

Homework 2

Garrett Allen

Friday, September 3rd, 2021

Problem 1

Task 2 (not graded, just here for setup for 3,4,5)

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

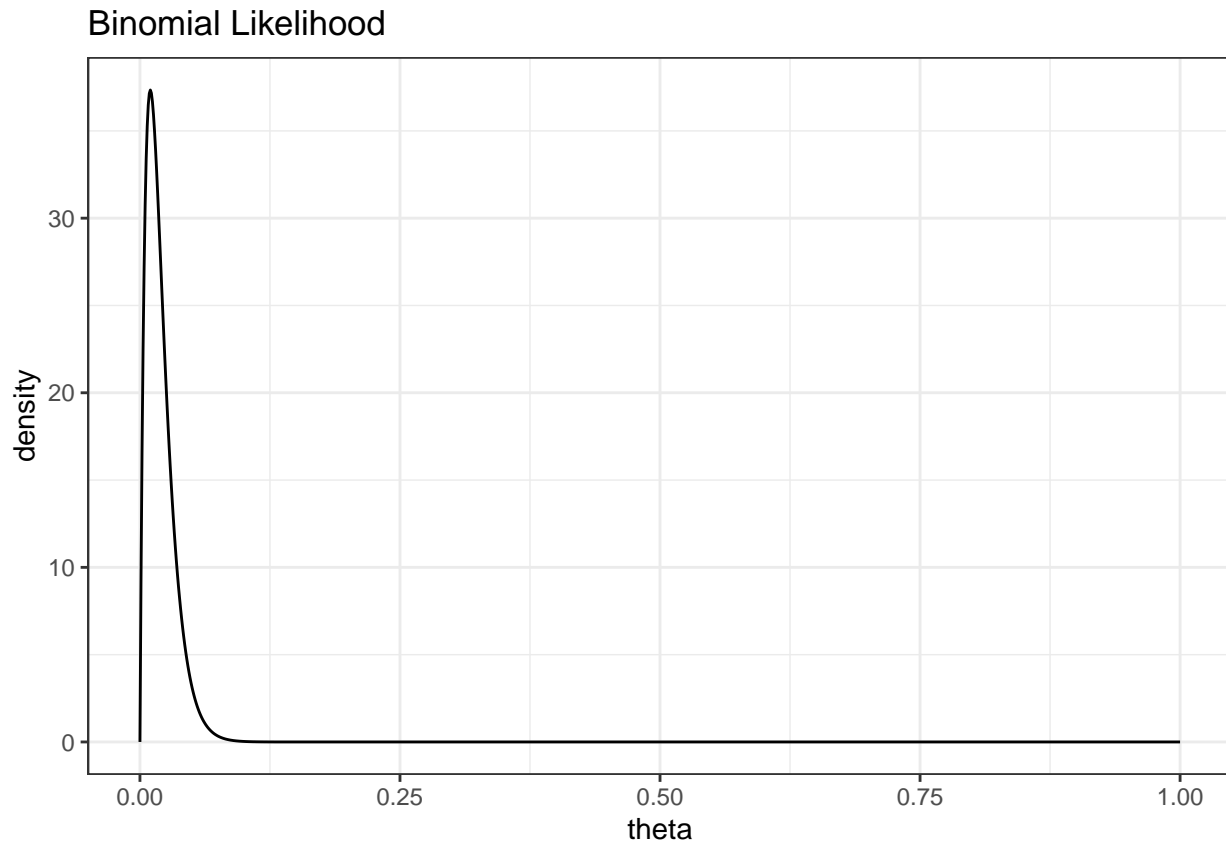
```
### Bernoulli LH Function ###
# Input: obs.data, theta
# Output: bernoulli likelihood

BernoulliLhFunction <- function(obs.data, theta) {
  n <- length((obs.data))
  x <- sum(obs.data)
  like <- dbeta(theta, x + 1, n - x + 1)
  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)
```

```
df <- data.frame(theta,like)

df %>%
  ggplot(aes(x = theta, y = like)) +
    geom_line() +
    labs(
      title = "Binomial Likelihood",
      x = "theta",
      y = "density") +
    theme_bw()
```



