

Homework 2

STA-360-602

Due at 5:00 PM EDT on Friday, 14 January

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.

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 TA's will go over the first two tasks and help you with the three other tasks if you're having trouble or need some help.

1. *Lab component* (30 points total) Please refer to lab 2 and complete tasks 3—5.

(10) Task 3

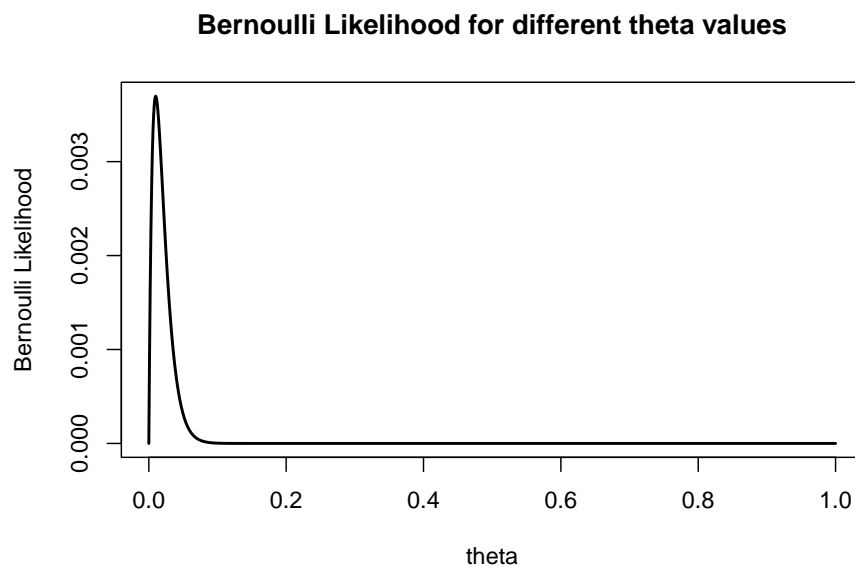
```
# set seed for reproducibility
set.seed(123)
obs.data <- rbinom(n = 100, size = 1, prob = 0.01) # observed data
```

```

### Bernoulli LH Function ###
# Input: obs.data, theta
# Output: bernoulli likelihood
Bernoulli_Likelihood <- function(obs, thetas) {
  N <- length(obs)
  x <- sum(obs)
  likelihood <- (thetas^x) * (1-thetas)^(N-x)
  return(likelihood)
}

### Plot LH for a grid of theta values ###
# Create the grid #
theta_seq <- seq(0,1,length=1000)
# Store the LH values
Likelihood_seq <- Bernoulli_Likelihood(obs.data, theta_seq)
# Create the Plot
plot(theta_seq, Likelihood_seq,
     type='l', lwd=2, # line
     main='Bernoulli Likelihood for different theta values',
     xlab='theta',
     ylab='Bernoulli Likelihood')

```



(10) Task 4

```
prior_to_posterior <- function(obs, a_prior, b_prior) {
  N <- length(obs)
  x <- sum(obs)
  a_posterior <- a_prior + x
  b_posterior <- b_prior + N - x
  return(c(a_posterior, b_posterior))
}
```

```
# non-informative prior 1, 1
prior_to_posterior(obs.data, 1, 1)
```

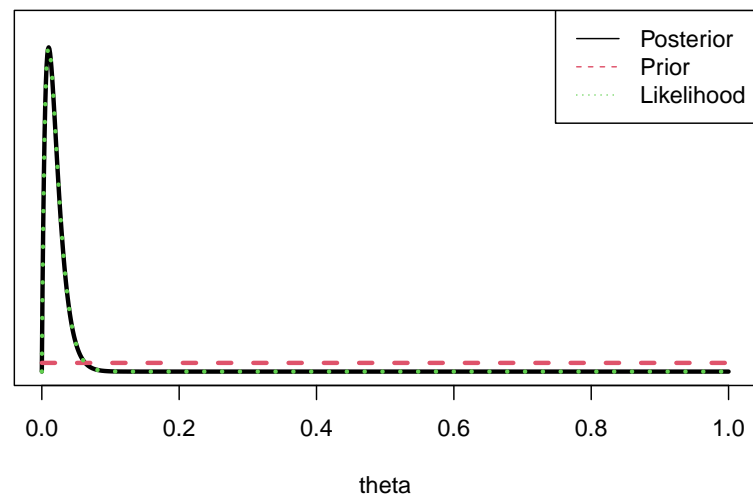
```
## [1] 2 100
```

```
# informative prior 3, 1
prior_to_posterior(obs.data, 3, 1)
```

