Homework #2 STA 360

Jerry Xin STA 360: Homework 2

Due Friday September 3rd, 5 PM EDT

```r
library(tidyverse)
library(knitr)
library(broom)
```

1. Lab component (30 points total) Please refer to lab 2 and complete tasks 3—5.

(a) (10) Task 3

(b) (10) Task 4

(c) (10) Task 5

```r
# 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
```

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

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

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

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

1(a) 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.

```r
### Bernoulli LH Function ###

# Input: obs.data, theta
# Output: bernoulli likelihood

myBernLH <- function(obs.data, theta){
```
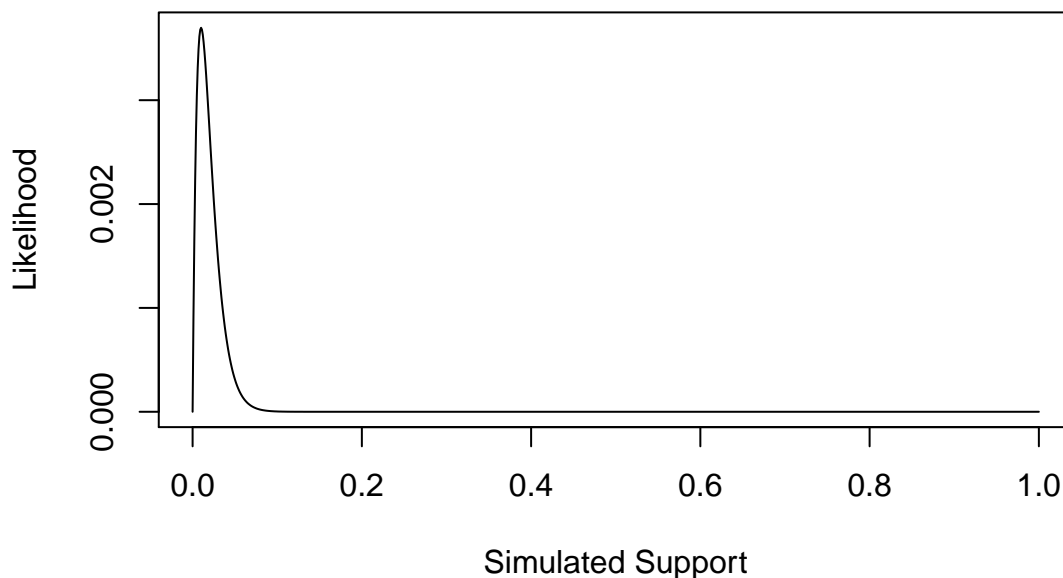
```
  N <- length(obs.data)
  x <- sum(obs.data)
  LH <- (theta^x) * ((1-theta)^(N-x))
  return(LH)
}


### Plot LH for a grid of theta values ###
# Create the grid #
# Store the LH values
# Create the Plot
theta.sim <- seq(from = 0, to = 1, length.out = 1000)
sim.LH <- myBernLH(obs.data, theta = theta.sim)
plot(theta.sim, sim.LH, type ="line", main = "Likelihood Profile", xlab = "Simulated Support", ylab = "
```

Likelihood Profile

Simulated Support

1(b) Task 4 (To be completed for homework)

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

```
posterior <- function(priorA, priorB, obs.data)
{
  N <- length(obs.data)
  x <- sum(obs.data)
  postA <- priorA + x
    postB <- priorB + N - x
    postparam <- list('postA' = postA,
      'postB' = postB)
    return(postparam)
}
```

```
nonInformative <- posterior(priorA = 1, priorB = 1, obs.data = obs.data)
informative <- posterior(priorA = 3, priorB = 1, obs.data = obs.data)
print(c(nonInformative, informative))
```

```
## $postA
## [1] 2
##
## $postB
## [1] 100
##
## $postA
## [1] 4
##
## $postB
## [1] 100
```

