

Homework 2

Shannon Murphy

2023-01-20

Lab Problems

Task 3

Write a function that takes as its inputs that data you simulated (or any data of the same type) and a sequence of θ values of length 1000 and produces Likelihood values based on the Binomial Likelihood. Plot your sequence and its corresponding Likelihood function.

The likelihood function is given below. Since this is a probability and is only valid over the interval from $[0, 1]$ we generate a sequence over that interval of length 1000.

You have a rough sketch of what you should do for this part of the assignment. Try this out in lab on your own.

```
#simulate data
#set a seed
set.seed(2023)

#create data
simulated_data <- rbinom(n = 100, size = 1, prob = 0.01)

#inspect the simulated data
table(simulated_data)
```

```
## simulated_data
##  0  1
## 97  3
```

```
head(simulated_data)
```

```
## [1] 0 0 0 0 0 0
```

```

### Bernoulli LH Function ###
# Input: obs_data, theta
# Output: bernoulli Likelihood

myBernLH <- function(obs_data, theta) {

  N <- length(obs_data)
  x <- sum(obs_data)
  LH <- (theta^x)*((1-theta)^(N-x))
  return(LH)

}

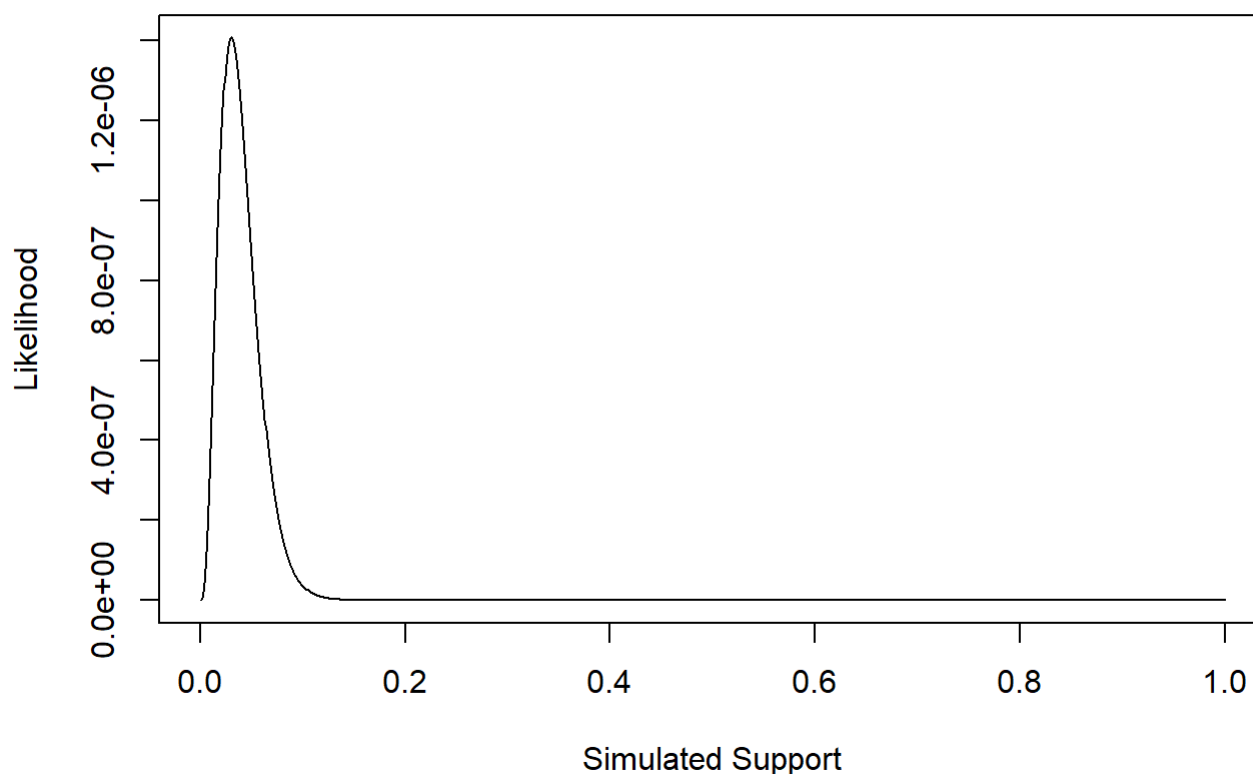
### Plot LH for a grid of theta values ###
# Create the grid #
theta_sim <- seq(from = 0, to = 1, length.out = 1000)

# Store the LH values
LH_values <- myBernLH(simulated_data, theta_sim)

# Create the Plot
plot(theta_sim, LH_values, type = 'l', main = 'Likelihood Profile', xlab = 'Simulated Support',
      ylab = 'Likelihood')