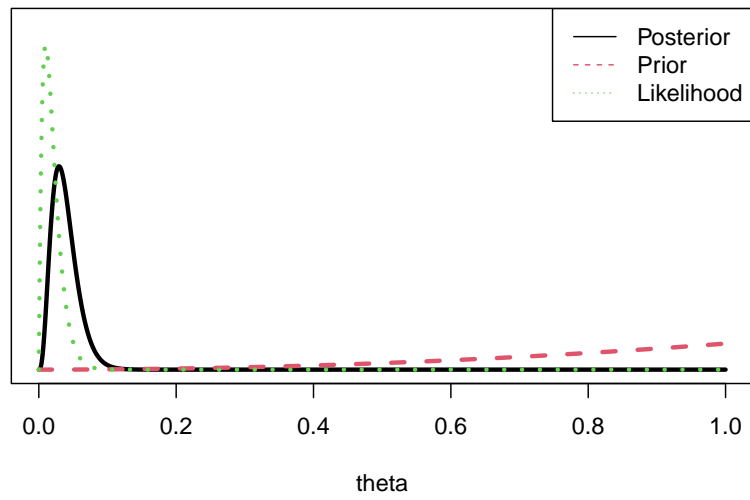
```
## [1] 4 100
```

(10) Task 5

Noninformative priors 1, 1



Informative priors 3, 1



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. (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

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

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) (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).

$$\begin{aligned} p(\theta|x_{1:n}) &= p(x_{1:n}|\theta)p(\theta) \\ &= \left(\prod_{i=1}^n \text{Exp}(x_i|\theta) \right) \text{Gamma}(\theta|a, b) \\ &= \left(\prod_{i=1}^n \theta \exp(-\theta x_i) \right) \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \mathbb{1}(x > 0) \\ &= \theta^n \exp(-\theta \sum_{i=1}^n x_i) \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \mathbb{1}(x > 0) \\ &= \frac{b^a}{\Gamma(a)} \theta^{a+n-1} \exp(-\theta(b + \sum_{i=1}^n x_i)) \mathbb{1}(x > 0) \\ &\propto \theta^{a+n-1} \exp(-\theta(b + \sum_{i=1}^n x_i)) \mathbb{1}(x > 0) \\ &\propto \text{Gamma}(a + n, b + \sum_{i=1}^n x_i) \end{aligned}$$

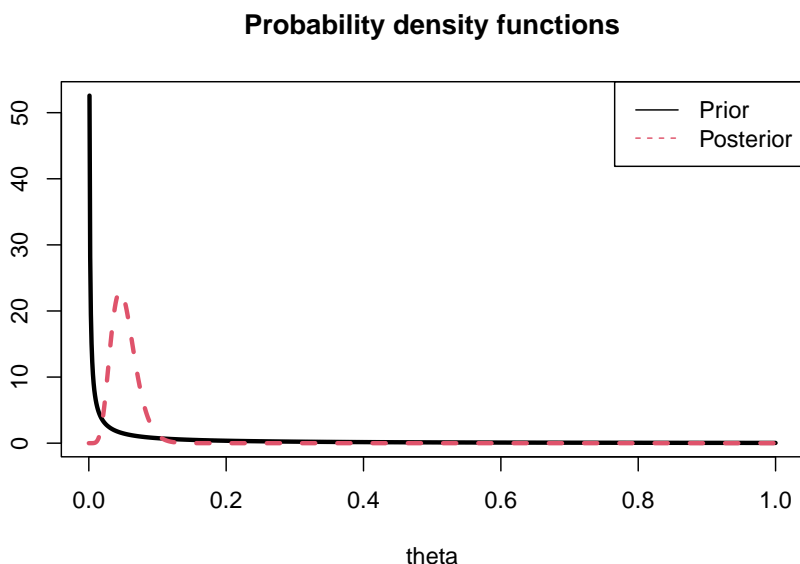
(b) (5) Why is the posterior distribution a *proper* density or probability distribution function?

- From the Gamma prior distribution, we should have parameters $a, b > 0$.
- From the Exponential distribution, we have $n > 0, \sum_{i=1}^n x_i > 0$.
- Hence, we have $a + n > 0, b + \sum_{i=1}^n x_i > 0$; both parameters satisfy the requirement of the Gamma function, so the posterior distribution is *proper*.