Non-Informative Posterior Parameters: PostA = 2, PostB = 100 Informative Posterior Parameters: PostA = 4 PostB = 100 # 1(c) Task 5 (To be completed for homework)

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.
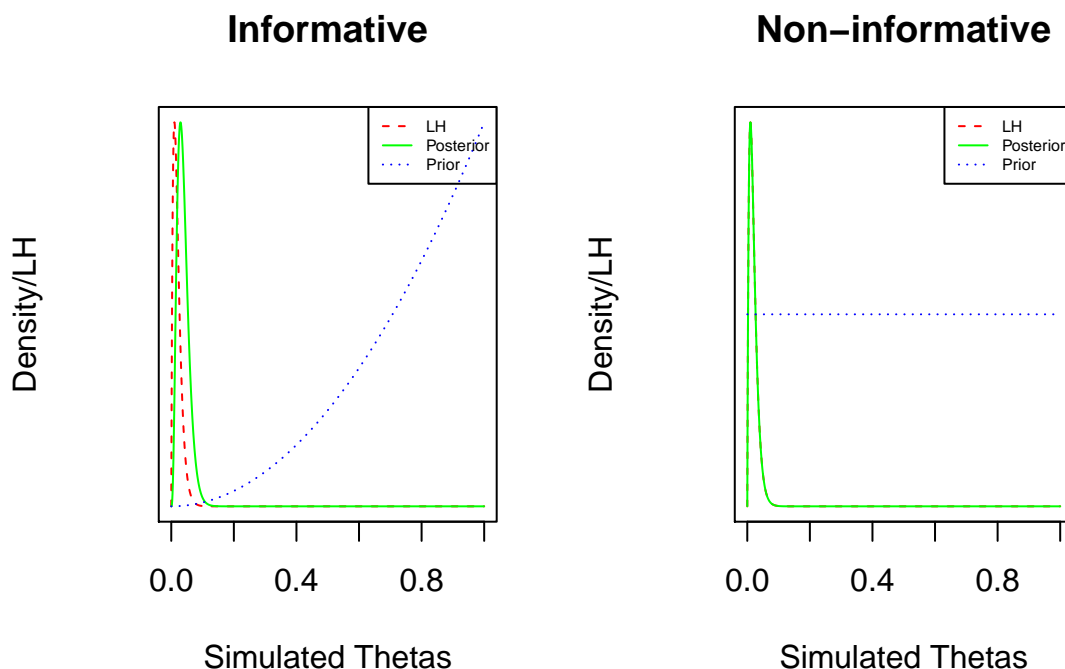
```
nonInformativeDen <- dbeta(x = theta.sim, shape1 = nonInformative$postA,
 shape2 =nonInformative$postB)
informativeDen <- dbeta(x = theta.sim, shape1 = informative$postA,
 shape2 = informative$postB)
priorInform <- dbeta(x = theta.sim, shape1 = 3,
 shape2 = 1)
priorNonInform <- dbeta(x = theta.sim, 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',
 , col = c("red"))
par(new = TRUE)
plot(theta.sim, informativeDen, lty = 1, axes = FALSE, xlab = '', ylab = '',
 type = 'l', , col = c("green"))
par(new = TRUE)
plot(theta.sim, priorInform, lty = 3, axes = FALSE, xlab = '', ylab = '',
 type = 'l', col = c("blue"))
legend('topright', lty=c(2,1,3), legend = c('LH', 'Posterior', 'Prior'),
       col = c("red", "green", "blue"), cex = 0.5)

plot(theta.sim, sim.LH, lty = 2, xlab = 'Simulated Thetas',
 ylab = 'Density/LH', type = 'l', yaxt = 'n', main = 'Non-informative',
 col = c("red"))
par(new = TRUE)
plot(theta.sim, nonInformativeDen, lty = 1, axes = FALSE, xlab = '', ylab = '',
 type = 'l', col = c("green"))
par(new = TRUE)
plot(theta.sim, priorNonInform, lty = 3, axes = FALSE, xlab = '', ylab = '',
 type = 'l', col = c("blue"))
legend('topright', lty=c(2,1,3), legend = c('LH', 'Posterior', 'Prior'),
```

