Lab 2 – Beta-Binomial Distribution

Rebecca C. Steorts

January 2018

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 1–3 will be completed in lab and tasks 3–5 should be completed in your weekly homework assignment. You should still upload task 3 even though this will be worked through in lab!

Task 1

Let's derive the Beta-Binomial distribution.

Assume that

$$X \mid \theta \sim \text{Binomial}(n, \theta),$$
$$\theta \sim \text{Beta}(a, b),$$

where $a, b > 0$ are assumed to be fixed, 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:

```
# 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 (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 such as (a,b) = (1,1) and for an informative case (a,b) = (3,1)}.

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 off using the command yaxt="none" in the plot() command. Why? Superimposing the distributions may make the scale non-sense.

Solution to Task 3

The likelihood function is given below. Since this is a probability and is only valid over the interval from $[0, 1]$, I will generate a sequence over that interval of length 1000.

```r
### Bernoulli LH Function ###
# Input - the data, theta grid #
# Produces likelihood values #
likelihood_function <- function(obs_data, theta) {
  n <- length(obs_data)
  x <- sum(obs_data)
  return(((theta)^x) * ((1 - theta)^(n - x)))
}

### 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 #
likelihood_sim <- likelihood_function(obs_data = obs_data, theta = theta_sim)
# Create the Plot #
plot(theta_sim, likelihood_sim, type = "l",
 main = "Likelihood Profile", xlab = "Simulated Support",
  ylab = "Likelihood")
```