```

Likelihood Profile



Task 4

Write a function that takes as its inputs prior parameters a and b for the Beta-Bernoulli model and the observed data, and produces the posterior parameters you need for the model. Generate and print the posterior parameters for a non-informative prior such as $(a,b) = (1,1)$ and for an informative case $(a,b) = (3,1)$.

$$X_1, X_2, \dots, X_n \sim \text{Bernoulli}(\theta)$$

$$\text{likelihood} = p(x_1, x_2, \dots, x_n | \theta) = p(x_{1:n} | \theta) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$$

Assume $\theta \sim \text{Beta}(a, b)$ where $a, b > 0$,

$$p(\theta) = \text{Beta}(\theta | a, b) = \frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1}$$

$$p(\theta) \propto \theta^{a-1} (1 - \theta)^{b-1}$$

Therefore, $p(\theta | x_{1:n}) \propto p(x_{1:n} | \theta) * p(\theta)$

$$p(\theta | x_{1:n}) \propto [\theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}] (\theta^{a-1} (1 - \theta)^{b-1})$$

$$p(\theta | x_{1:n}) \propto \theta^{\sum x_i + a - 1} (1 - \theta)^{n + b - \sum x_i - 1}$$

This is of the form $\text{Beta}(\sum x_i + a, n - \sum x_i + b)$

For the simulation, the input will be the data, `simulated_data`, from the Bernoulli distribution, and the parameters (a,b) of the prior.

```
#create function to return the parameters of the posterior beta distribution.
beta_bernoulli <- function(a,b,data) {
  sumofx <- sum(data)
  sample_size <- length(data)

  a0 <- a + sumofx
  b0 <- sample_size + b - sumofx

  return(paste('a0 is', a0, ' b0 is', b0))
}
```

```
#run function with uninformative prior
uninformative_ex <- beta_bernoulli(1,1,simulated_data)
uninformative_ex
```

```
## [1] "a0 is 4  b0 is 98"
```

```
#run function with informative prior
informative_ex <- beta_bernoulli(3,1,simulated_data)
informative_ex
```

```
## [1] "a0 is 6  b0 is 98"
```

