Homework 2

Rebecca C. Steorts

Problem 1 (Solution to Lab 2). We repeat Tasks 1–2, so the problem is fully reproducible.

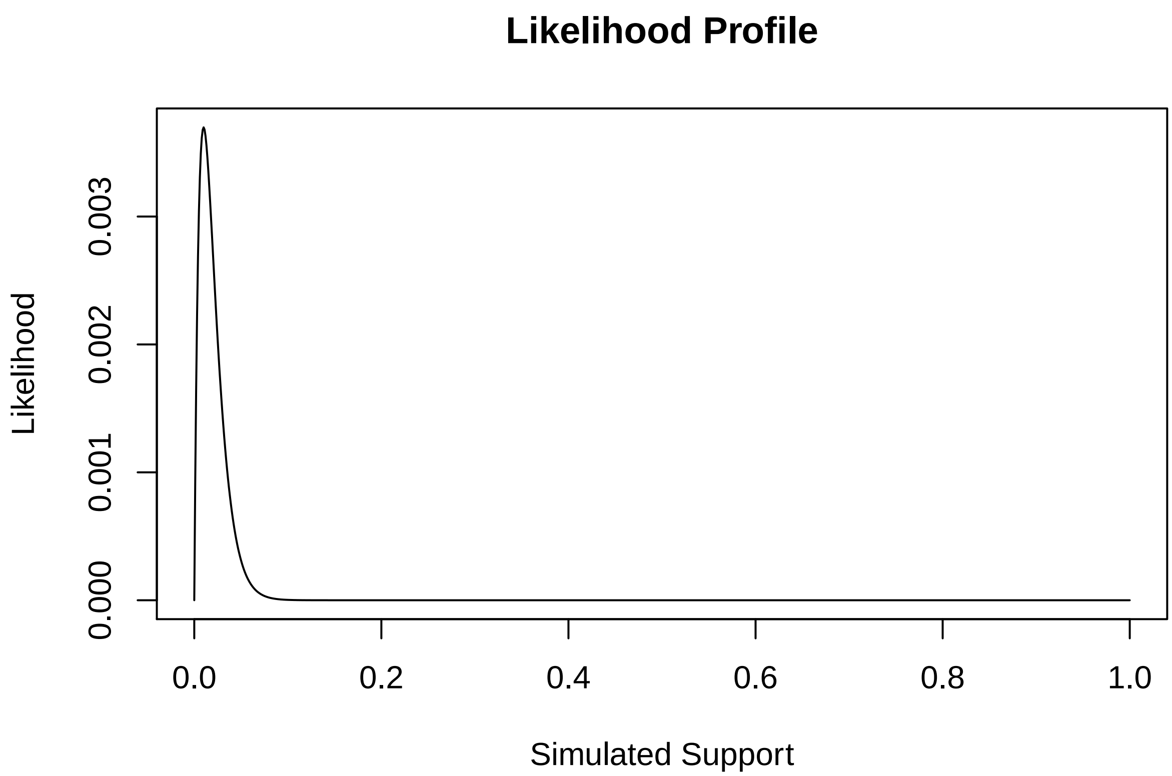
The data can be simulated as follows:

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

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.

```r
### Bernoulli LH Function ###
# Input - the data, theta grid #
# Produces likelihood values #
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 #
sim.LH <- myBernLH(obs.data = obs.data, theta = theta.sim)
# Create the Plot #
plot(theta.sim, sim.LH, type = 'l',
 main = 'Likelihood Profile', xlab = 'Simulated Support',
  ylab = 'Likelihood')
```



Task 3

The function used to generate the paraters is given as follows. The parameters themselves are printed out and stores as well for later use.

```
### Function to determine posterior parameters based on ###
### observed data and prior assumptions ###
# Inputs - Prior Paramters, observed data #
myPosteriorParam <- function(pri.a, pri.b, obs.data){
    N <- length(obs.data)
    x <- sum(obs.data)
    post.a <- pri.a + x
    post.b <- pri.b + N - x
    post.param <- list('post.a' = post.a,
     'post.b' = post.b)
    return(post.param)
}

# Find posterior parameters for two different priors #
# a = 1, b = 1 #
non.inform <- myPosteriorParam(pri.a = 1, pri.b = 1, obs.data = obs.data)
# a = 3, b = 1 #
inform <- myPosteriorParam(pri.a = 3, pri.b = 1, obs.data = obs.data)
print(non.inform)
```

```
## $post.a
## [1] 2
##
## $post.b
## [1] 100
```

```r
print(inform)
```

```
## $post.a
## [1] 4
##
## $post.b
## [1] 100
```

