20180130 - Bayes Binomial

Evan L. Ray

Janary 30, 2018

Introduction

Let's estimate the proportion of M&M's that are blue. Call this proportion θ. Suppose we take a sample of n M&M's and let the random variable X denote a count of how many are blue in that sample. Our model is $X \sim \text{Binomial}(n, \theta)$.

We have developed two approaches to inference for θ:

1. The maximum likelihood estimate $\hat{\theta}_{MLE} = \frac{x}{n}$, which maximizes the likelihood function $\mathcal{L}(\theta|x)$.
2. A Bayesian approach with conjugate prior distribution given by $\Theta \sim Beta(a, b)$. The posterior distribution is given by $\Theta|n, x \sim Beta(a + x, b + n - x)$. One option for a point estimate based on this distribution is the posterior mean: $\hat{\theta}_{Bayes} = \frac{a+x}{}$

We have three related goals in this lab:

1. To see what effect the prior distribution has on Bayesian inferences, and how this changes as the sample size n increases.
2. To see what the posterior distribution looks like in Bayesian inference, and how this changes as the sample size n increases.
3. To compare maximum likelihood and Bayesian estimates of θ, and see how these estimates change as the sample size n increases.

Procedure

Step 1. Take some samples and record data

In order to compare the estimates above, let's take samples of a few different sizes and plot the likelihood function and point estimates from each method for each sample size.

Take a sample of about 50 M&M's – the exact number is not important.

Record the following:

Was your first M&M blue? (Pick a random M&M from your sample that you will count as your first draw)

Out of your first 10 M&M's, how many were blue?

Out of your first 20 M&M's, how many were blue?

Out of all of the M&M's you picked, how many were blue? Also, how big was your total sample size?

Step 2. Put your prior parameters and data in to R and explore the results

Open up the Jupyter notebook file for this lab that you have cloned onto Gryd from GitHub. You will see code like this:

```
################################################################################
## You will modify prior parameter specifications and observed data values in here
beta_prior_params <- data.frame(
    prior_a = c(1, 2), # add a new value of a by appending to this vector
    prior_b = c(1, 10) # add a new value of b by appending to this vector
)

observed_data <- data.frame(
    x = c(0), # add new values of x by appending to this vector
    n = c(0) # add new values of n by appending to this vector
)
## You don't need to modify anything below this point
################################################################################
```

Add your prior parameters a and b for Θ to the first couple of lines. For example, if your prior was Beta$(3, 4)$ the first lines will look like this:

```
beta_prior_params <- data.frame(
    prior_a = c(1, 2, 3), # add a new value of a by appending to this vector
    prior_b = c(1, 10, 4) # add a new value of b by appending to this vector
)
```

Similarly, add each x and n value to the observed data. If your first M&M was blue, and in the first 10 you had a total of 3 blue M&M's your next lines would look like this (but also add the values for larger sample sizes)

```
observed_data <- data.frame(
    x = c(0), # add new values of x by appending to this vector
    n = c(0) # add new values of n by appending to this vector
)
```

Then, run the code in that R cell, scroll to the bottom and answer the questions.

To submit the lab, save the notebook file, then use git to add/commit/push the file to GitHub.