

Homework 2

Chavez Cheong

1/13/2022

1 Lab Component

Task 3

```
# 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)

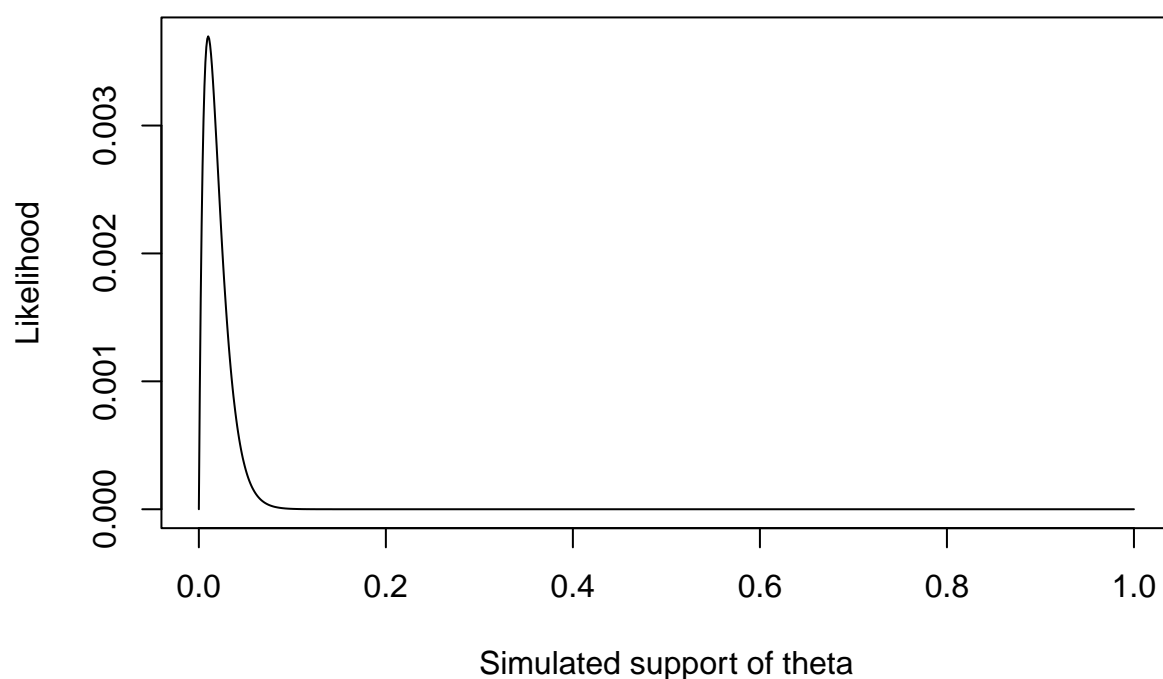
## [1] 100

### Bernoulli LH Function ###
# Input: obs.data, theta
# Output: bernoulli likelihood

bernoulliLHFunction <- function(obs.data, theta) {
  x <- sum(obs.data)
  n <- length(obs.data)
  like <- theta ^ x * (1 - theta) ^ (n-x)
  return(like)
}

### Plot LH for a grid of theta values ###
# Create the grid #
# Store the LH values
# Create the Plot
theta <- seq(0, 1, length = 1000)
like <- bernoulliLHFunction(obs.data, theta)
plot(theta, like, type = "l", ylab = "Likelihood", lty = 1, lwd = 1, xlab = "Simulated support of theta")
```

Likelihood Profile



Task 4

```
### Beta Bernoulli Posterior Function
# Input: a, b, obs.data
# Output: posterior.a, posterior.b (Hyperparameters for the posterior distribution)

betabernoulliPosteriorFunction <- function(a, b, obs.data) {
  x <- sum(obs.data)
  n <- length(obs.data)
  posterior.a <- x + a
  posterior.b <- n - x + b
  return(c(posterior.a, posterior.b))
}
```

```
betabernoulliPosteriorFunction(1, 1, obs.data)
```

