

Homework 2

STA 360, Fall 2020

Due Friday August 28, 5 PM EDT

Lab component

Task 1

Let's start by quickly deriving the Beta-Binomial distribution.

We assume that

$$X \mid \theta \sim \text{Binomial}(\theta)$$

,

$$\theta \sim \text{Beta}(a, b),$$

where a, b are assumed to be known parameters. What is the posterior distribution of $\theta \mid X$?

$$p(\theta \mid X) \propto p(X \mid \theta)p(\theta) \tag{1}$$

$$\propto \theta^x (1 - \theta)^{(n-x)} \times \theta^{(a-1)} (1 - \theta)^{(b-1)} \tag{2}$$

$$\propto \theta^{x+a-1} (1 - \theta)^{(n-x+b-1)}. \tag{3}$$

This implies that

$$\theta \mid X \sim \text{Beta}(x + a, n - x + b).$$

Task 2

Simulate some data using the `rbinom` function of size $n = 100$ and probability equal to 1%. Remember to `set.seed(123)` so that you can replicate your results.

The data can be simulated as follows:

```
# set a seed
set.seed(123)
# create the observed data
obs.data <- rbinom(n = 100, size = 1, prob = 0.01)
# inspect the observed data
head(obs.data)
```

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

```
tail(obs.data)
```

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

```
length(obs.data)
```

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

Task 3

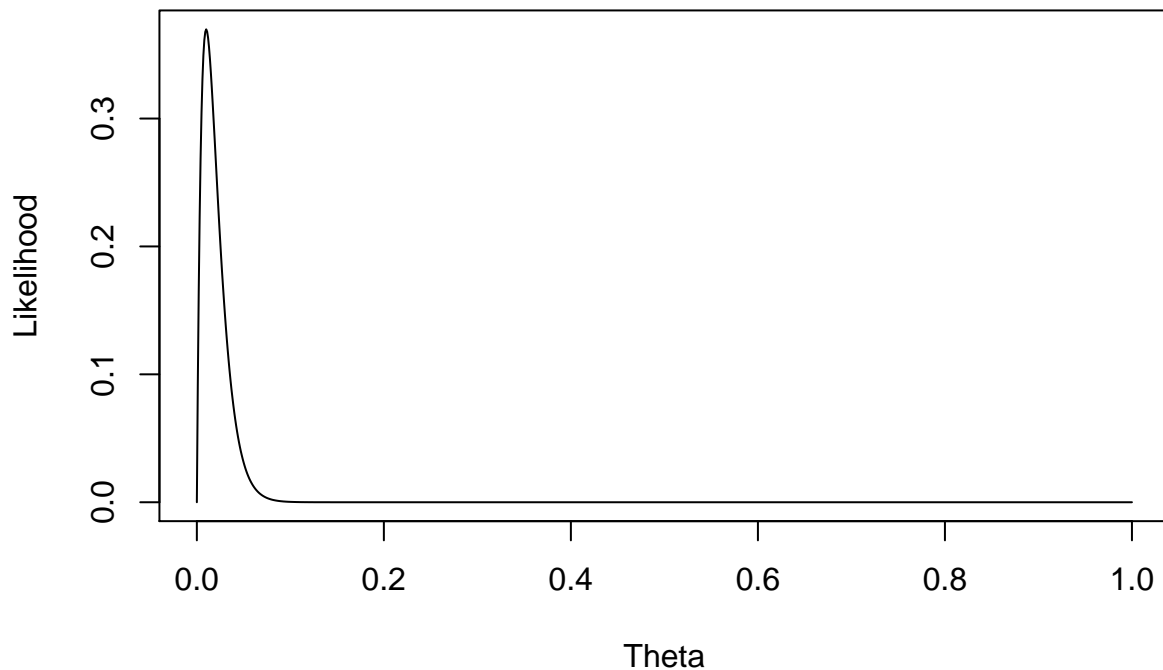
Write a function that takes as its inputs that data you simulated (or any data of the same type) and a sequence of θ values of length 1000 and produces Likelihood values based on the Binomial Likelihood. Plot your sequence and its corresponding Likelihood function.