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 $(a,b) = (3,1)$.

```
#generate posterior beta appropriate parameters
posteriorGenerate <- function(a,b, obs.data) {
  n <- length(obs.data)
  x <- sum(obs.data)
  a.post <- x + a
  b.post <- n - x + b
  return(c(a.post,b.post))
}
#generate non-informative paramater, and informative parameter
```

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

## [1] 2 100

posteriorGenerate(3,1,obs.data)

## [1] 4 100

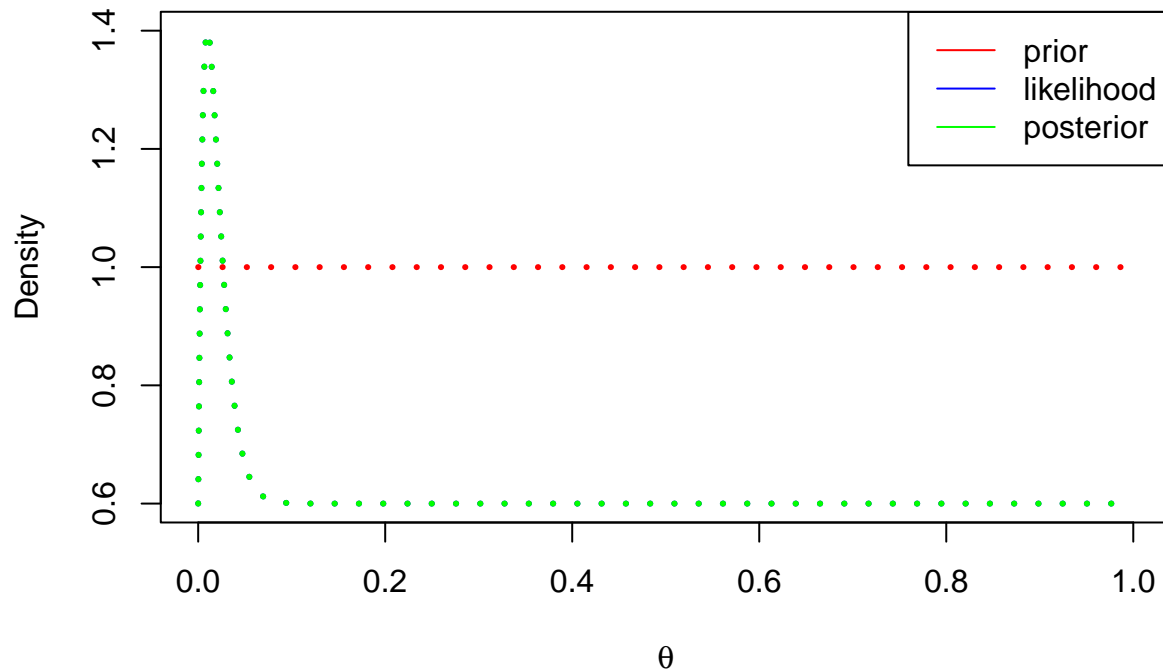
noninform_param <- posteriorGenerate(1,1,obs.data)
inform_param <- posteriorGenerate(3,1,obs.data)
```

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.

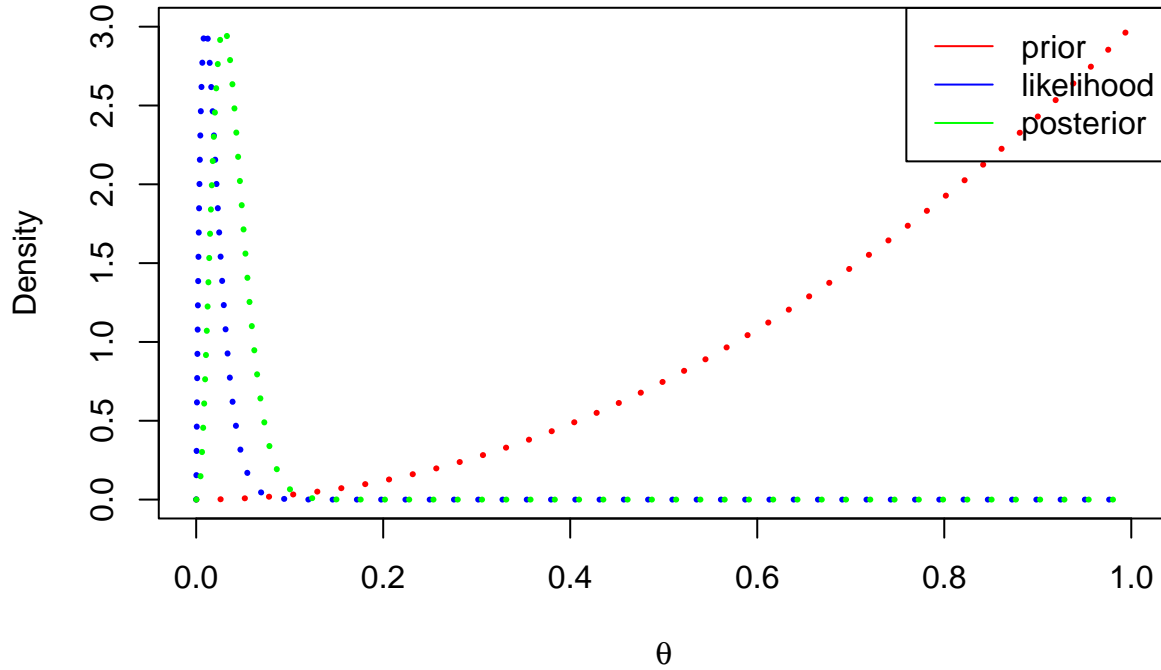
```
#creating noninform_prior, likelihood, and posterior resulting
noninform_prior <- dbeta(theta, 1, 1)
like <- BernoulliLhFunction(obs.data, theta)
noninform_posterior <- dbeta(theta, noninform_param[1], noninform_param[2])