Non-Informative Posterior

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

```
betabernoulliPosteriorFunction(3, 1, obs.data)
```

Informative Posterior

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

Task 5

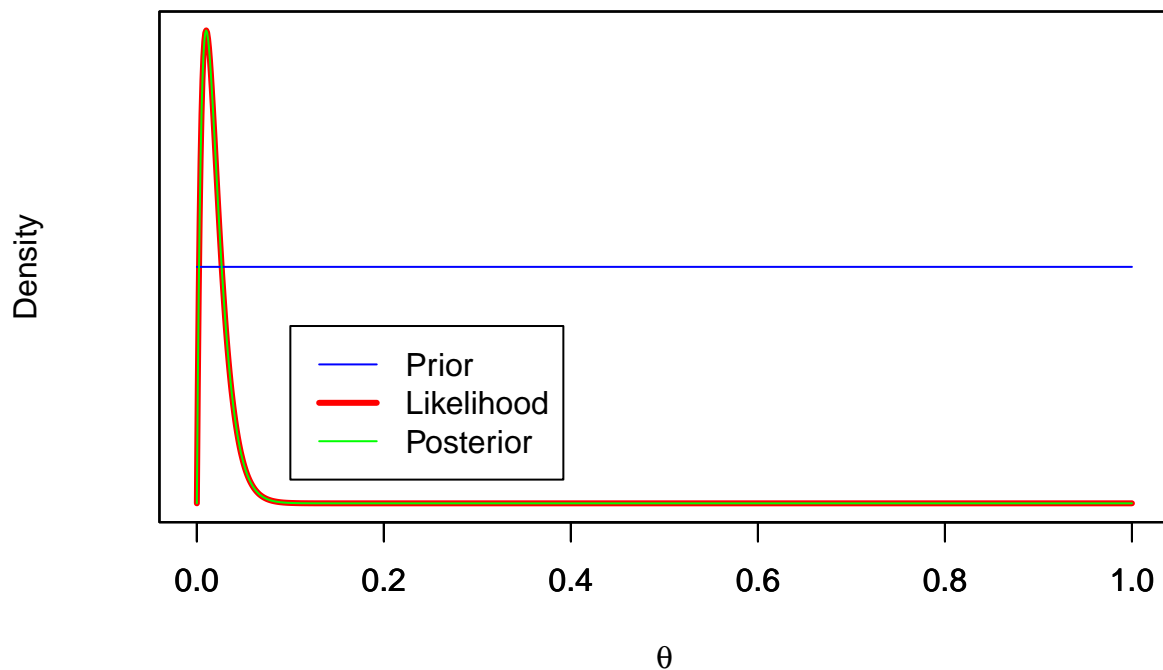
Generating Parameters and Distributions

```
informative.posterior.a <- betabernoulliPosteriorFunction(3, 1, obs.data)[1]
informative.posterior.b <- betabernoulliPosteriorFunction(3, 1, obs.data)[2]
noninformative.posterior.a <- betabernoulliPosteriorFunction(1, 1, obs.data)[1]
noninformative.posterior.b <- betabernoulliPosteriorFunction(1, 1, obs.data)[2]
informative.posterior <- dbeta(theta, informative.posterior.a, informative.posterior.b)
noninformative.posterior <- dbeta(theta, noninformative.posterior.a, noninformative.posterior.b)
informative.prior <- dbeta(theta, 3, 1)
noninformative.prior <- dbeta(theta, 1, 1)
```

Plot Non-Informative Prior

```
plot(theta, like, type = "l", ylab = "Density", yaxt = "n", col = "red", lwd = 3, xlab = expression(theta),
      par(new = TRUE))
plot(theta, noninformative.prior, ylab = "", yaxt = "n", col = "blue", lwd = 1, xlab = "", type = "l")
par(new = TRUE)
plot(theta, noninformative.posterior, ylab = "", yaxt = "n", col = "green", lwd = 1, xlab = "", type = "l")
legend(0.1, 14, c("Prior", "Likelihood", "Posterior"), col = c("blue", "red", "green"), lwd = c(1, 3, 1))
```