Informative   **Non−informative**

These graphs both show the posterior distribution is a result of combining the Likelihood and the Prior. It is graphically seen as an average between the Likelihood and Prior. A non-informative prior is graphically seen as a flat line and seems to have less/ almost neglible effect on the posterior, as it seems to have no effect on the posterior graph. The effect is less when compared to a prior which is informative, which is graphically a upwards sloping curve, and is shown to have a larger effect on moving the posterior rightwards.

2. (20 points total) The Exponential-Gamma Model

We write $X Exp(\theta)$ to indicate that X has the Exponential distribution, that is, its p.d.f. is

$$p(x|\theta) = Exp(x|\theta) = \theta \exp(-\theta x)(x > 0).$$

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 (such as neuron spikes, website hits, neutrinos captured in a detector), extreme values such as maximum daily rainfall over a period of one year, or the amount of time until a product fails (lightbulbs are a standard example).

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 $\theta Ga(a, b)$, that is,

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

$$\theta^n \exp(-\theta \sum x_i) \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta)$$

Remove $\frac{b^a}{\Gamma(a)}$ because it doesn't have theta and can be treated as a constant to get:

$$\theta^n \exp(-\theta \sum x) \theta^{a-1} \exp(-b\theta)$$

Simplify to get:

$$\theta^{a+n-1} \exp(-\theta * (b + \sum x))$$

4

$$Ga(\theta|a+n, b+\sum x)$$

The posterior is a Gamma Distribution with Gamma(a + n, b + sum(x)).

(a) (5) 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 (Bernoulli, Beta, Exponential, or Gamma).

$$\theta^n \exp(-\theta \sum x_i) \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta)$$

Remove $\frac{b^a}{\Gamma(a)}$ because it doesn't have theta and can be treated as a constant to get:

$$\theta^n \exp(-\theta \sum x_i) \theta^{a-1} \exp(-b\theta)$$

Simplify to get:

$$\theta^{a+n-1} \exp(-\theta * (b + \sum x_i))$$

$$Ga(\theta|a+n, b+\sum x_i)$$

The posterior is a Gamma Distribution with Gamma(a + n, b + sum(x)).

(b) (5) Why is the posterior distribution a *proper* density or probability distribution function?

The posterior distribution is a proper density, or probability distribution function, because it integrates to 1, which means the area under the function adds up to 1.

(c) (5) Now, suppose you are measuring the number of seconds between lightning strikes during a storm, your prior is $Ga(0.1, 1.0)$, and your data is