Task 5

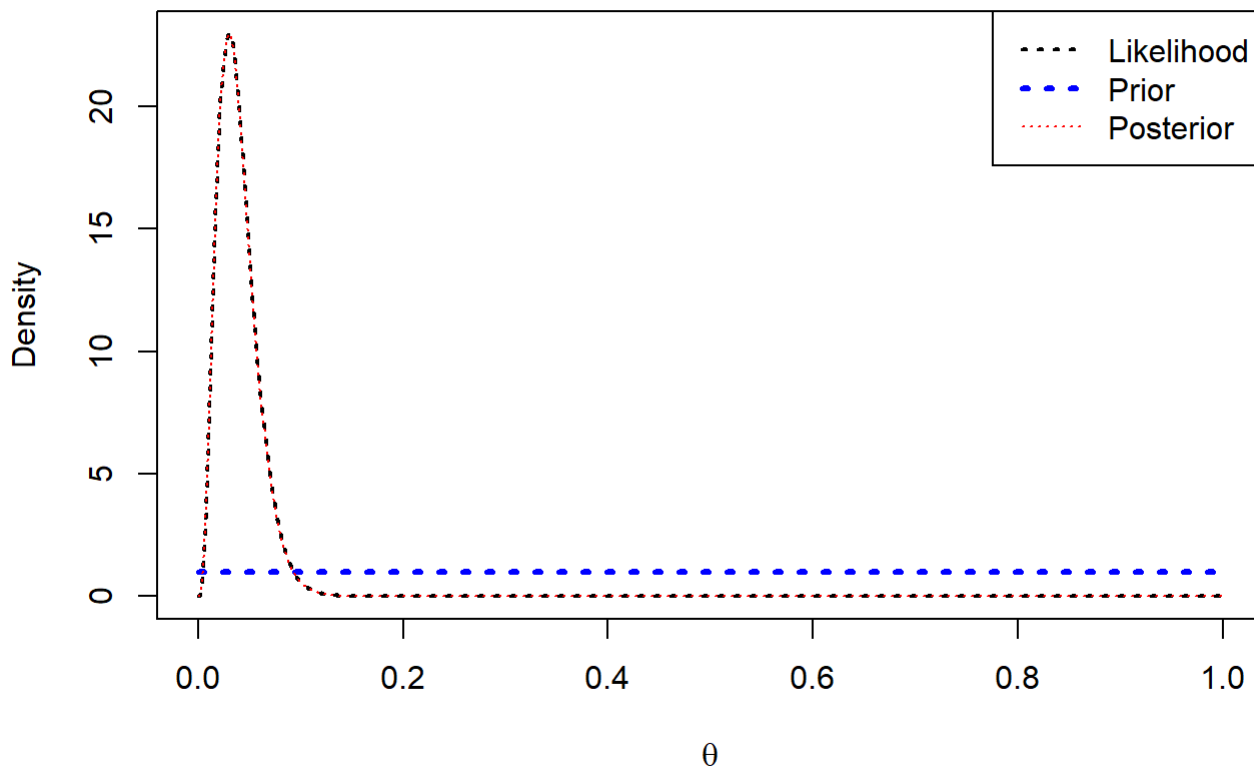
Create two plots, one for the informative and one for the non-informative case to show the posterior distribution and superimpose the prior distributions on each along with the likelihood. What do you see? Remember to turn the y-axis ticks off since superimposing may make the scale non-sense.

```
sampsize <- length(simulated_data)
sumofx <- sum(simulated_data)
a <- 1
b <- 1

likelihood <- dbeta(theta_sim, sumofx + 1, sampsize - sumofx + 1)
prior <- dbeta(theta_sim, a, b)
posterior <- dbeta(theta_sim, sumofx + a, sampsize - sumofx + b)

plot(theta_sim, likelihood, type = 'l', ylab = 'Density', lty = 3, lwd = 2, xlab = expression(theta), col = 'black', main = 'Density Plots with Uninformative Prior')
lines(theta_sim, prior, type = 'l', ylab = 'Density', lty = 3, lwd = 3, col = 'blue')
lines(theta_sim, posterior, type = 'l', ylab = 'Density', lty = 3, lwd = 1, col = 'red')
legend('topright', legend = c('Likelihood','Prior','Posterior'), lwd = c(2,3,1), lty = 3, col = c('black','blue','red'))
```