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.

```
### Code taken from lab session
### Bernoulli LH Function ###
# Input: obs.data, theta
# Output: bernoulli likelihood
bernoulli_likelihood <- function (obs.data, theta) {
  n = length(obs.data)
  x = sum(obs.data)
  return (choose(n, x) * theta ^ (x) * (1-theta) ^ (n-x))
}

### Plot LH for a grid of theta values ###
# Create the grid #
# Store the LH values
# Create the Plot
theta = seq(0, 1, length.out = 1000)
plot(theta, bernoulli_likelihood(obs.data, theta),
      type = "l", ylab = "Likelihood", xlab = "Theta")
```



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 i.e. $(a,b) = (1,1)$ and for an informative case $(a,b) = (3,1)$.

```
### Based on code provided in lab session
### Function for generating posterior parameters
posterior_params <- function(a, b, obs.data) {
  n = length(obs.data)
  x = sum(obs.data)
  new_a = x+a
  new_b = n-x+b
  return(c(new_a, new_b))
}
### Uninformative case
posterior_params(1, 1, obs.data)
```

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

```
### Informative case
posterior_params(3, 1, obs.data)
```

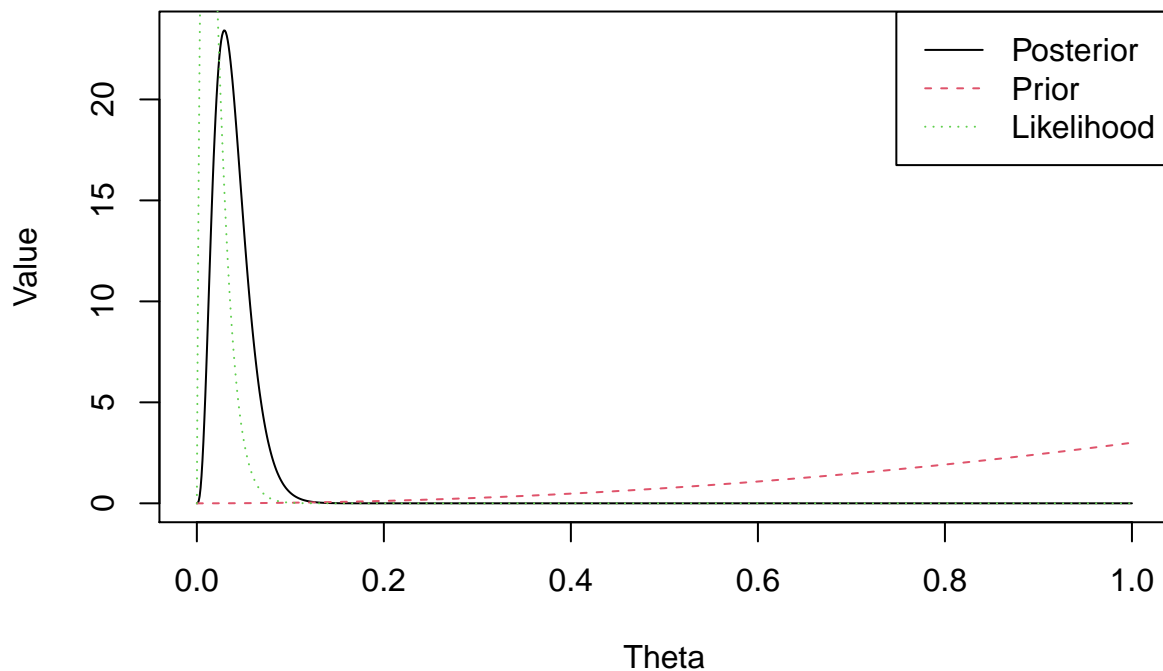
```
## [1] 4 100
```

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.

```
### Based on code provided in lab session
### Generate theta
theta = seq(0, 1, length.out = 1000)
### Find parameters of posterior distributions
informative_params = posterior_params(3, 1, obs.data)
uninformative_params = posterior_params(1, 1, obs.data)
### Plot the posterior (informative case)
plot(theta, dbeta(theta, shape1 = informative_params[1], shape2 = informative_params[2]),
     type = "l", xlab = "Theta", ylab = "Value", main = "Posterior Generated Using
     Informative Prior")
### Plot the prior (informative case)
lines(theta, dbeta(theta, shape1 = 3, shape2 = 1), col = 2, lty = 2)
### Plot the likelihood
normalized_likelihood = 1000*bernoulli_likelihood(obs.data, theta)/
    sum(bernoulli_likelihood(obs.data, theta))
lines(theta, normalized_likelihood, col = 3, lty = 3)
legend("topright", legend = c("Posterior", "Prior", "Likelihood"),
     lty = c(1,2,3), col = c(1,2,3))
```