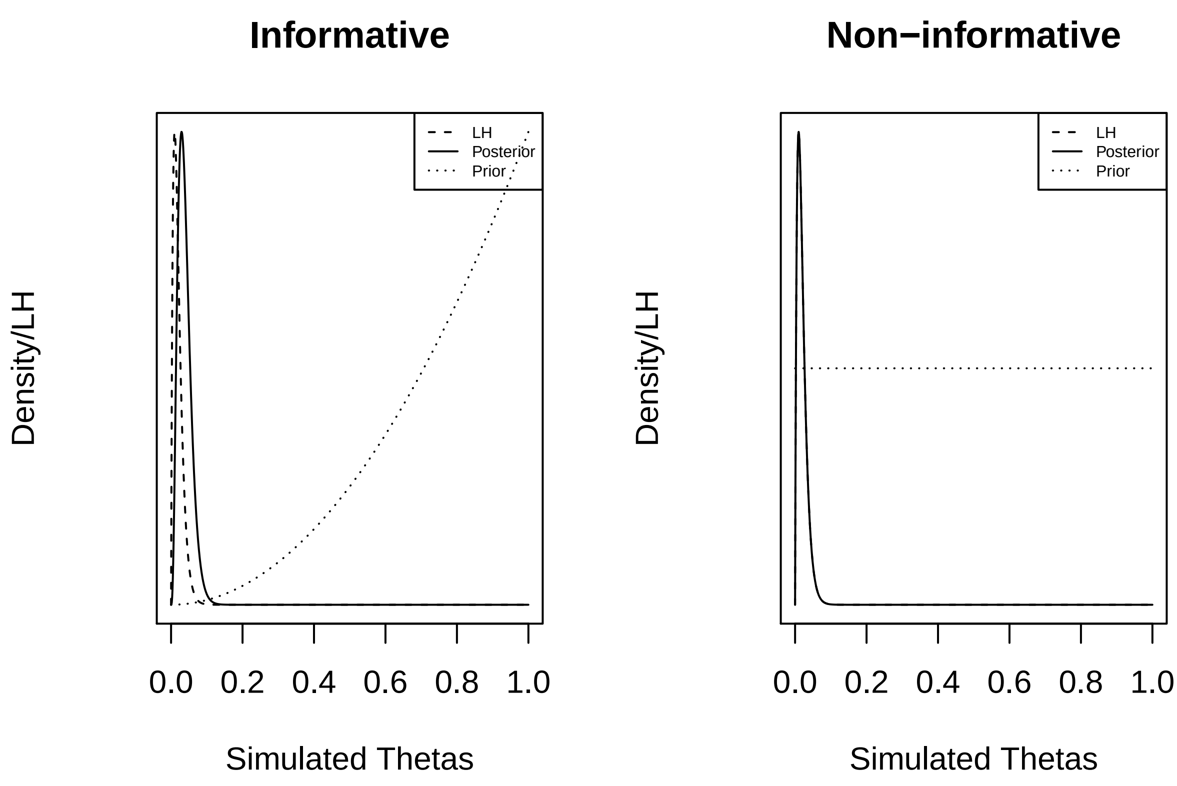
Task 4 and 5

The desired plots are given below, along with the code used to generate them.

```r
### Create a plot of LH, Pri, Posterior using the simulated seq ###
non.inform.den <- dbeta(x = theta.sim, shape1 = non.inform$post.a,
 shape2 = non.inform$post.b)
inform.den <- dbeta(x = theta.sim, shape1 = inform$post.a,
 shape2 = inform$post.b)
pri.inform <- dbeta(x = theta.sim, shape1 = 3,
 shape2 = 1)
pri.non.inform <- dbeta(x = theta.sim, shape1 = 1,
 shape2 = 1)


par(mfrow=c(1, 2))
plot(theta.sim, sim.LH, lty = 2, xlab = 'Simulated Thetas',
 ylab = 'Density/LH', type = 'l', yaxt = 'n', main = 'Informative')
par(new = TRUE)
plot(theta.sim, inform.den, lty = 1, axes = FALSE, xlab = '', ylab = '',
 type = 'l')
par(new = TRUE)
plot(theta.sim, pri.inform, lty = 3, axes = FALSE, xlab = '', ylab = '',
 type = 'l')
legend('topright', lty=c(2,1,3), legend = c('LH', 'Posterior', 'Prior'),
cex = 0.5)

plot(theta.sim, sim.LH, lty = 2, xlab = 'Simulated Thetas',
 ylab = 'Density/LH', type = 'l', yaxt = 'n', main = 'Non-informative')
par(new = TRUE)
plot(theta.sim, non.inform.den, lty = 1, axes = FALSE, xlab = '', ylab = '',
 type = 'l')
par(new = TRUE)
plot(theta.sim, pri.non.inform, lty = 3, axes = FALSE, xlab = '', ylab = '',
 type = 'l')
legend('topright', lty=c(2,1,3), legend = c('LH', 'Posterior', 'Prior'),
cex=0.5)
```