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

Plot the prior and posterior p.d.f.s. (Be sure to make your plots on a scale that allows you to clearly see the important features.)

```
set.seed(123)
obs.data2 <- c(20.9, 69.7, 3.6, 21.8, 21.4, 0.4, 6.7, 10.0)
head(obs.data2)
```

```
## [1] 20.9 69.7  3.6 21.8 21.4  0.4
```

```
tail(obs.data2)
```

```
## [1]  3.6 21.8 21.4  0.4  6.7 10.0
```

```
length(obs.data2)
```

```
## [1] 8
```

```r
posterior2 <- function(priorA, priorB, obs.data)
{
  N <- length(obs.data2)
  x <- sum(obs.data2)
  postA <- priorA + N
    postB <- priorB + x
    postparam <- list('postA' = postA,
     'postB' = postB)
    return(postparam)
}
```

```r
informative1<- posterior2(priorA = 0.1, priorB = 1.0, obs.data = obs.data2)
print(c(informative))
```

```
## $postA
## [1] 4
##
## $postB
## [1] 100
```

```r
gamma.sim <- seq(from = 0, to = 1, length.out = 1000)
```

```r
informativeDen1 <- dgamma(x = gamma.sim, shape = informative1$postA,
 rate = informative1$postB)
priorInform1 <- dgamma(x = gamma.sim, shape = 0.1,
 rate = 1)

plot(gamma.sim, informativeDen1, lty = 1, xlab = expression(theta), ylab = 'Density',
 type = 'l', , col = c("green"))
par(new = TRUE)
plot(gamma.sim, priorInform1, lty = 3, axes = FALSE, xlab = '', ylab = '',
 type = 'l', col = c("blue"))
legend('topright', lty=c(2,1,3), legend = c( 'Posterior', 'Prior'),
       col = c("green", "blue"), cex = 0.5)
```
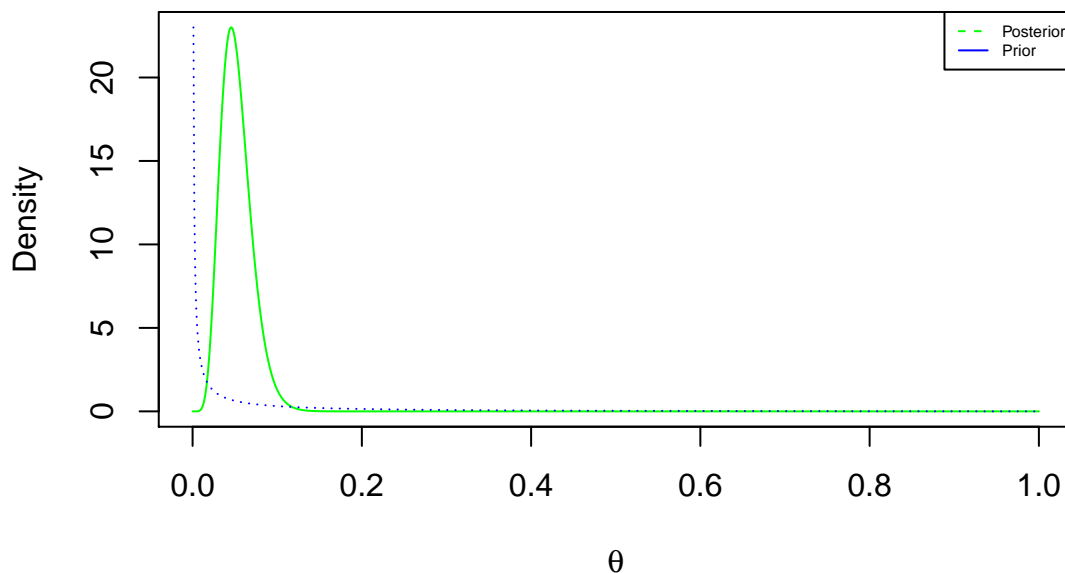
(d) (5) Give a specific example of an application where an Exponential model would be reasonable. Give an example where an Exponential model would NOT be appropriate, and explain why.

Exponential functions assuming a constant multiplicative rate of change, so a specific application would be studying a lightbulb, which has the memorylessness property of an exponential distribution there is also a constant multiplicative rate of change here, 2. An example where the exponential model would not be appropiate is modeling the risk of death, because it doesn't have the property of memorylessness, and there is not a constant multiplication rate of change here. The probability is relatively close to 0 for young people from 0-40 years old, but skyrockets when people grow older and will level out at around 1 for those who get older and older. This is a logistic regression model instead.

Constant Rate Assumption with the Exponential Model, time in between events. Assuming constant time in between events, memorylessness property.

3. (40 points total) { Priors, Posteriors, Predictive Distributions (Hoff, 3.9)}

An unknown quantity Y has a Galenshore(a, θ) distribution if its density is given by

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

for $y > 0, \theta > 0, a > 0$. Assume for now that a is known. For this density,

$$E[Y] = \frac{\Gamma(a + 1/2)}{\theta \Gamma(a)}$$

and

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

(a) (10) Identify a class of conjugate prior densities for θ. Assume the prior parameters are c and d. Plot a few members of this class of densities.