Density Plots with Uninformative Prior

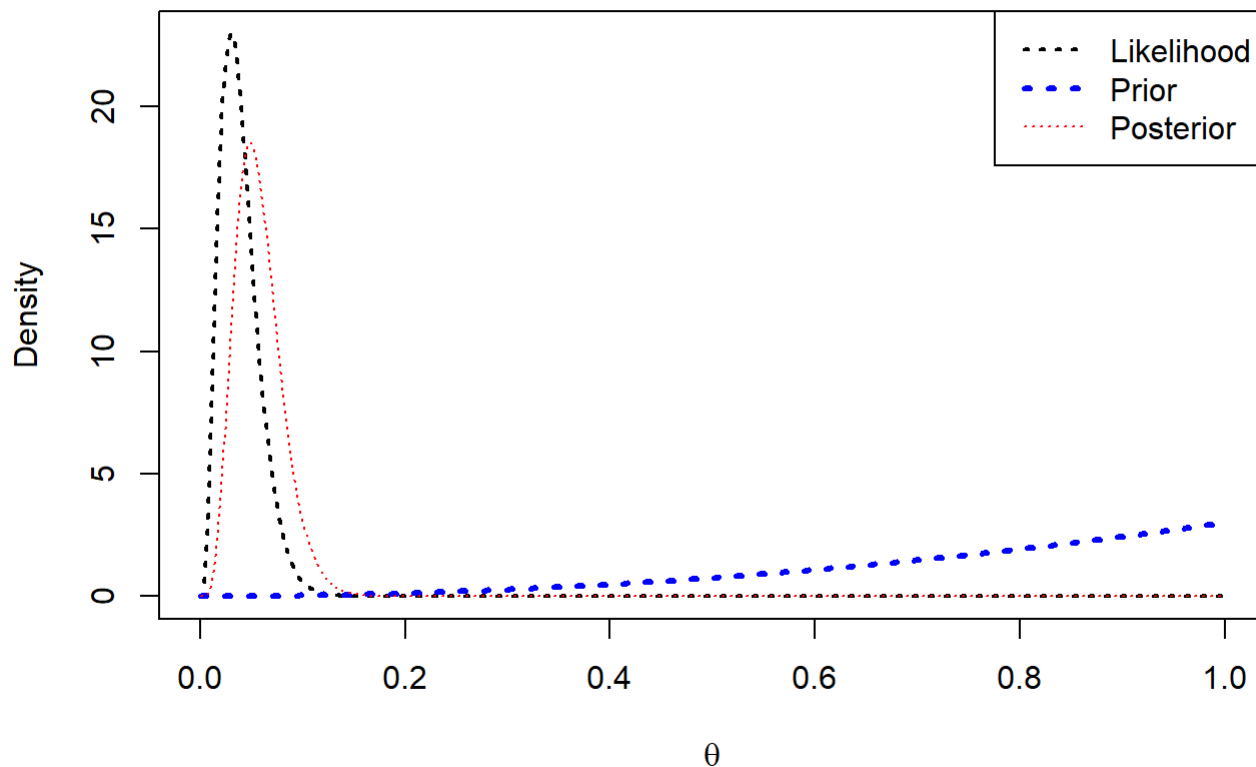


```
sampsize <- length(simulated_data)
sumofx <- sum(simulated_data)
a <- 3
b <- 1

likelihood <- dbeta(theta_sim, sumofx + 1, sampsize - sumofx + 1)
prior <- dbeta(theta_sim, a, b)
posterior <- dbeta(theta_sim, sumofx + a, sampsize - sumofx + b)

plot(theta_sim, likelihood, type = 'l', ylab = 'Density', lty = 3, lwd = 2, xlab = expression(theta), col = 'black', main = 'Density Plots with Informative Prior')
lines(theta_sim, prior, type = 'l', ylab = 'Density', lty = 3, lwd = 3, col = 'blue')
lines(theta_sim, posterior, type = 'l', ylab = 'Density', lty = 3, lwd = 1, col = 'red')
legend('topright', legend = c('Likelihood', 'Prior', 'Posterior'), lwd = c(2, 3, 1), lty = 3, col = c('black', 'blue', 'red'))
```

Density Plots with Informative Prior



Homework Problems

Question 2: The Exponential-Gamma Model

We write $X \sim \text{Exp}(\theta)$ to indicate that X has the Exponential distribution, that is, its pdf is

$$p(x|\theta) = \text{Exp}(x|\theta) = \theta \exp(-\theta x)$$

for $x > 0$. 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, extreme values such as maximum daily rainfall over a period of one year, or the amount of time until a product fails.

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

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

for $\theta > 0$.

Part a) 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.