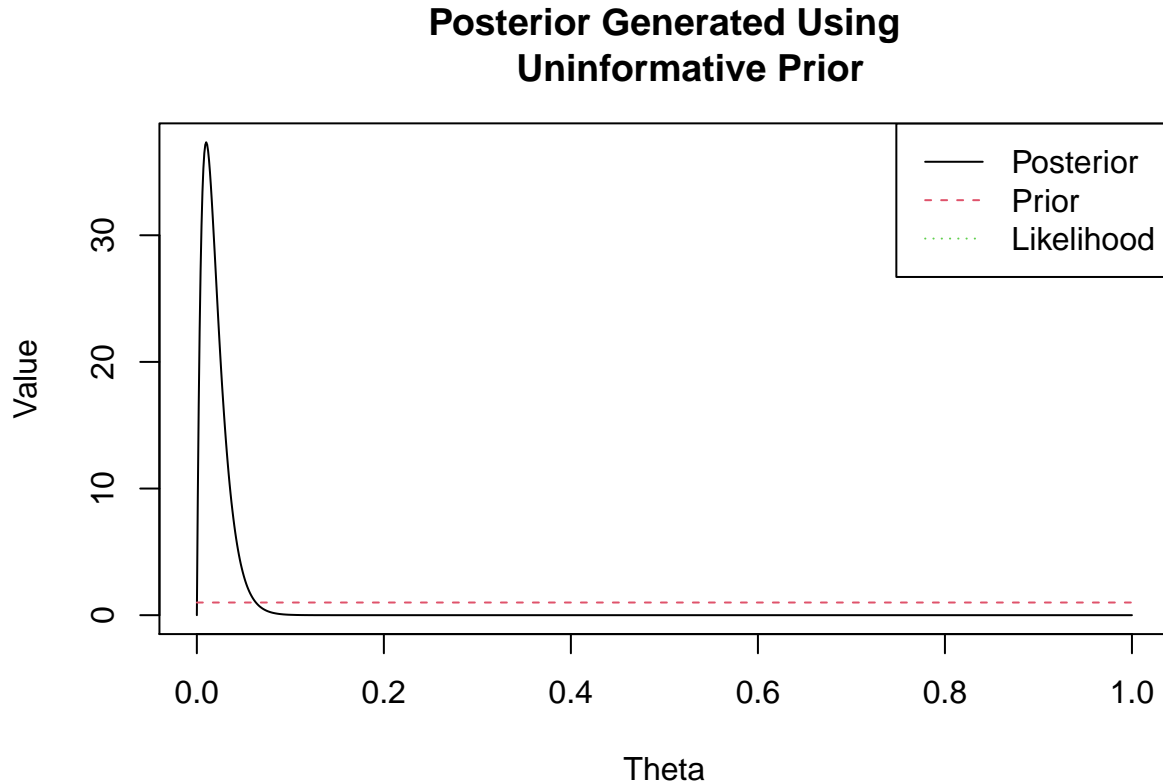
Posterior Generated Using Informative Prior



```

### Plot the posterior (uninformative case)
plot(theta, dbeta(theta, shape1 = uninformative_params[1], shape2 = uninformative_params[2]),
     type = "l", xlab = "Theta", ylab = "Value", main = "Posterior Generated Using
     Uninformative Prior")
### Plot the prior (uninformative case)
lines(theta, dbeta(theta, shape1 = 1, shape2 = 1), col = 2, lty = 2)
legend("topright", legend = c("Posterior", "Prior", "Likelihood"),
     lty = c(1,2,3), col = c(1,2,3))

```



Exponential Gamma Model

a.