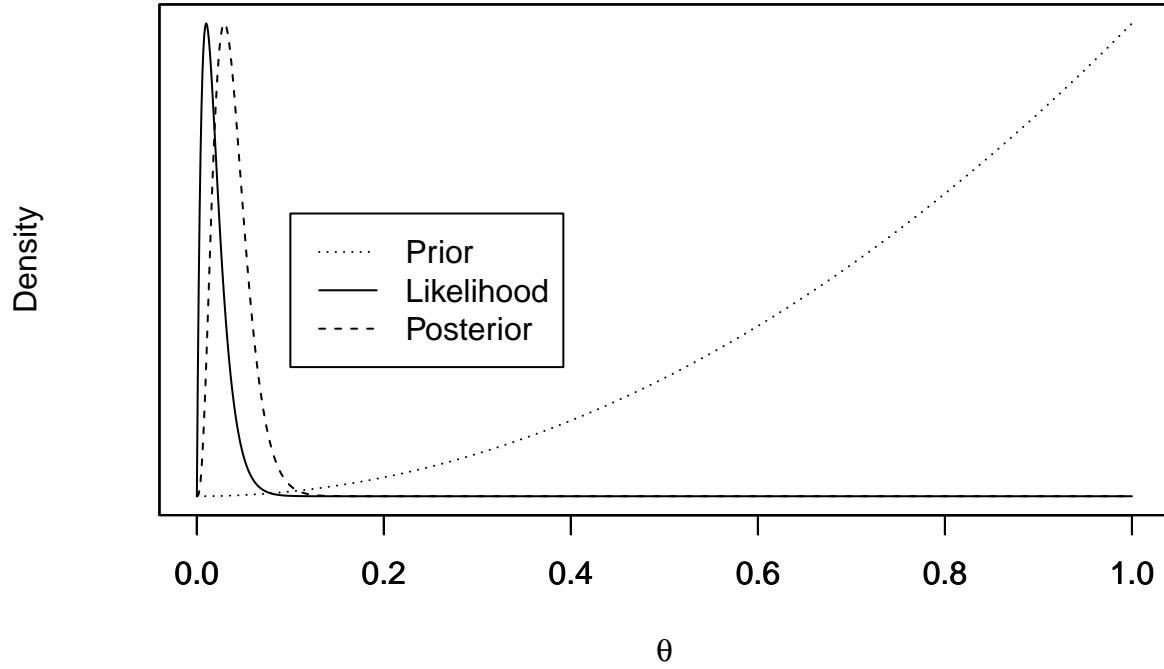
Non-Informative Prior



Plot Informative Prior Case

```
plot(theta, like, type = "l", ylab = "Density", yaxt = "n", lty = 1, lwd = 1, xlab = expression(theta),
      par(new = TRUE))
plot(theta, informative.prior, ylab = "", yaxt = "n", lty = 3, lwd = 1, xlab = "", type = "l")
par(new = TRUE)
plot(theta, informative.posterior, ylab = "", yaxt = "n", lty = 2, lwd = 1, xlab = "", type = "l")
legend(0.1, 14, c("Prior", "Likelihood", "Posterior"), lty = c(3, 1, 2), lwd = c(1, 1, 1))
```

Informative Prior



Analysis When the prior is non-informative, we see that the posterior is almost identical to the likelihood from the sample data. However, when the prior is more informative relative to the likelihood from the sample data, the posterior distribution will be more heavily affected by the prior distribution.