plot(theta, noninform_prior, type = "l", ylab = "Density",
      lty = 3, lwd = 3, xlab = expression(theta), col = "red")
par(new = TRUE)
plot(theta, like, type = "l",
      lty = 3, lwd = 3, axes = FALSE, col = "blue",
      xlab = "", ylab = "")
par(new = TRUE)
plot(theta, noninform_posterior, type = "l", xlab = "", ylab = "",
      lty = 3, lwd = 3, axes = FALSE, col = "green")
par(new = TRUE)
legend('topright',
      legend = c("prior", "likelihood", "posterior"),
      col = c("red", "blue", "green"), lty = 1)
```



```
inform_prior <- dbeta(theta,3,1)
inform_posterior <- dbeta(theta, inform_param[1], inform_param[2])

plot(theta, inform_prior,type = "l", ylab = "Density",
      lty = 3, lwd = 3, xlab = expression(theta), col = "red")
par(new = TRUE)
plot(theta, like, type = "l",
      lty = 3, lwd = 3, axes = FALSE, col = "blue",
      xlab = "", ylab = "")
par(new = TRUE)
plot(theta, inform_posterior, type = "l", xlab = "", ylab = "",
      lty = 3, lwd = 3, axes = FALSE, col = "green")
par(new = TRUE)
legend('topright',
      legend = c("prior","likelihood","posterior"),
      col = c("red","blue","green"), lty = 1)
```



Problem 2

2a, 2b

We start by applying Baye's formula. Then, we apply proportionality to recognize that the posterior is proportional to a known distribution.

$$p(\theta|x_1, \dots, x_n) = \frac{p(x_1, \dots, x_n|\theta)p(\theta)}{p(x)}$$

Then:

$$\begin{aligned} p(\theta|x_1, \dots, x_n) &\propto p(x_1, \dots, x_n|\theta)p(\theta) \\ p(\theta|x_1, \dots, x_n) &\propto \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \theta^n * e^{-\theta * \sum_i x_i} \\ p(\theta|x_1, \dots, x_n) &\propto \theta^{a+n-1} e^{-\theta(\sum_i x_i + b)} \\ p(\theta|x_1, \dots, x_n) &\propto \text{Gamma}(\theta|a+n, \sum_i x_i + b) \end{aligned}$$

Thus the posterior distribution is proportional to a gamma distribution, and thus it is a proper distribution (integrates) to 1. We can calculate the normalizing constant by finding the value for $p(x)$.

$$\begin{aligned} p(x) &= \int_0^1 \frac{b^a}{\Gamma(a)} \text{Gamma}(\theta|a+n, \sum_i x_i + b) d\theta \\ p(x) &= \frac{b^a}{\Gamma(a)} \frac{\Gamma(a+n)}{(\sum_i x_i + b)^{a+n}} \int_0^1 \frac{(\sum_i x_i + b)^{a+n}}{\Gamma(a+n)} \text{Gamma}(\theta|a+n, \sum_i x_i + b) d\theta \end{aligned}$$

Since the integral is now a proper distribution, it integrates to 1. Thus we have:

$$p(x) = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a+n)}{(\sum_i x_i + b)^{a+n}}$$

and our full distribution:

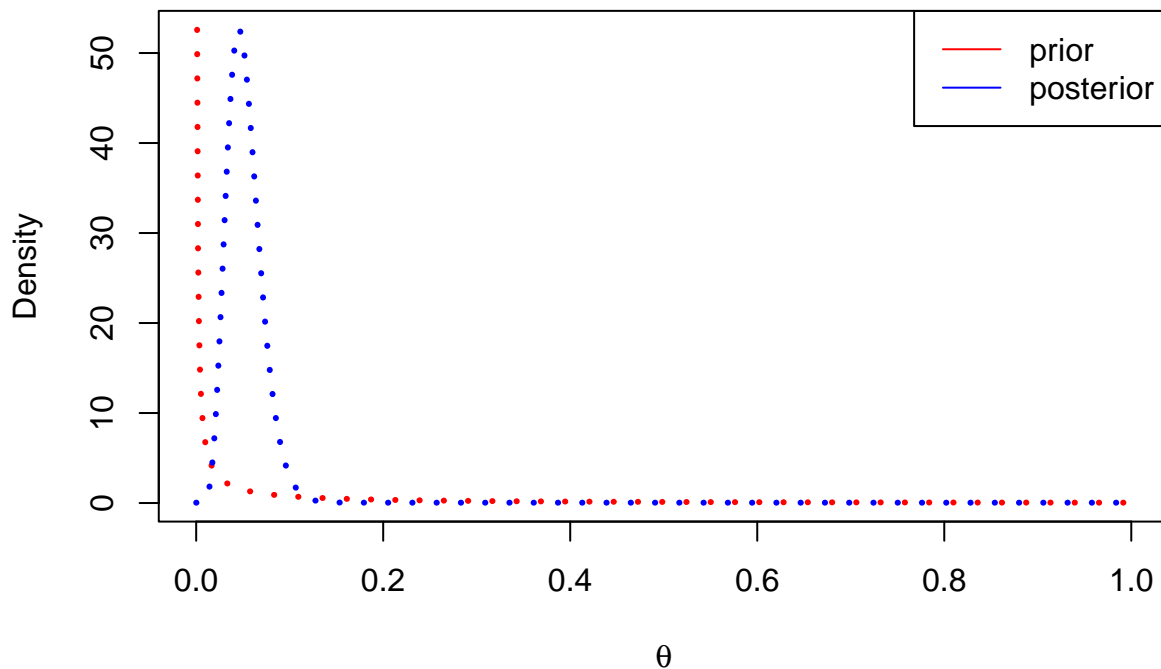
$$p(\theta|x_1, \dots, x_n) = \text{Gamma}(\theta|a + n, \sum_i x_i + b)$$

2c