In the plots given above, the first think to note is Shrinkage. The posterior distribution is averaged between the prior and the likelihood. As a result, when we used the flat prior, the influence of the likelihood is much, much greater than in the case where we use an informative prior.

Problem 2. Suppose you have data $x_1, \ldots, x_n$ which you are modeling as i.i.d. observations from an Exponential distribution, and suppose that your prior is $\boldsymbol{\theta} \sim \text{Gamma}(a, b)$, 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. Derive the formula for the posterior density, $p(\theta|x_{1:n})$.

$$\begin{aligned} p(\theta|x_{1:n}) &= \prod_{i=1}^{n}[\theta exp(-\theta x)] \times \frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta) \\ &\propto \theta^n exp(-\theta\sum_i x_i)\theta^{a-1}\exp(-b\theta) \\ &\propto \theta^{n+a-1}\exp\{-\theta(b+\sum_i x_i)\} \end{aligned}$$

It immediately follows that $\theta|x_{1:n}$ is Gamma$(n+a, b+\sum_i x_i)$.

b. The posterior mentioned above is a proper distribution since we restrict $n+a > 0$ and $b+\sum_i x_i$, which insures that the Gamma is a proper probability distribution.

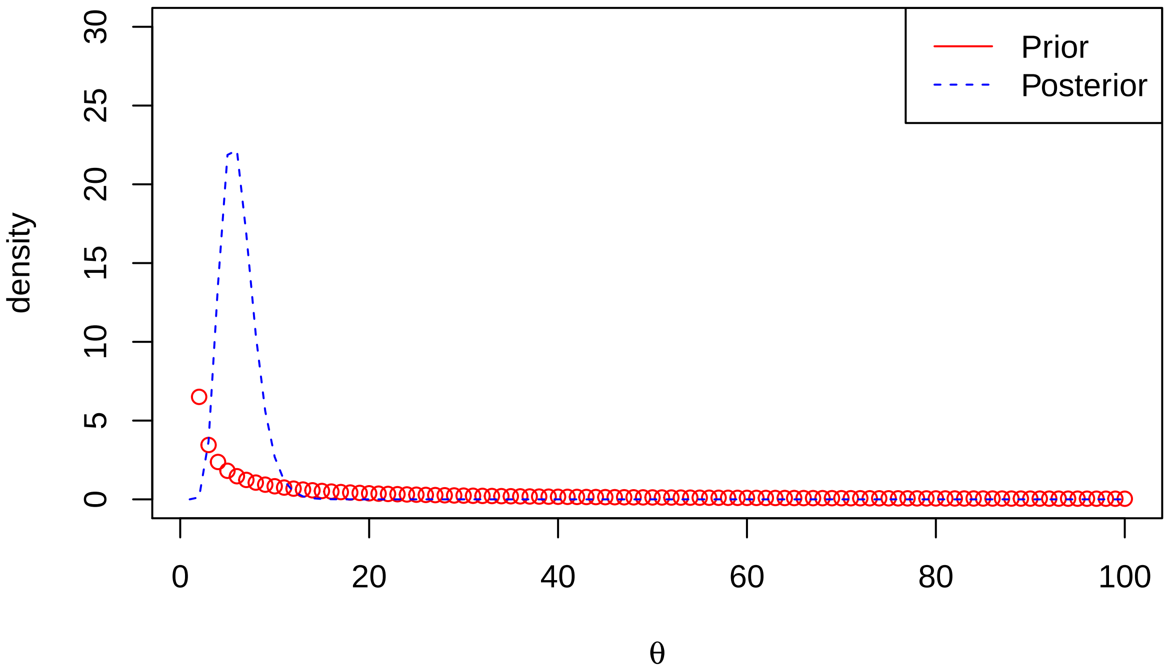
c.

```
# input data
data <- c(20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0)
# sequence of theta values
theta <- seq(0, 1, length.out=100)
x <- sum(data)
n <- length(data)
```