2 The Exponential-Gamma Model

a

Likelihood:

$$\begin{aligned}
 p(x_{1:n}|\theta) &= P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n|\theta) \\
 &= \prod_{i=1}^n P(X_i = x_i|\theta) \\
 &= \prod_{i=1}^n p(x_i|\theta) \\
 &= \prod_{i=1}^n \theta e^{-\theta x_i} \\
 &= \theta^n e^{-\theta \sum_{i=1}^n x_i}
 \end{aligned}$$

Posterior:

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

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

b

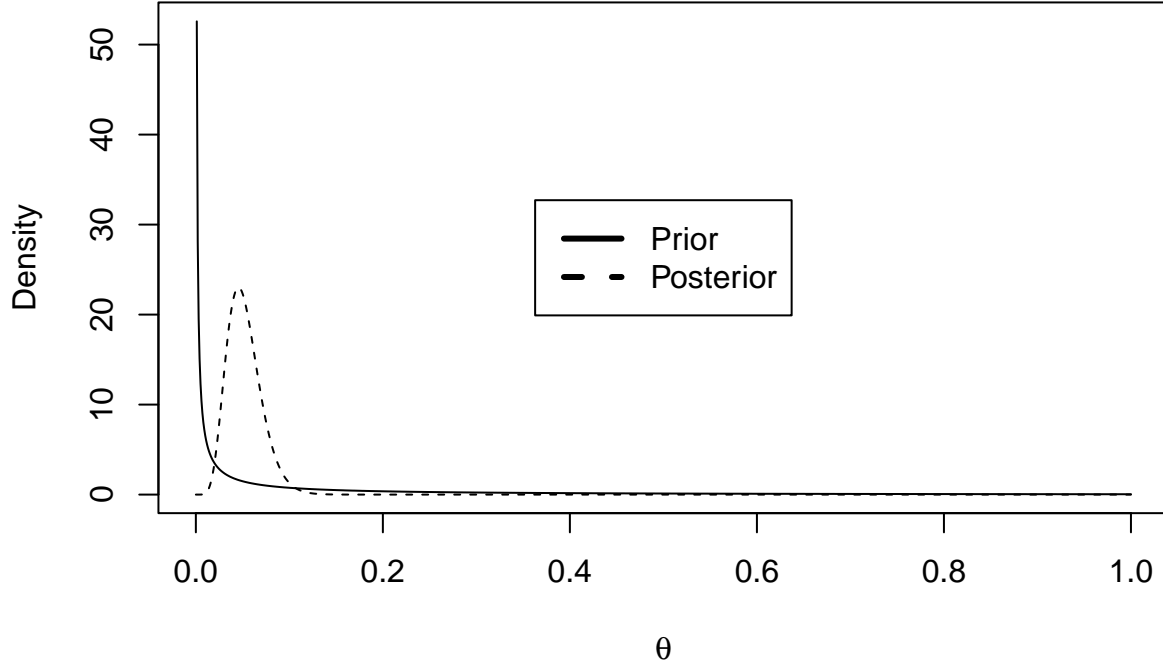
With a proper prior distribution ($\text{Gamma}(a, b)$, $a > 0, b > 0$) and a proper sample likelihood distribution (Exponential), the resulting posterior distribution must be a proper distribution. Additionally, we see after multiplying the prior and likelihood to obtain an equation proportional to the posterior distribution, that the resulting equation fits the kernel of a Gamma distribution with both hyperparameters $\alpha = n + a > a > 0$ and $\beta = b + \sum_{i=1}^n x_i > b > 0$, which fits the condition of a Gamma distribution, hence we see once again that the posterior Gamma distribution is a proper density.

c

```
### Model sample data
strike.data <- c(20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0)
x <- sum(strike.data)
n <- length(strike.data)
gamma_theta <- seq(0, 1, length = 1000)
### Calculate prior
prior.alpha <- 0.1
prior.beta <- 1
prior <- dgamma(gamma_theta, shape = prior.alpha, rate = prior.beta)
### Calculate posterior
posterior.alpha <- n + prior.alpha
posterior.beta <- x + prior.beta
posterior <- dgamma(gamma_theta, shape = posterior.alpha, rate = posterior.beta)

### Plot prior and posterior
plot(gamma_theta, prior, type = "l", ylab = "Density", lty = 1, lwd = 1, xlab = expression(theta), main = "", yaxp = 1, yaxt = "n")
lines(gamma_theta, posterior, ylab = "", yaxt = "n", lty = 2, lwd = 1, xlab = "", type = "l")
legend("center", c("Prior", "Posterior"), lty = c(1,2), lwd = c(3,3))
```

Exponential–Gamma Model



d

An exponential model would be reasonable to model the time until a car engine fails for the first time. It would not be reasonable to use an exponential model to model the time between car engine failures, as one of the key properties of the exponential distribution is that the rate, which does not hold here as after the first car engine failure, it is likely that some part of the car engine is damaged and thus the time between subsequent car engine failures is likely to be shorter.