```
obs.data = c(20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0)
theta = seq(0,1,length = 1000)
prior = dgamma(theta, shape = .1, rate = 1)
sum = sum(obs.data)
posterior = dgamma(theta, shape = 8.1, rate = (sum + 1))
max(posterior)

## [1] 23.0042

plot(theta, prior, type = "l", ylab = "Density",
      lty = 3, lwd = 3, xlab = expression(theta), col = "red")
par(new = TRUE)
plot(theta, posterior, type = "l",
      lty = 3, lwd = 3, axes = FALSE, col = "blue",
      xlab = "", ylab = "")
par(new = TRUE)
legend('topright', legend = c("prior", "posterior"),
      col = c("red", "blue"), lty = 1)
```



2d

An appropriate usage of the exponential distribution would be to model the arrival time between customers in a store, where the rate is fixed. . An event that would not be appropriate to model with an exponential distribution is the interarrival time between customers, where the rate is also variable, as the exponential distribution works under the assumption that the underlying rate of arrivals is fixed.

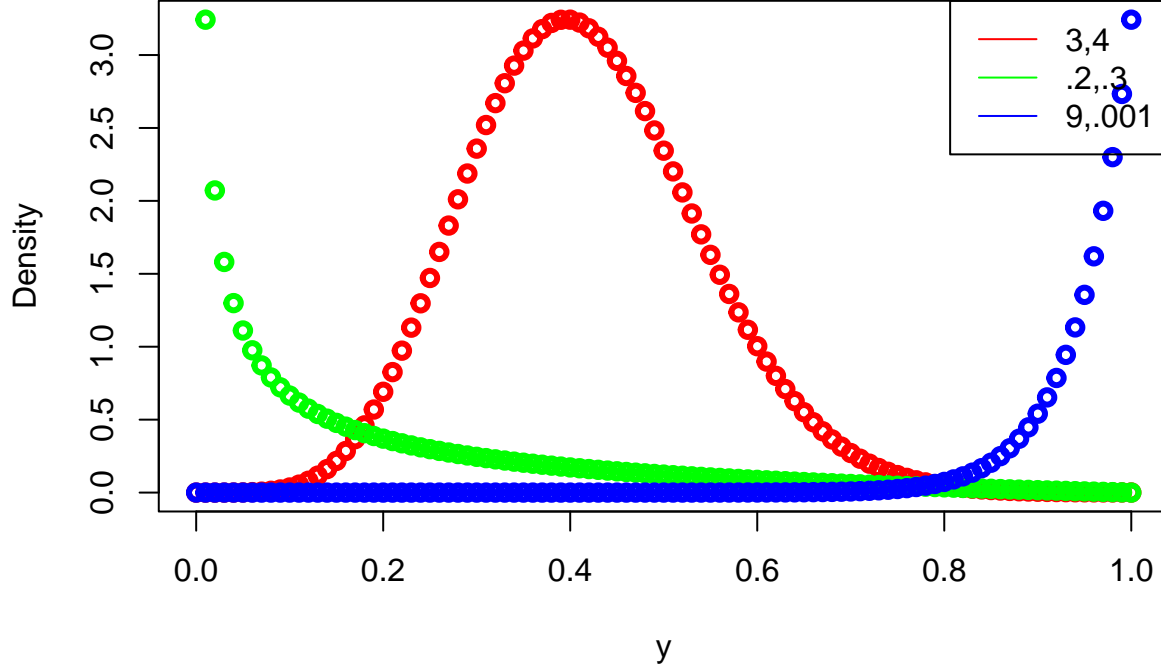
Problem 3

3a