Step 1: Define the likelihood

$$p(x_1, \dots, x_n|\theta) = \prod_{i=1}^n \theta \exp(-\theta x_i)$$

$$p(x_1, \dots, x_n|\theta) = \theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right)$$

Step 2: Find posterior

$$p(\theta|x_{1:n}) \propto \left[\theta^n \exp\left(-\theta \sum_{i=1}^n x_i\right) \right] \left(\frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta) \right)$$

$$p(\theta|x_{1:n}) \propto \theta^{n+a-1} \exp\left(-\left(b + \sum_{i=1}^n x_i\right) \theta\right)$$

This is in the form of a Gamma distribution, so $p(\theta|x_{1:n}) \sim \text{Gamma}(n + a, b + \sum x_i)$. Therefore, we can assume that

$$p(\theta|x_{1:n}) = \frac{(b + \sum x_i)^{n+a}}{\Gamma(n + a)} \theta^{n+a-1} \exp\left(-\left(b + \sum_{i=1}^n x_i\right) \theta\right)$$

Part b) Why is the posterior distribution a proper density or probability distribution function?

The posterior distribution is a proper density because it is derived by dividing a joint distribution by a marginal likelihood of the data. In this specific case, we were dealing with a conjugate prior, meaning the posterior is of the same distributional family as the prior.

Proper densities integrate to 1 over their entire support.

Part c) 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 listed below. Plot the prior and posterior pdfs (be sure to make sure your plots on a scale that allows you to clearly see the important features).

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

```

q2_data <- c(20.9,69.7,3.6,21.8,21.4,0.4,6.7,10.0)

q2_sum <- sum(q2_data)
q2_n <- length(q2_data)
q2_a <- 0.1
q2_b <- 1.0

q2_a0 <- q2_n + q2_a
q2_b0 <- q2_b + q2_sum

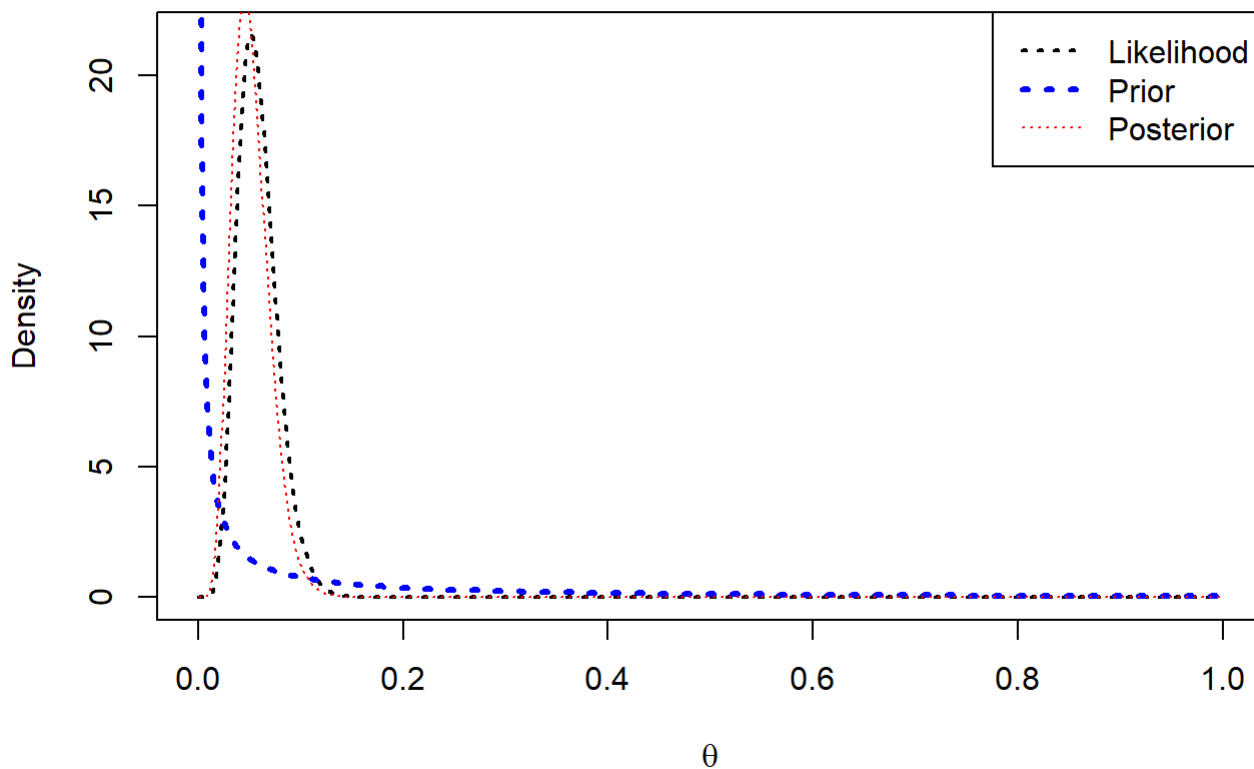
q2_theta <- seq(from = 0, to = 1, length.out = 1000)

likelihood <- dgamma(q2_theta, q2_n+1, q2_sum)
prior <- dgamma(q2_theta, q2_a, q2_b)
posterior <- dgamma(q2_theta, q2_a0, q2_b0)

plot(q2_theta, likelihood, type = 'l', ylab = 'Density', lty = 3, lwd = 2, xlab = expression(theta), col = 'black', main = 'Density Plots')
lines(q2_theta, prior, type = 'l', ylab = 'Density', lty = 3, lwd = 3, col = 'blue')
lines(q2_theta, posterior, type = 'l', ylab = 'Density', lty = 3, lwd = 1, col = 'red')
legend('topright', legend = c('Likelihood','Prior','Posterior'), lwd = c(2,3,1), lty = 3, col = c('black','blue','red'))