3 The Galenshore Distribution

a

The conjugate distribution is most likely a Galenshore distribution, as when we look at kernels of the Galenshore distribution ($\theta^{2c-1}e^{-d^2\theta^2}$), we see elements like e raised to the power of the square of the parameter of interest (e^{θ^2}) and the parameter itself raised to the power of a hyperparameter, which suggests that when we multiply two Galenshore distributions, one from the likelihood and one from the prior, and take the subsequent distribution proportional to θ , we should be able to obtain a kernel of another Galenshore distribution, which makes the Galenshore a suitable class of conjugate prior densities for θ .

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

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

Four Galenshore Distributions, $\theta_a \sim \text{Galenshore}(1, 1)$, $\theta_b \sim \text{Galenshore}(1, 5)$, $\theta_c \sim \text{Galenshore}(5, 1)$ and $\theta_d \sim \text{Galenshore}(5, 5)$ have been plotted.

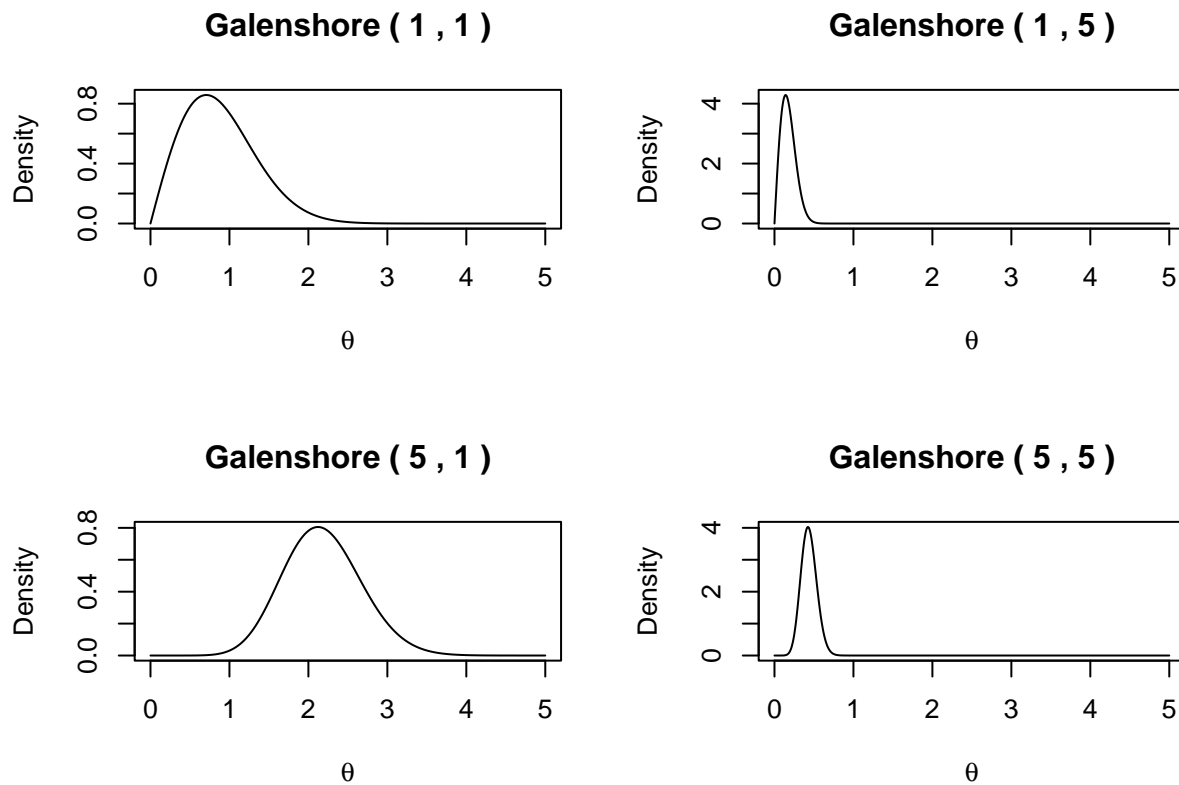
```

### Galenshore Plotting Function ###
# Input: c, d, theta
# Output: Galenshore Density Plot

galenshorePlotter <- function(c, d) {
  theta <- seq(0,5, length = 1000)
  density <- (2 / gamma(c)) * (d ^ (2 * c)) * theta ^ (2 * c - 1) * exp(-(d ^ 2) * (theta ^ 2))
  plot(theta, density, type = "l", lty = 1, xlab = expression(theta), ylab = "Density", main = paste("Galenshore (", c, ", ", d, ")"))
}

### Plot four Galenshores
par(mfrow = c(2,2))
galenshorePlotter(1,1)
galenshorePlotter(1,5)
galenshorePlotter(5,1)
galenshorePlotter(5,5)