Now we generate the likelihood for each θ value and calculate the prior and posterior distributions.

```
# calculate prior and posterior distributions
prior <- dgamma(theta, shape = 0.1, rate = 1.0)
posterior <- dgamma(theta, shape = 0.1 + n, rate = 1.0 + x)

plot(prior, xlab=expression(theta), ylab="density", col="red",lty=1, ylim=c(0,30))
lines(posterior, col="blue",lty=2)
legend('topright', c("Prior", "Posterior"), lty=c(1,2), col=c("red","blue"))
```



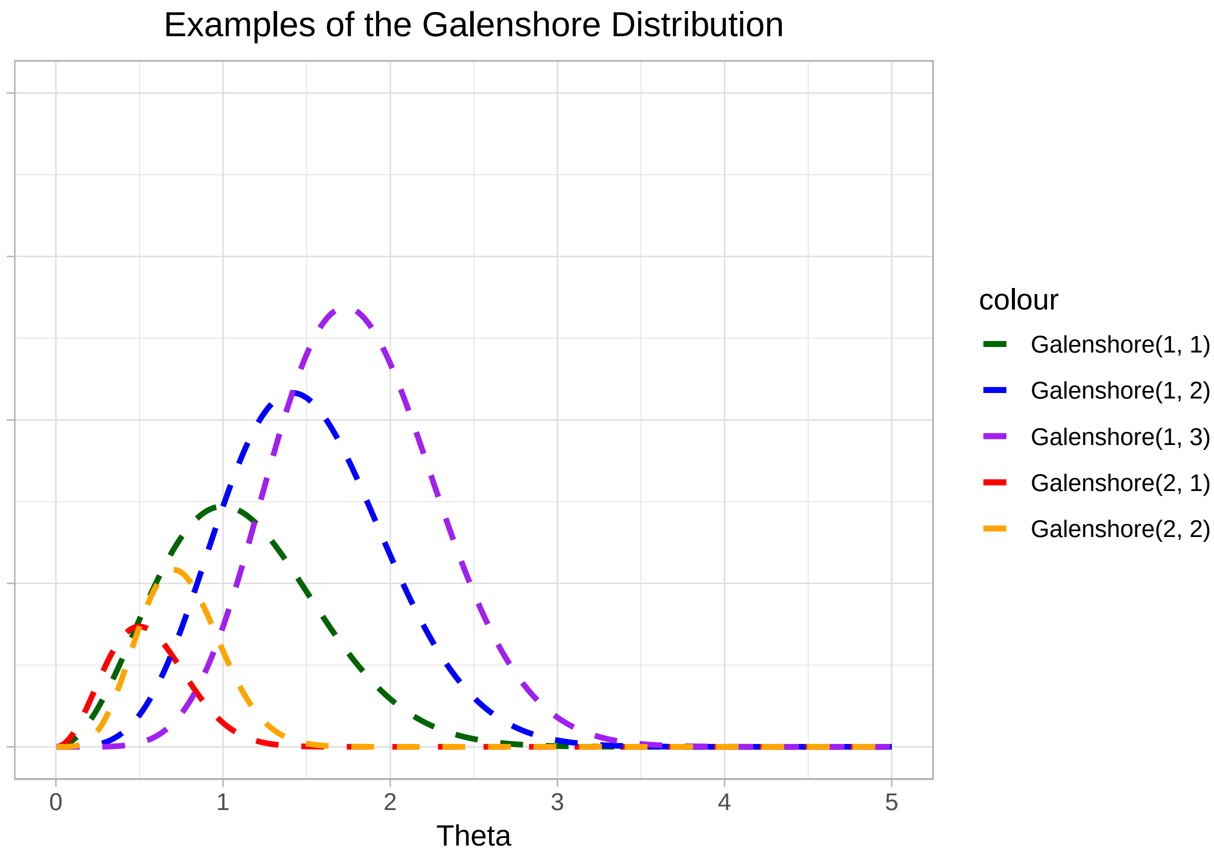
d. There are many examples where this model should be realistic in practice. Examples include:

- The time it takes before one's next telephone call.
- The time it takes for a bank teller to help a customer (assuming wait times are constant).
- The time until a radioactive particle decays.

One example of where this model should not be realistic in practice is any situation where the assumption of a constant rate is not satisfied. Often the rate of incoming phone calls differs according to the **time of day**. In such situations, the model would not be realistic here.

Problem 3.

a. The family of Galenshore distributions itself is conjugate. The only conjugate prior for a Galenshore distribution is another Galenshore distribution


Examples of the Galenshore Distribution
0
1
2
3
4
5
Theta
colour
Galenshore(1, 1)
Galenshore(1, 2)
Galenshore(1, 3)
Galenshore(2, 1)
Galenshore(2, 2)

b. Let's now consider plotting a few Galenshore densities.