Let the prior also be a Galenshore($c, $d) distribution:

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

The likelihood a Galenshore(a, θ) distribution:

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

Because you have θ^{2a} , and $e^{-\theta^2 y^2}$, only a Galenshore distribution would work as a conjugate prior density.

```
dgalenshore <- function(theta, c, d) {
    # takes numeric vector (elements should be >0) and returns the galenshore
    # probability density at each point. Parameters a and theta should all be
    # > 0.
    return((2/gamma(c)) * theta^(2*c-1) * d^(2 * c) * exp(-d^2 * theta^2))
```

```
}
x <- seq(0.01, 3, 0.01)
plot(x, dgalenshore(x, 1, 1), type = "l", ylab = "Galenshore Density(x)", col = "black",
    ylim = c(0, 5), lwd = 2)
lines(x, dgalenshore(x, 1, 1.5), type = "l", col = "red", lwd = 2, lty = 2)
lines(x, dgalenshore(x, 3, 1), type = "l", col = "blue", lwd = 2, lty = 3)
legend("topright", c("Galenshore(1,1)", "Galenshore(1,2)", "Galenshore(3,1)"),
    col = c("black", "red", "blue"), lwd = c(2, 2, 2), lty = c(1, 2, 3))
```
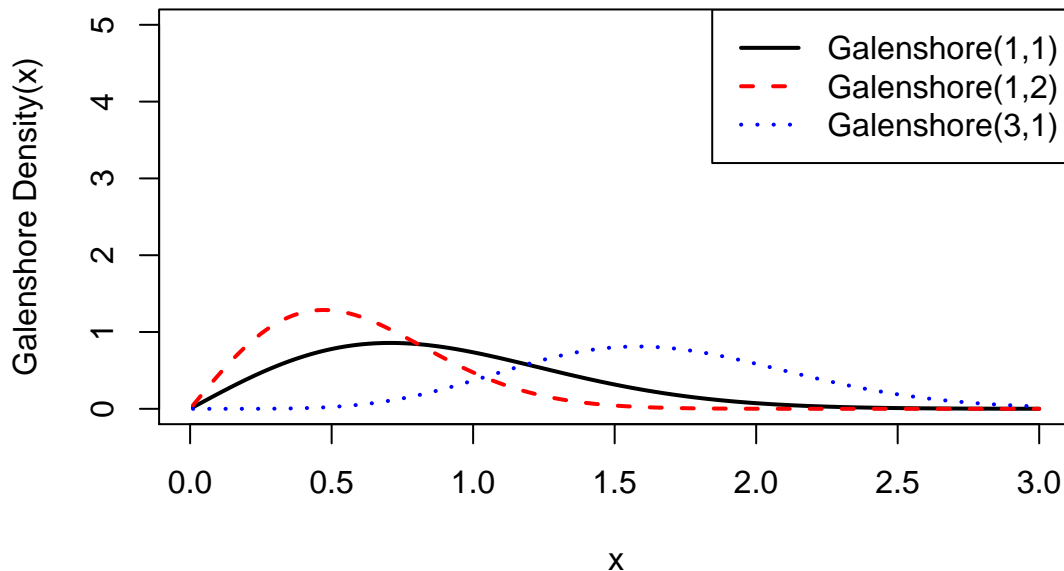

(b) (5) Let $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Galenshore(a, θ). Find the posterior distribution of $\theta \mid y_{1:n}$ using a prior from your conjugate class.

You have a sample of ys, $y_{1:n}$ See attached document.

(c) (10) Write down

$$\frac{p(\theta_a \mid y_{1:n})}{p(\theta_b \mid y_{1:n})}$$

and simplify. Identify a sufficient statistic.

Compared the ratio of 2 different likelihood distributions, one has parameter a and one has parameter b. Normalizing constants will cancel out, some simplification. See Attached document.

(d) (5) Determine

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

.

See Attached Document

(e) (10) Show that the form of the posterior predictive density

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

See Attached Document.