```
y <- seq(0,1,by = .01) #initializing y values
galenshore <- function(y,a,theta){ #creating a galenshore from parameters

  galenshore <- (2/gamma(a))*theta^(2*a)*y^(2*a-1)*exp((-theta^2)*y^2)
  return(galenshore)
}

plot(x = y,
     y = galenshore(y,3,4), #<- plotting this
     ylab = "Density",
     xlab = "y",
     lty = 3,
     lwd = 3,
     col = "red")
par(new = TRUE)
plot(x = y,
     y = galenshore(y,.2,.3), # <- plotting this
     axes = FALSE,
     lty = 3,
     lwd = 3,
     col = "green",
     xlab = "",
     ylab = "")
par(new = TRUE)
plot(x = y,
     y = galenshore(y,9,.001), # <- plotting this
     axes = FALSE,
     lty = 3,
     lwd = 3,
     col = "blue",
     xlab = "",
     ylab = "")
par(new = TRUE)
legend('topright',legend = c("3,4",".2,.3","9,.001"), col = c("red","green","blue"), lty = 1)
```



I suspected that a Galenshore distribution will be a conjugate prior for itself, so I decided to pick this as my prior distribution. It turns out this is a good choice, as seen by the next part of the question. Above the plot of a few Galenshore distributions with $0 < y < 1$.

3b

We will prove that the Galenshore distribution is a conjugate prior for itself, and prove the form of the posterior distribution. First we start with Baye's rule, with $\theta \sim \text{Galenshore}(c, d)$ and $Y_1, \dots, Y_n | \theta \sim \text{Galenshore}(a, \theta)$.

$$p(\theta | y_1, \dots, y_n) \propto p(y_1, \dots, y_n | \theta) p(\theta)$$

$$p(\theta | y_1, \dots, y_n) \propto \left(\frac{2}{\Gamma(a)} \right)^n \theta^{2an} \prod_i y_i^{2a-1} e^{-\theta^2 (\sum_i y_i^2)} \frac{2}{\Gamma(c)} d^{2c} \theta^{2c-1} e^{-d^2 \theta^2}$$

This looks very complicated at first, but most of these things are constants if we treat this as a function of θ . After removing constants, we have:

$$p(\theta | y_1, \dots, y_n) \propto \theta^{2an} e^{-\theta^2 (\sum_i y_i^2)} \theta^{2c-1} e^{-d^2 \theta^2}$$

If we combine like terms, we get:

$$p(\theta | y_1, \dots, y_n) \propto \theta^{2(an+c)-1} e^{-\theta^2 ((\sum_i y_i^2) + d^2)}$$

We recognize this as the form of a Galenshore density, and we find that:

$$p(\theta | y_1, \dots, y_n) \propto \text{Galenshore} \left(an + c, \sqrt{\sum_i y_i^2 + d^2} \right)$$

3c

We need not know the constants for the Galenshore distribution when computing $\frac{p(\theta_a|y_1, \dots, y_n)}{p(\theta_b|y_1, \dots, y_n)}$ since they will be divided out anyways, so we can ignore writing them entirely. We write $s = \sum_i y_i^2$ to simplify notation. Now we simplify:

$$\frac{p(\theta_a|y_1, \dots, y_n)}{p(\theta_b|y_1, \dots, y_n)} = \frac{\theta_a^{2(an+c)-1} e^{-\theta_a^2(s+d^2)}}{\theta_b^{2(an+c)-1} e^{-\theta_b^2(s+d^2)}}$$

Combining terms, we get:

$$\frac{p(\theta_a|y_1, \dots, y_n)}{p(\theta_b|y_1, \dots, y_n)} = \left(\frac{\theta_a}{\theta_b}\right)^{2(an+c)-1} \left(e^{(\theta_b^2 - \theta_a^2)(s+d^2)}\right)$$