$$\begin{aligned}
 p(x_{1:n}|\theta) &= \theta^n e^{-\theta \sum_{i=1}^n x_i} I(x > 0) \\
 p(\theta|x_{1:n}) &= \frac{p(x_{1:n}|\theta)p(\theta)}{p(x)} \\
 p(\theta|x_{1:n}) &\propto \theta^n e^{(-\theta \sum_{i=1}^n x_i)} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} I(x > 0) \\
 p(\theta|x_{1:n}) &\propto \theta^{n+a-1} e^{(-b\theta - \theta \sum_{i=1}^n x_i)} I(x > 0) \\
 p(\theta|x_{1:n}) &\propto \theta^{n+a-1} e^{-(b + \sum_{i=1}^n x_i)\theta} I(x > 0)
 \end{aligned}$$

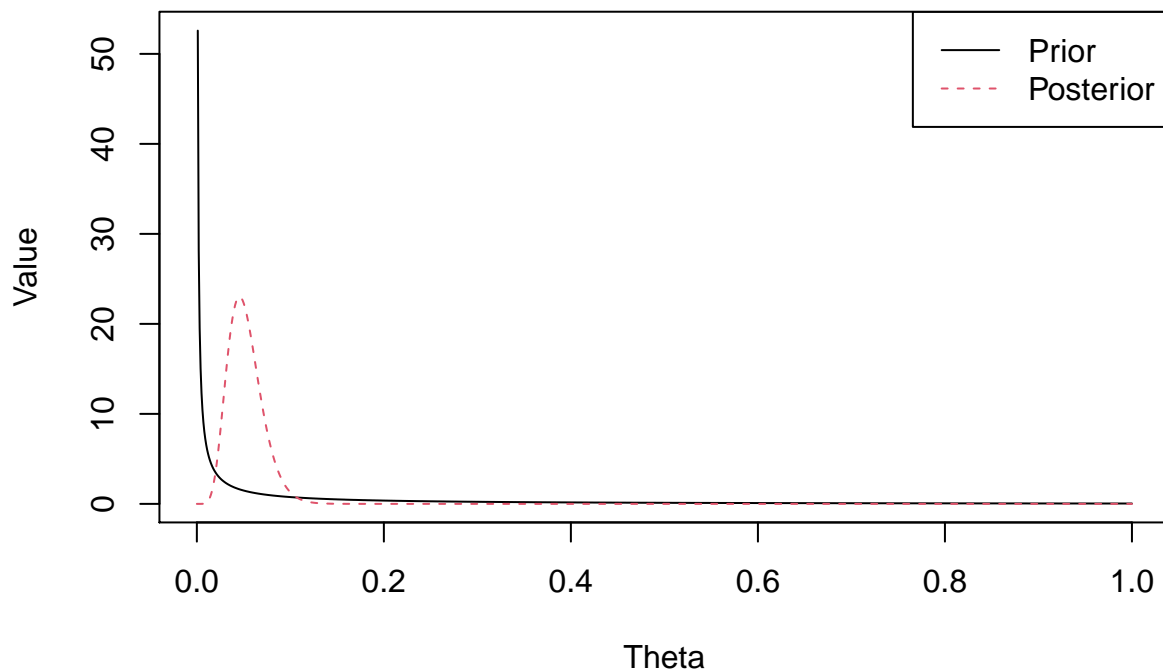
As can be seen above:

$$p(\theta|x_{1:n}) \sim \text{Gamma}(\theta|n + a, b + \sum_{i=1}^n x_i)$$

b. The posterior distribution is a proper probability distribution function because as a probability function, it must integrate to 1.

c.

```
### Record data
x <- c(20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0)
### Theta
theta = seq(0, 1, length.out = 1000)
n = length(x)
x_sum = sum(x)
a = 0.1
b = 1
plot(theta, dgamma(theta, shape = a, rate = b), type = "l", ylab = "Value", xlab = "Theta")
lines(theta, dgamma(theta, shape = n + a, rate = x_sum+b), col = 2, lty = 2)
legend("topright", legend = c("Prior", "Posterior"),
      lty = c(1,2), col = c(1,2))
```

