Homework 2

STA 360, Fall 2020

Due Friday August 28, 5 PM EDT

In class, you saw the Binomial-Beta model. We will now use this to solve a very real problem! Suppose I wish to determine whether the probability that a worker will fake an illness is truly 1%. Your task is to assist me! Tasks 1–2 will be completed in lab and tasks 3–5 should be completed in your weekly homework assignment.

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 $\theta \mid X$?

$$\begin{aligned}
p(\theta \mid X) &\propto p(X \mid \theta)p(\theta) && (1)\\
&\propto \theta^x (1-\theta)^{(n-x)} \times \theta^{(a-1)}(1-\theta)^{(b-1)} && (2)\\
&\propto \theta^{x+a-1}(1-\theta)^{(n-x+b-1)}. && (3)
\end{aligned}$$

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:

```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
```

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


### Plot LH for a grid of theta values ###
# Create the grid #
# Store the LH values
# Create the Plot
```

Task 4

Task 5

Exponential Gamma Model

a.

$$p(x_{1:n}|\theta) = \theta^n e^{-n\theta \sum_{i=1}^n x_i} I(x > 0)$$

$$p(\theta|x_{1:n}) = \frac{p(x_{1:n}|\theta)p(\theta)}{p(x)}$$

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

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

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

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