This shows that, for fixed parameters a, c, d we need only know s to compute the probability density at θ_a relative to θ_b . This means that s is a sufficient statistic for our model.

3d

To compute the $E(\theta|y_1, \dots, y_n)$ we need to compute:

$$\begin{aligned} E(\theta|y_1, \dots, y_n) &= \int_{-\infty}^{\infty} \theta p(\theta|y_1, \dots, y_n) d\theta \\ E(\theta|y_1, \dots, y_n) &= \int_{-\infty}^{\infty} \theta^{2(an+c)} e^{-\theta^2(s+d^2)} \left(\sqrt{d^2 + s}\right)^{2(an+c)} \frac{2}{\Gamma(an+c)} d\theta \end{aligned}$$

We recognize that the integrand is the kernel of Galenshore($an + c + 1/2, \sqrt{s + d^2}$). Therefore, if we multiply and divide by the constant necessary to make the integrand a proper distribution, the integral will be 1. Thus we will multiply and divide by

$$\frac{2 \left(\sqrt{d^2 + s}\right)^{2(an+c+1/2)}}{\Gamma(an + c + 1/2)}$$

After doing this, we will just be left with the reciprocal of the term above multiplied by the other constants we had outside the integrand. Thus we are left with:

$$\begin{aligned} E(\theta|y_1, \dots, y_n) &= \frac{\Gamma(an + c + 1/2)}{2 \left(\sqrt{s + d^2}\right)^{2(an+c+1/2)}} \frac{2 \left(\sqrt{s + d^2}\right)^{2(an+c)}}{\Gamma(an + c)} \\ E(\theta|y_1, \dots, y_n) &= \frac{\Gamma(an + c + 1/2)}{\left(\sqrt{s + d^2}\right) \Gamma(an + c)} \end{aligned}$$

3e

First we start by writing down the formula for our predictive posterior distribution.

$$p(y_{n+1}|y_1, \dots, y_n) = \int_{-\infty}^{\infty} p(y_{n+1}|\theta) p(\theta|y_1, \dots, y_n) d\theta$$

Then, we substitute in our distributions in the integral.

$$p(y_{n+1}|y_1, \dots, y_n) = \int_{-\infty}^{\infty} \frac{2}{\Gamma a} y_{n+1}^{2a-1} \theta^{2a} e^{-\theta^2 y_{n+1}^2} \frac{2}{\Gamma an + c} \left(\sqrt{s + d^2} \right)^{2(an+c)} \theta^{2(an+c)-1} e^{\theta^2(s+d^2)} d\theta$$

This looks quite messy. However, by factoring out constants and combining like terms, we get:

$$p(y_{n+1}|y_1, \dots, y_n) = \frac{2^2 y_{n+1}^{2a-1} (\sqrt{s + d^2})^{2(an+c)}}{\Gamma(a)\Gamma(an + c)} \int_{-\infty}^{\infty} \theta^{2a+2(an+c)-1} e^{-\theta^2(y_{n+1}^2 + d^2 + s)} d\theta$$

We recognize the integrand as the kernel of a Galenshore($an + a + c, \sqrt{y_{n+1}^2 + d^2 + s}$). If we multiply and divide by the proper constant for this distribution, the integral will integrate to 1, and we will just be left with our predictive posterior distribution. The constant we will multiply and divide by is:

$$\frac{2 (y_{n+1}^2 + d^2 + s)^{an+a+c}}{\Gamma(an + a + c)}$$

We will just be left with the reciprocal of the constant above (as the other constant will go away when we evaluate the integral, as it will then be a proper distribution) and thus we will have the reciprocal of the constant above multiplied by the other constants we had outside the integral. Thus our distribution is:

$$p(y_{n+1}|y_1, \dots, y_n) = \frac{2 y_{n+1}^{2a-1} (s + d^2)^{2(an+c)} \Gamma(an + a + c)}{(y_{n+1}^2 + d^2 + s)^{an+a+c} \Gamma(a)\Gamma(an + c)}$$