(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.

- An Exponential model is reasonable when modeling time because the distribution is over $(0, +\infty)$; moreover, the Exponential distribution is

memoryless, i.e., $P(X > t + s | X > t) = P(X > s)$, making it suitable for modelling the time between next randomly occurring events such as neural spikes when the time that has elapsed is irrelevant to when the next event will occur.

- An Exponential model is NOT reasonable when modeling the fluctuation in the stock market (rises and falls), because the model can never achieve negative values.
- Source used: “STA 360: Reference Sheet for Distributions”

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. (40 points total) *Priors, Posteriors, Predictive Distributions (Hoff, 3.9)*

An unknown quantity $Y|\theta$ has a Galenshore(a, θ) 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 θ is unknown and a random variable. For this density,

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

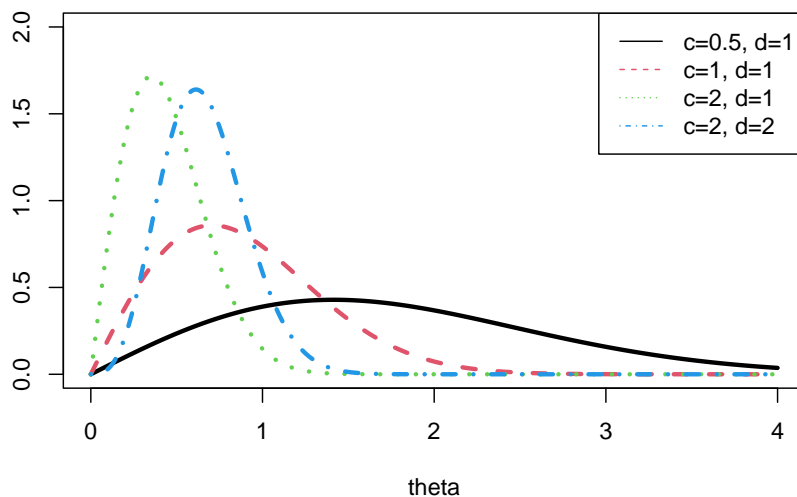
and

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

- (a) (10) Identify a class of conjugate prior densities for θ . Assume the prior parameters are c and d . That is, state the distribution that θ should have with parameters c, d such that the resulting posterior is conjugate. Plot a few members of this class of densities.

- $\theta \sim \text{Galenshore}(c, d)$

Galenshore density functions



- (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.

$$\begin{aligned}
p(\theta|y_{1:n}) &= p(y_{1:n}|\theta)p(\theta) \\
&= \left(\prod_{i=1}^n \text{Galenshore}(y_i|a, \theta) \right) \text{Galenshore}(\theta|c, d) \\
&= \left(\prod_{i=1}^n \frac{2}{\Gamma(a)} \theta^{2a} y_i^{2a-1} e^{-\theta^2 y_i^2} \right) \frac{2}{\Gamma(c)} d^{2c} \theta^{2c-1} e^{-d^2 \theta^2} \\
&= \left(\prod_{i=1}^n y_i^{2a-1} \right) \frac{2^n}{(\Gamma(a))^n} \theta^{2an} e^{-\theta^2 \sum y_i^2} \frac{2}{\Gamma(c)} d^{2c} \theta^{2c-1} e^{-d^2 \theta^2} \\
&\propto \theta^{2an} e^{-\theta^2 \sum y_i^2} \theta^{2c-1} e^{-d^2 \theta^2} \\
&\propto \theta^{2(c+an)-1} e^{-\theta^2 (d^2 + \sum y_i^2)}
\end{aligned}$$

Therefore, $\theta|y_{1:n} \sim \text{Galenshore}(c + an, \sqrt{d^2 + \sum y_i^2})$.

- Given that $a, c, d, n, y > 0$, this posterior distribution is proper because $c + an > 0$ and $\sqrt{d^2 + \sum y_i^2} > 0$.