```



b

Prior:

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

$$\propto \theta^{2c-1} e^{-d^2 \theta^2}$$

Likelihood:

$$\begin{aligned}
p(y_{1:n}|\theta) &= (Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n|\theta) \\
&= \prod_{i=1}^n P(Y_i = y_i|\theta) \\
&= \prod_{i=1}^n p(y_i|\theta) \\
&= \prod_{i=1}^n \frac{2}{\Gamma(a)} \theta^{2a} y_i^{2a-1} e^{-\theta^2 y_i^2} \\
&\propto \theta^{2an} e^{-\theta^2 \sum y_i^2}
\end{aligned}$$

Posterior:

$$\begin{aligned}
p(\theta|y_{1:n}) &\propto p(\theta)p(y_{1:n}|\theta) \\
&\propto \theta^{2c-1} e^{-d^2 \theta^2} \theta^{2an} e^{-\theta^2 \sum y_i^2} \\
&\propto \theta^{2(an+c)-1} e^{-\theta^2 (d^2 + \sum y_i^2)} \\
\theta|y_{1:n} &\sim \text{Galenshore}(an + c, \sqrt{d^2 + \sum y_i^2})
\end{aligned}$$

c

We have

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

Thus, from the posterior distribution of $\theta|y_{1:n}$ we found in part (b), we find that

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

For ease of calculation, let $x = an + c$ and $z = \sqrt{d^2 + \sum y_i^2}$

$$\begin{aligned}
\frac{p(\theta_a|y_{1:n})}{p(\theta_b|y_{1:n})} &= \frac{\frac{2}{\Gamma(x)} z^{2x} \theta_a^{2x-1} e^{-z^2 \theta_a^2}}{\frac{2}{\Gamma(x)} z^{2x} \theta_b^{2x-1} e^{-z^2 \theta_b^2}} \\
&= \left(\frac{\theta_a}{\theta_b}\right)^{2x-1} e^{-(\theta_a^2 - \theta_b^2) z^2}
\end{aligned}$$

Substituting back $x = an + c$ and $z = \sqrt{d^2 + \sum y_i^2}$

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

as required.

The sufficient statistic $T(y) = \boxed{\sum y_i^2}$.

d

We are given that the expected value for a Galenshore distribution $Y \sim \text{Galenshore}(a, \theta)$ is $E[Y] = \frac{\Gamma(a+1/2)}{\theta\Gamma(a)}$.

Thus

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

e

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

As required. Note that $\int \text{Galenshore}(an + n_c, (d^2 + \sum y_i^2 + y_{n+1}^2)^{(an+a+c)}) d\theta = 1$ because the Galenshore distribution is a proper distribution.