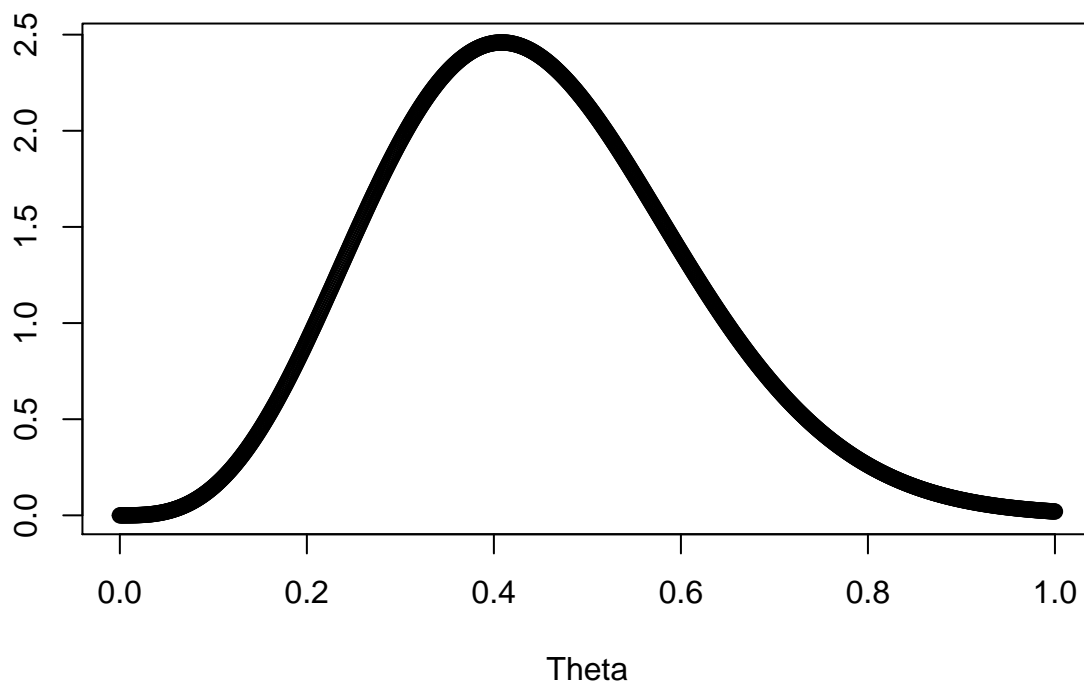


d. An Exponential model would be reasonable when, for example, modeling the time it takes for the next plane to arrive at an airport. An Exponential model would not be appropriate when, for example, modeling the time it takes for leaves to fall off of a tree because this is highly dependent on the time of year. There would need to be some sort of memory included in the system to make an effective model, but the Exponential model is memoryless.

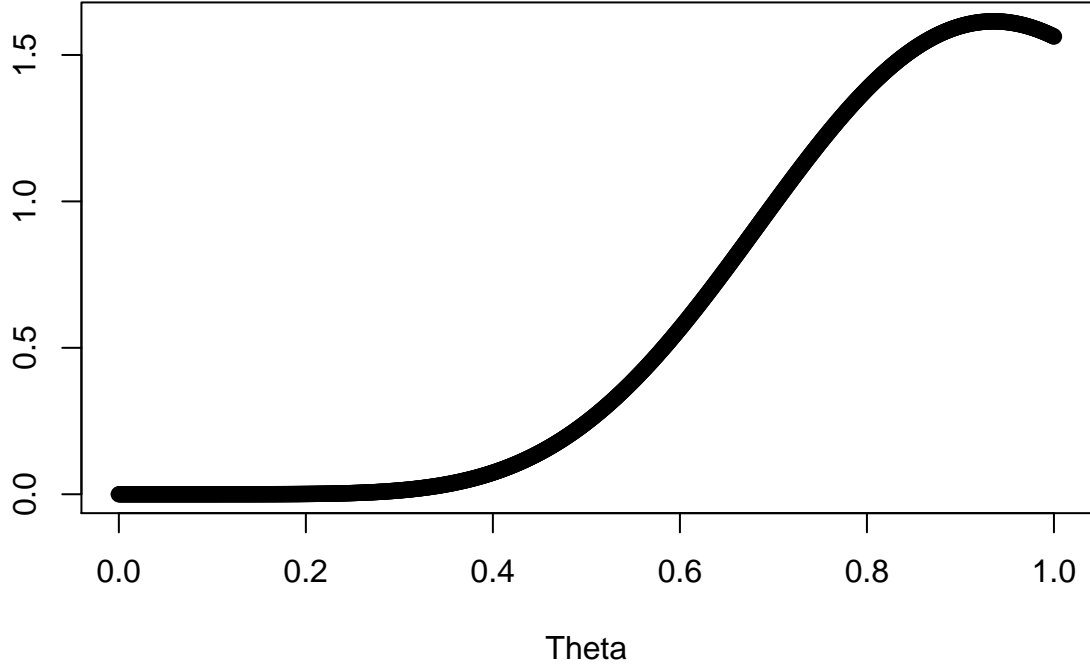
Priors, Posteriors, Predictive Distributions