- (c) (10) 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.

$$\begin{aligned}
\frac{p(\theta_a|y_{1:n})}{p(\theta_b|y_{1:n})} &= \frac{p(y_{1:n}|\theta_a)p(\theta_a)}{p(y_{1:n}|\theta_b)p(\theta_b)} \\
&= \frac{\left(\prod y_i^{2a-1} \right) \frac{2^n}{(\Gamma(a))^n} \theta_a^{2an} e^{-\theta_a^2 \sum y_i^2} \frac{2}{\Gamma(c)} d^{2c} \theta_a^{2c-1} e^{-d^2 \theta_a^2}}{\left(\prod y_i^{2a-1} \right) \frac{2^n}{(\Gamma(a))^n} \theta_b^{2an} e^{-\theta_b^2 \sum y_i^2} \frac{2}{\Gamma(c)} d^{2c} \theta_b^{2c-1} e^{-d^2 \theta_b^2}} \\
&= \frac{\theta_a^{2an} e^{-\theta_a^2 \sum y_i^2} \theta_a^{2c-1} e^{-d^2 \theta_a^2}}{\theta_b^{2an} e^{-\theta_b^2 \sum y_i^2} \theta_b^{2c-1} e^{-d^2 \theta_b^2}} \\
&= \frac{\theta_a^{2(an+c)-1} e^{-\theta_a^2 (d^2 + \sum y_i^2)}}{\theta_b^{2(an+c)-1} e^{-\theta_b^2 (d^2 + \sum y_i^2)}} \\
&= \left(\frac{\theta_a}{\theta_b} \right)^{2(an+c)-1} e^{(\theta_b^2 - \theta_a^2)(d^2 + \sum y_i^2)}
\end{aligned}$$

Therefore, $\sum y_i^2$ is sufficient.

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

- It is given that $E[Y] = \frac{\Gamma(a+1/2)}{\theta\Gamma(a)}$ for $Y \sim \text{Galenshore}(a, \theta)$.
- Since $p(\theta|y_{1:n}) \sim \text{Galenshore}(c + an, \sqrt{d^2 + \sum y_i^2})$,

$$E[\theta|y_{1:n}] = \frac{\Gamma(c + an + 1/2)}{\sqrt{d^2 + \sum y_i^2} \Gamma(c + an)}.$$

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

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

$$\begin{aligned} p(y_{n+1}|y_{1:n}) &= \int p(y_{n+1}|\theta)p(\theta|y_{1:n})d\theta \\ &= \int \frac{2}{\Gamma(a)}\theta^{2a}y_{n+1}^{2a-1}e^{-\theta^2 y_{n+1}^2} \frac{2}{\Gamma(c + an)}(d^2 + \sum y_i^2)^{c+an}\theta^{2(c+an)-1}e^{-(d^2 + \sum y_i^2)\theta^2}d\theta \\ &= \frac{4y_{n+1}^{2a-1}(d^2 + \sum y_i^2)^{c+an}}{\Gamma(a)\Gamma(c + an)} \int \theta^{2(an+a+c)-1}e^{-(d^2 + \sum y_i^2 + y_{n+1}^2)\theta^2}d\theta \\ &= \frac{4y_{n+1}^{2a-1}(d^2 + \sum y_i^2)^{c+an}}{\Gamma(a)\Gamma(c + an)} \frac{\Gamma(an + a + c)}{2(d^2 + \sum y_i^2 + y_{n+1}^2)^{(an+a+c)}} \\ &\quad \cdot \int \frac{2}{\Gamma(an + a + c)}(d^2 + \sum y_i^2 + y_{n+1}^2)^{(an+a+c)}\theta^{2(an+a+c)-1}e^{-(d^2 + \sum y_i^2 + y_{n+1}^2)\theta^2}d\theta \\ &= \frac{2y_{n+1}^{2a-1}\Gamma(an + a + c)}{\Gamma(a)\Gamma(an + c)} \frac{(d^2 + \sum y_i^2)^{an+c}}{(d^2 + \sum y_i^2 + y_{n+1}^2)^{(an+a+c)}} \\ &\quad \cdot \int p(\theta|an + a + c, \sqrt{d^2 + \sum y_i^2 + y_{n+1}^2})d\theta \\ &\quad \text{where } \theta \sim \text{Galenshore}(an + a + c, \sqrt{d^2 + \sum y_i^2 + y_{n+1}^2}) \\ &= \frac{2y_{n+1}^{2a-1}\Gamma(an + a + c)}{\Gamma(a)\Gamma(an + c)} \frac{(d^2 + \sum y_i^2)^{an+c}}{(d^2 + \sum y_i^2 + y_{n+1}^2)^{(an+a+c)}}. \end{aligned}$$