

# Homework 2

STA-360-602

Total pts: 3 (reproducibility) + 9 (Q1) + 10 (Q2) + 15 (Q3) = 37

Follow the homework instructions per the syllabus and in class instructions. Late homeworks will not be accepted.

The goal of this problem is to make sure that you have an understanding of simple Bayesian models and how to work with them in R markdown, which is the point of the lab assignment this week. Your teaching assistants will go over the first two tasks and you will finish the remainder for your homework assignment.

1. *Lab component* (3 points each) Please refer to lab 2 and complete tasks 3—5.

The goal of this problem is to see how a conjugate model relates to real data. You will get practice deriving a posterior distribution that you have not seen before, plotting densities as we did in class, and seeing a connection to real data. Finally, you will get practice thinking about when the model below might be appropriate in practice.

2. (10 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

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

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) (3) Derive (step by step) 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) (2) Why is the posterior distribution a *proper* density or probability distribution function?
- (c) (3) 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) (2) 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.

The goal of this problem is to introduce you to a new family of distributions, get more practice deriving the posterior, and work with a posterior predictive distribution on your own for the first time. This will be an intense problem, so reach out if you're having trouble!

3. (15 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) (3) Identify a class of conjugate prior densities for  $\theta$ . **Assume the prior parameters are  $c$  and  $d$ , which are fixed and known.** That is, state the prior distribution for  $\theta$ , which will have **known and fixed** parameters  $c, d$  such that the resulting posterior is conjugate. Plot a few members of this class of densities.
- (b) (3) 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) (5) Show that

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

where

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

Identify a sufficient statistic.

- (d) (1) Determine  $E[\theta|y_{1:n}]$ .
- (e) (3) Show that the form of the posterior predictive density

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