- a. A class of conjugate prior densities for θ would have the kernel $\theta^{2b-1}e^{-\theta^2 c^2}$, so $\theta \sim \text{Galenshore}(b, c)$.

```
### Plotting prior class
theta = seq(0,1, length.out = 1000)
galenshore <- function (b, c, theta) {
  return (2/(factorial(b-1))* c^(2*b) * theta^(2*b-1) * exp(-theta^2 * c^2))
}
# As an example, choose b=2, c=3
b = 2
c = 3
plot(theta, galenshore(b, c, theta), xlab = "Theta", ylab = "")
```



```
# Another example, choose b=4, c=2
b = 4
c = 2
plot(theta, galenshore(b, c, theta), xlab = "Theta", ylab = "")
```



b.

$$\begin{aligned}
p(y \mid \theta) &= \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2} \\
p(y_{1:n} \mid \theta) &= \left(\frac{2}{\Gamma(a)} \right)^n \theta^{2an} e^{(-\theta^2 \sum_{i=1}^n y_i^2)} \left(\prod_{i=1}^n y_i^{2a-1} \right) \\
p(\theta) &\propto \theta^{2b-1} e^{-\theta^2 c^2} \\
p(\theta \mid y_{1:n}) &\propto p(y_{1:n} \mid \theta) p(\theta) \\
p(\theta \mid y_{1:n}) &\propto \left(\frac{2}{\Gamma(a)} \right)^n \theta^{2an} e^{(-\theta^2 \sum_{i=1}^n y_i^2)} \left(\prod_{i=1}^n y_i^{2a-1} \right) \theta^{2b-1} e^{-\theta^2 c^2} \\
p(\theta \mid y_{1:n}) &\propto \theta^{2an} e^{(-\theta^2 \sum_{i=1}^n y_i^2)} \theta^{2b-1} e^{-\theta^2 c^2} \\
p(\theta \mid y_{1:n}) &\propto \theta^{2an+2b-1} e^{(-\theta^2 \sum_{i=1}^n y_i^2) - \theta^2 c^2} \\
p(\theta \mid y_{1:n}) &\propto \theta^{2(an+b)-1} e^{-\theta^2 (\sum_{i=1}^n y_i^2 + c^2)}
\end{aligned}$$

We can see that

$$p(\theta \mid y_{1:n}) \sim \text{Galenshore} \left(an + b, \sqrt{c^2 + \sum_{i=1}^n (y_i^2)} \right)$$

c.

$$\begin{aligned}\frac{p(\theta_a \mid y_{1:n})}{p(\theta_b \mid y_{1:n})} &= \frac{\frac{2}{\Gamma(an+b)} ((\sum_{i=1}^n y_i^2) + c^2)^{(an+b)} \theta_a^{2(an+b)-1} e^{-\theta_a^2((\sum_{i=1}^n y_i^2)+c^2)}}{\frac{2}{\Gamma(an+b)} ((\sum_{i=1}^n y_i^2) + c^2)^{(an+b)} \theta_b^{2(an+b)-1} e^{-\theta_b^2((\sum_{i=1}^n y_i^2)+c^2)}} \\ \frac{p(\theta_a \mid y_{1:n})}{p(\theta_b \mid y_{1:n})} &= \frac{\theta_a^{2(an+b)-1} e^{-\theta_a^2((\sum_{i=1}^n y_i^2)+c^2)}}{\theta_b^{2(an+b)-1} e^{-\theta_b^2((\sum_{i=1}^n y_i^2)+c^2)}} \\ \frac{p(\theta_a \mid y_{1:n})}{p(\theta_b \mid y_{1:n})} &= \frac{\theta_a}{\theta_b} e^{(-\theta_a^2+\theta_b^2)((\sum_{i=1}^n y_i^2)+c^2)}\end{aligned}$$