3b

$$p(y_{1:n}|\theta) = \left(\frac{2}{\Gamma(a)}\right)^n \theta^{2an} \prod_{i=1}^{n} y_i^{2a-1} e^{-\theta^2 \sum y_i^2}$$

Let our prior $p(\theta) \sim Galenshore(c, d) \qquad (c, d > 0)$

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

$$= \left(\frac{2}{\Gamma(a)}\right)^n \theta^{2an} \left(\prod_{i=1}^{n} y_i^{2a-1}\right) e^{-\theta^2 \sum y_i^2} * \frac{2}{\Gamma(c)} d^{2c} \theta^{2c-1} e^{-\theta^2 d^2}$$

$$\propto \theta^{2(an+c)-1} e^{-(d^2 + \sum y_i^2)\theta^2}$$

$$\sim Galenshore(an + c, \sqrt{d^2 + \sum y_i^2})$$

c.

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

$$= \frac{\theta_a^{2an} e^{-\theta_a^2 \sum y_i^2} * \theta_a^{2c-1} e^{-\theta_a^2 d^2}}{\theta_b^{2an} e^{-\theta_b^2 \sum y_i^2} * \theta_b^{2c-1} e^{-\theta_b^2 d^2}}$$

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

Therefore, $\sum y_i^2$ is a sufficient statistic to calculate the ratio above.

d.

This is just the mean of the posterior distribution. Since we know the posterior follows a Galenshore distribution, we know from the beginning of this problem that:

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

e. We now compute the posterior predictive density. Recall that

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

We assume since $y_1, \cdots, y_n$ are i.i.d. that y_{n+1} is also i.i.d $\Rightarrow p(y_{n+1}|\theta, y_{1:n}) = p(y_{n+1}|\theta)$

$$\Rightarrow p(y_{n+1}|y_{1:n})$$

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

$$= \int_\theta \frac{2}{\Gamma(a)}\theta^{2a}y_{n+1}^{2a-1}e^{-\theta^2 y_{n+1}^2} * \frac{2}{\Gamma(an+c)}\left(\sqrt{d^2+\sum y_i}\right)^{2(an+c)}\theta^{2(an+c)-1}e^{-(\sqrt{d^2+\sum y_i})^2\theta^2}\,d\theta$$

$$= \frac{2y_{n+1}^{2a-1}}{\Gamma(a)}\frac{2(d^2+\sum y_i^2)^{an+c}}{\Gamma(an+c)}\int_\theta \theta^{2(an+a+c)-1}e^{-(d^2+\sum y_i^2+y_{n+1}^2)\theta^2}\,d\theta$$

$$= \frac{2y_{n+1}^{2a-1}}{\Gamma(a)}\frac{2(d^2+\sum y_i^2)^{an+c}}{\Gamma(an+c)}$$
$$* \int_\theta \theta^{2(an+a+c)-1}e^{-(d^2+\sum y_i^2+y_{n+1}^2)\theta^2} * \frac{(d^2+\sum y_i^2+y_{n+1}^2)^{(an+a+c)}}{(d^2+\sum y_i^2+y_{n+1}^2)^{(an+a+c)}} * \frac{2}{\Gamma(an+a+c)}\frac{\Gamma(an+a+c)}{2}\,d\theta$$

$$= \frac{2y_{n+1}^{2a-1}}{\Gamma(a)}\frac{2(d^2+\sum y_i^2)^{an+c}}{\Gamma(an+c)}\frac{\Gamma(an+a+c)}{2}\frac{1}{(d^2+\sum y_i^2+y_{n+1}^2)^{(an+a+c)}}$$
$$* \int_\theta \frac{2}{\Gamma(an+a+c)}\theta^{2(an+a+c)-1}e^{-(d^2+\sum y_i^2+y_{n+1}^2)\theta^2} * (d^2+\sum y_i^2+y_{n+1}^2)^{(an+a+c)}\,d\theta$$

The term inside the integral is $\sim Galenshore\left(an+a+c, \sqrt{d^2+\sum y_i^2+y_{n+1}^2}\right)$ and therefore now integrates to 1.

$$\Rightarrow p(y_{n+1}|y1:n) = \frac{2y_{n+1}^{2a-1}}{\Gamma(a)}\frac{2(d^2+\sum y_i^2)^{an+c}}{\Gamma(an+c)}\frac{\Gamma(an+a+c)}{2}\frac{1}{(d^2+\sum y_i^2+y_{n+1}^2)^{(an+a+c)}}$$

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