamma(a)}$$

and

$$E[Y^2] = \frac{a}{\theta^2}.$$

- (a) (10) Identify a class of conjugate prior densities for  $\theta$ . **Assume the prior parameters are  $c$  and  $d$ .** That is, state the distribution that  $\theta$  should have with parameters  $c, d$  such that the resulting posterior is conjugate. Plot a few members of this class of densities.
- (b) (5) Let  $Y_1, \dots, Y_n \stackrel{iid}{\sim} \text{Galenshore}(a, \theta)$ . Find the posterior distribution of  $\theta | y_{1:n}$  using a prior from your conjugate class.

(c) (10) Show that

$$\frac{p(\theta_a \mid y_{1:n})}{p(\theta_b \mid y_{1:n})} = \left( \frac{\theta_a}{\theta_b} \right)^{2(an+c)-1} e^{(\theta_b^2 - \theta_a^2)(d^2 + \sum y_i^2)},$$

where

$$\theta_a, \theta_b \sim \text{Galenshore}(c, d).$$

Identify a sufficient statistic.

(d) (5) Determine  $E[\theta \mid y_{1:n}]$ .

(e) (10) Show that the form of the posterior predictive density

$$p(y_{n+1} \mid y_{1:n}) = \frac{2y_{n+1}^{2a-1} \Gamma(an + a + c)}{\Gamma(a) \Gamma(an + c)} \frac{(d^2 + \sum y_i^2)^{an+c}}{(d^2 + \sum y_i^2 + y_{n+1}^2)^{(an+a+c)}}.$$