d. The expected value of a Galenshore distribution with parameters a and θ is $\frac{\Gamma(a+\frac{1}{2})}{\theta\Gamma(a)}$. Because $p(\theta \mid y_{1:n})$ is a Galenshore distribution with parameters $an+b$ and $\sqrt{c^2 + \sum_{i=1}^n (y_i^2)}$, its expected value can be expressed as:

$$E[\theta \mid y_{1:n}] = \frac{\Gamma(na+b+\frac{1}{2})}{\left(\sqrt{c^2 + \sum_{i=1}^n (y_i^2)}\right) \Gamma(an+b)}$$

e.

$$\begin{aligned}p(y_{n+1} \mid y_{1:n}) &= \int p(y_{n+1} \mid \theta) p(\theta \mid y_{1:n}) d\theta \\ p(y_{n+1} \mid y_{1:n}) &= \int \left[\frac{2}{\Gamma(a)} \theta^{2a} y_{n+1}^{2a-1} e^{-\theta^2 y_{n+1}^2} \right] \left[\frac{2}{\Gamma(an+b)} \left(\sqrt{c^2 + \sum_{i=1}^n (y_i^2)} \right)^{2(an+b)} \theta^{2(an+b)-1} e^{-\theta^2 (\sqrt{c^2 + \sum_{i=1}^n (y_i^2)})^2} \right] d\theta \\ p(y_{n+1} \mid y_{1:n}) &= \frac{4y_{n+1}^{2a-1} (c^2 + \sum_{i=1}^n (y_i^2))^{(an+b)}}{\Gamma(a)\Gamma(an+b)} \int \theta^{2a} e^{-\theta^2 y_{n+1}^2} \theta^{2(an+b)-1} e^{-\theta^2 (c^2 + \sum_{i=1}^n (y_i^2))} d\theta \\ p(y_{n+1} \mid y_{1:n}) &= \frac{4y_{n+1}^{2a-1} (c^2 + \sum_{i=1}^n (y_i^2))^{(an+b)}}{\Gamma(a)\Gamma(an+b)} \int \theta^{2(an+a+b)-1} e^{-\theta^2 (y_{n+1}^2 + c^2 + \sum_{i=1}^n (y_i^2))} d\theta \\ p(y_{n+1} \mid y_{1:n}) &= \frac{4y_{n+1}^{2a-1} (c^2 + \sum_{i=1}^n (y_i^2))^{(an+b)}}{\Gamma(a)\Gamma(an+b)} \int \theta^{2(an+a+b)-1} e^{-\theta^2 (y_{n+1}^2 + c^2 + \sum_{i=1}^n (y_i^2))} d\theta \times \\ &\quad \left[\frac{2}{\Gamma(an+a+b)} \sqrt{y_{n+1}^2 + c^2 + \sum_{i=1}^n (y_i^2)} \right] \left[\frac{\Gamma(an+a+b)}{2} \frac{1}{\sqrt{y_{n+1}^2 + c^2 + \sum_{i=1}^n (y_i^2)}^{2(an+a+b)}} \right] \\ p(y_{n+1} \mid y_{1:n}) &= \frac{4y_{n+1}^{2a-1} (c^2 + \sum_{i=1}^n (y_i^2))^{(an+b)}}{\Gamma(a)\Gamma(an+b)} \left[\frac{\Gamma(an+a+b)}{2} \frac{1}{(y_{n+1}^2 + c^2 + \sum_{i=1}^n (y_i^2))^{an+a+b}} \right] \\ p(y_{n+1} \mid y_{1:n}) &= \frac{2y_{n+1}^{2a-1} (c^2 + \sum_{i=1}^n (y_i^2))^{(an+b)}}{\Gamma(a)\Gamma(an+b)} \frac{\Gamma(an+a+b)}{(y_{n+1}^2 + c^2 + \sum_{i=1}^n (y_i^2))^{an+a+b}}\end{aligned}$$