```

Density Plots



Part d) 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 exponential distribution assumes a constant rate of events. For example, an exponential model could be used to estimate time until radioactive particle decay. It would not be appropriate to use an exponential model for time between snowfalls over the course of a full year because there will be a different rate in the winter than the summer.

Question 3: Priors, Posteriors, Predictive Distributions

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

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

Part a) 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.

The Galenshore family is a conjugate prior.

$$p(\theta) = \frac{2}{\Gamma(c)} d^{2c} \theta^{2c-1} e^{-d^2 \theta^2}$$

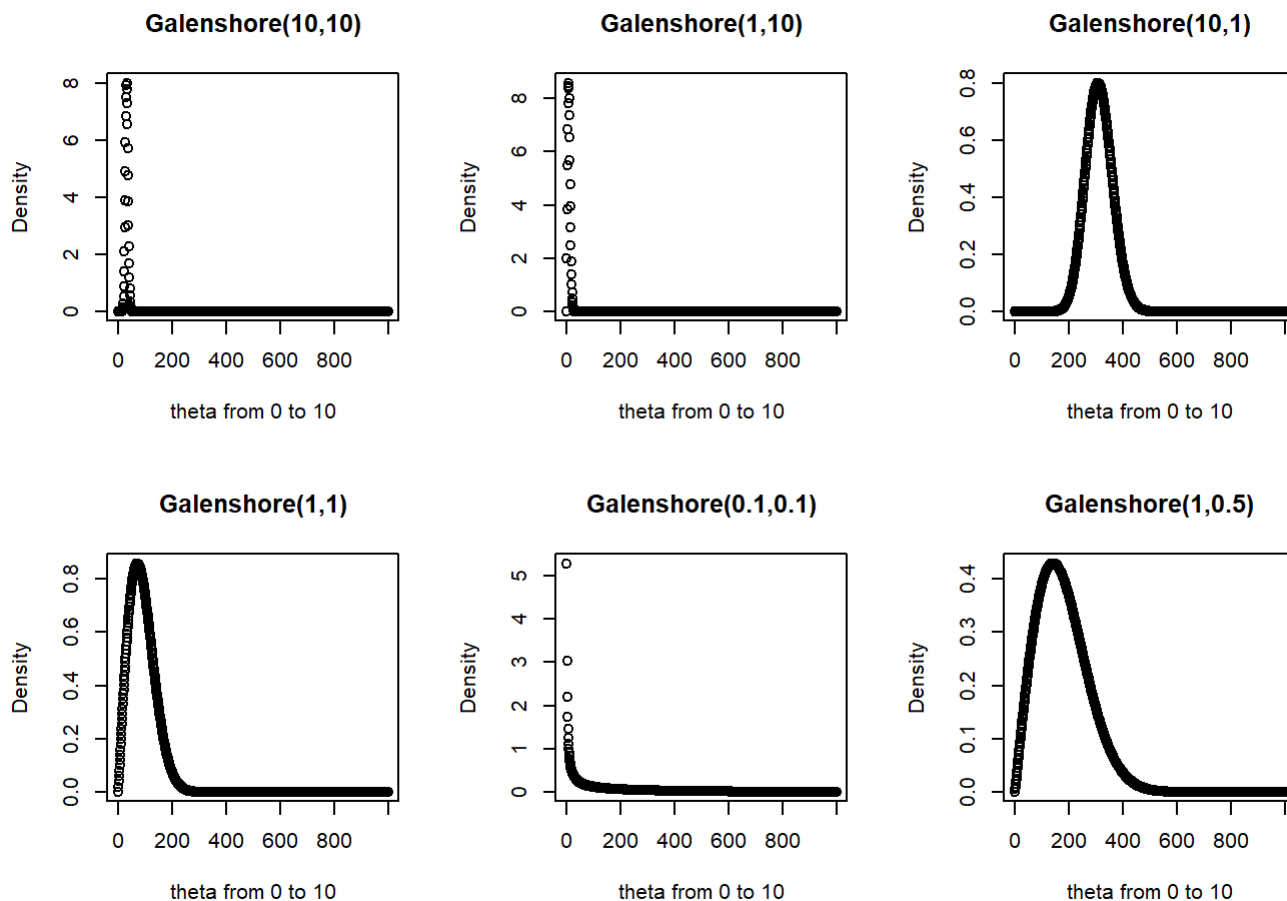
```

#Plot a few members of the __ class of densities
theta_3 <- seq(from = 0, to = 10, length.out = 1000)

galenshore <- function(c,d,theta) {
  prob <- (2/factorial(c-1))*(d^(2*c))*(theta^(2*c-1))*exp(-(d^2)*(theta^2))
  return(prob)
}

par(mfrow = c(2,3))
plot(galenshore(10,10,theta_3), main = 'Galenshore(10,10)', xlab = 'theta from 0 to 10', ylab = 'Density')
plot(galenshore(1,10,theta_3), main = 'Galenshore(1,10)', xlab = 'theta from 0 to 10', ylab = 'Density')
plot(galenshore(10,1,theta_3), main = 'Galenshore(10,1)', xlab = 'theta from 0 to 10', ylab = 'Density')
plot(galenshore(1,1,theta_3), main = 'Galenshore(1,1)', xlab = 'theta from 0 to 10', ylab = 'Density')
plot(galenshore(0.1,0.1, theta_3), main = 'Galenshore(0.1,0.1)', xlab = 'theta from 0 to 10', ylab = 'Density')
plot(galenshore(1,0.5, theta_3), main = 'Galenshore(1,0.5)', xlab = 'theta from 0 to 10', ylab = 'Density')

```



Part b)

Let $Y_1, \dots, Y_n \sim \text{Galenshore}(a, \theta)$ iid. Find the posterior distribution of $\theta|y_{1:n}$ using a prior from your conjugate class.

Step 1: Find the likelihood

$$p(y_1, y_2, \dots, y_n | \theta) = \prod_{i=1}^n \frac{2}{\Gamma(a)} \theta^{2a} y_i^{2a-1} \exp(-\theta^2 y_i^2)$$

$$p(y_1, y_2, \dots, y_n | \theta) \propto \theta^{2an} \exp(-\theta^2 \sum y_i^2)$$

Step 2: Find the prior

$$p(y_{1:n} | \theta) p(\theta) \propto \left(\theta^{2an} e^{-\theta^2 \sum y_i^2} \right) [\theta^{2c-1} e^{-d^2 \theta^2}]$$

$$p(\theta | y_{1:n}) \propto \theta^{2(na+c)-1} e^{-\theta^2 (d^2 + \sum y_i^2)}$$

This is of the form $\text{Galenshore}(na + c, (d^2 + \sum_{i=1}^n y_i^2)^{1/2})$

Part c) Show the following, and identify a sufficient statistic.

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

where

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

Part 1)

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

where C is the constant on the posterior. It then follows

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

Part 2) Find a sufficient statistic

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

Consider the definition of regular exponential class being anything of the form $p(y | \phi) = h(y) c(\phi) e^{q(\phi) t(y)}$. It then follows that for a random sample Y_1, Y_2, \dots, Y_n of REC(q), the statistic $S = \sum_{i=1}^n t(y_i)$ is a complete sufficient statistic for ϕ .

Define $h(y) = y^{2a-1}$, $c(\theta) = \frac{2}{\Gamma(a)}\theta^{2a}$, $q(\theta) = -\theta^2$, and $t(y) = y^2$. This shows $p(y|\theta)$ is a member of the regular exponential class, and thus,

$$S = \sum_{i=1}^n y_i^2$$

is a complete and sufficient statistic for θ .

Part d) Determine the expected value

$$E[\theta|y_{1:n}]$$

It was given that $E[Y] = \frac{\Gamma(a+1/2)}{\theta\Gamma(a)}$ for $Galenshore(a, \theta)$. It then follows that for $Galenshore(na + c, (d^2 + \sum_{i=1}^n y_i^2)^{1/2})$

$$E[\theta|y_{1:n}] = \frac{\Gamma(na + c + 1/2)}{(d^2 + \sum_{i=1}^n y_i^2)^{1/2}\Gamma(na + c)}$$

Part e) Show that the form of the posterior predictive density is

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

By definition of posterior predictive density,

$$p(y_{n+1}|y_{1:n}) = \int p(y_{n+1}|\theta)p(\theta|y_{1:n})d\theta$$

Using the likelihood and posterior found above,

$$p(y_{n+1}|\theta) = \frac{2}{\Gamma(a)}\theta^{2a}y_{n+1}^{2a-1}\exp(-\theta^2y_{n+1}^2)$$

$$p(\theta|y_{1:n}) = \frac{2}{\Gamma(na + c)}(d^2 + \sum y_i^2)^{na+c}(\theta^{2(na+c)-1})e^{-\theta^2(d^2 + \sum y_i^2)}$$

Find the product and simplify:

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

Define $A = y_{n+1}^{2a-1}(d^2 + \sum y_i^2)^{na+c}$. Then using the posterior predictive density equation,

$$p(y_{n+1}|y_{1:n}) = A \int \theta^{2(an+a+c)-1}e^{-(d^2 + y_{n+1}^2 + \sum y_i^2)\theta^2} d\theta$$

Define $B = an + a + c$ and $C = d^2 + y_{n+1}^2 + \sum y_i^2$, then

$$p(y_{n+1}|y_{1:n}) = A \int \theta^{2B-1} e^{-C\theta^2} d\theta$$

The integrand, $\theta^{2B-1} e^{-C\theta^2} d\theta$, looks suggests $\theta \sim \text{Galenshore}(B, \sqrt{C})$. The true pdf would be $p(\theta) = \frac{2}{\Gamma(B)} C^B \theta^{2B-1} e^{-C\theta^2}$. Therefore, we will multiply $p(y_{n+1}|y_{1:n})$ by

$$1 = \frac{2C^B/\Gamma(B)}{2C^B/\Gamma(B)}$$

$$p(y_{n+1}|y_{1:n}) = A \frac{\Gamma(B)}{2C^B} \int \frac{2}{\Gamma(B)} C^B \theta^{2B-1} e^{-C\theta^2} d\theta$$

Because Galenshore is a proper distribution, the integral = 1 and only the constant in front is leftover.

$$p(y_{n+1}|y_{1:n}) = A \frac{\Gamma(B)}{2C^B}$$

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

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