

Lab 2

Parker DeBruyne - V00837207

May 12, 2023

Likelihood methods for the Binomial distribution

- Put your name in the author section above.
- Execute each chunk of code to ensure that your code works properly. The TA will review the code with you.
- Sometimes one of your chunks of code will not compile properly, but you must hand your document in. In that case, 'Comment' out the R code that is not working properly using `#` as the first character in your lines of code. Do that now for the chunk of code in Section number 4, below. Put `#` at the start of each line of R code.
- Complete question 6 below. Execute your question 6 code chunk to ensure that it works properly. If you cannot compile it without errors before the end of the class, comment out the chunk.
- Save this .Rmd file to the (Windows) C: drive so that there are no knitting issues and then knit the entire document to pdf.
- Copy your pdf and Rmd files to your personal directory (see the Lab 1 instructions if needed).
- Upload the pdf file to the Lab 2 Practice Assignment Activity in the Lab section of BrightSpaces before Thursday 11:59pm.

1. First we generate 1 observation from the Binomial(100, theta=.07) distribution

```
#This is a comment inside of a chunk of R code.

set.seed(54321)    #setting the random number generator seed
                   #SHOULD always be done before simulating.
# Ask your Lab instructor to explain set.seed()

bdat <- rbinom(1, size=100, prob=.07 )    # see ?rbinom for information
bdat

## [1] 6
```

2. We know that the MLE for the probability of success, theta, is the sample proportion.

```
thetahat <- bdat/100
thetahat

## [1] 0.06
```

The MLE of theta is 0.06.

3. Compute the vector of Log-Likelihoods for a sequence of theta values given the one data point bdat.

```
#create a vector of values from 0.02 to 0.14 in steps of 0.005
theta <- seq(0.02, 0.14, by=.005)

#compute the log of binomial probabilities
bloglike <- dbinom(bdat, size=100, prob=theta, log=TRUE)

#combine the theta values and the log probabilities as column vectors
head(cbind(theta, bloglike))

##      theta  bloglike
## [1,] 0.020 -4.472250
## [2,] 0.025 -3.614208
## [3,] 0.030 -3.003571
## [4,] 0.035 -2.564456
## [5,] 0.040 -2.251580
## [6,] 0.045 -2.035745
```

4. Plot the Log-Likelihood function versus theta.

There are several ways to call the *plot* function in R. Here are two ways to plot **Y versus X**.

1. `plot(x, y)`
2. `plot(y ~ x)` Where y is plotted on the vertical axis and x is plotted on horizontal axis.

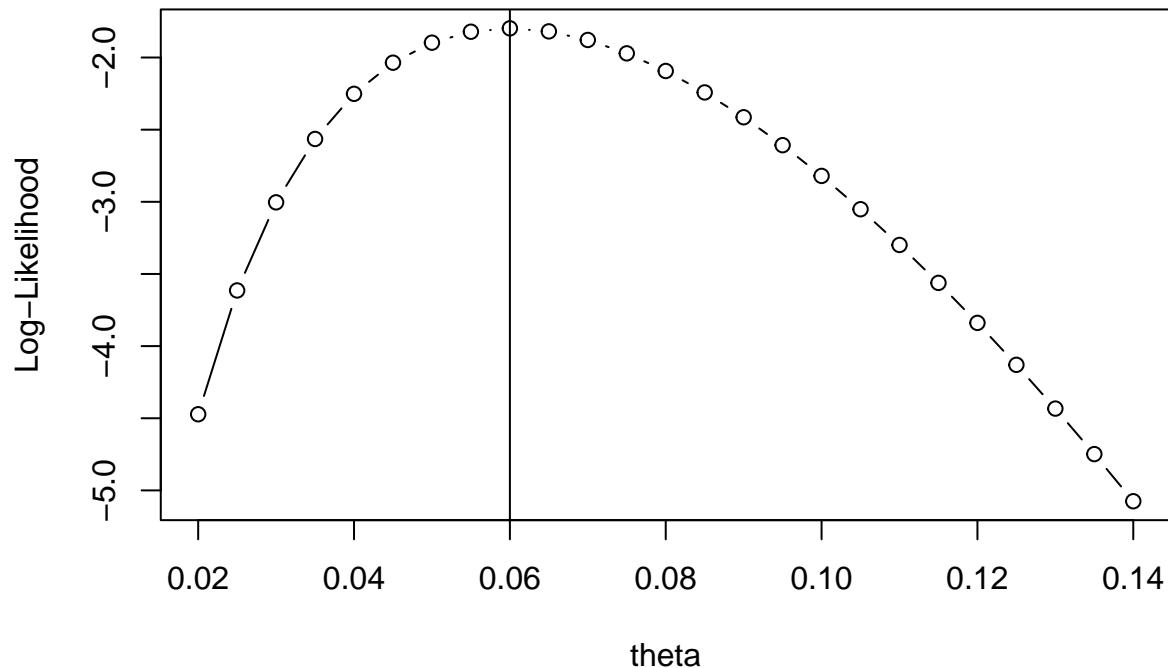
```
#plot(bloglike ~ theta, ylab='Log-Likelihood', xlab='theta', type='b')
#title('Binomial Log-likelihood for Lab 2')

#optional
#abline(v=0.06)
```

```
plot(bloglike ~ theta, ylab='Log-Likelihood', xlab='theta', type='b')
title('Binomial Log-likelihood for Lab 2')

#optional
abline(v=0.06) #showing the maximum point with a verticle line
```

Binomial Log-likelihood for Lab 2



5. Generate $n=10$ observations from the same distribution and plot the log-likelihood function given the n observations.

```
set.seed(54321)
n <- 10
bdat2 <- rbinom(n, size=100, prob=.07) # this generates n binomial observations
bdat2
```

```
## [1] 6 7 5 5 5 10 3 5 6 6
```

```
theta <- seq(0.02, 0.14, by=.005)
```

```
bloglike2 <- 0
for (i in 1:n){
  bloglike2 <- bloglike2 + dbinom(bdat2[i], size=100, prob=theta, log=TRUE)
  #computes the sum of the log of binomial probabilities over the observations
}
```

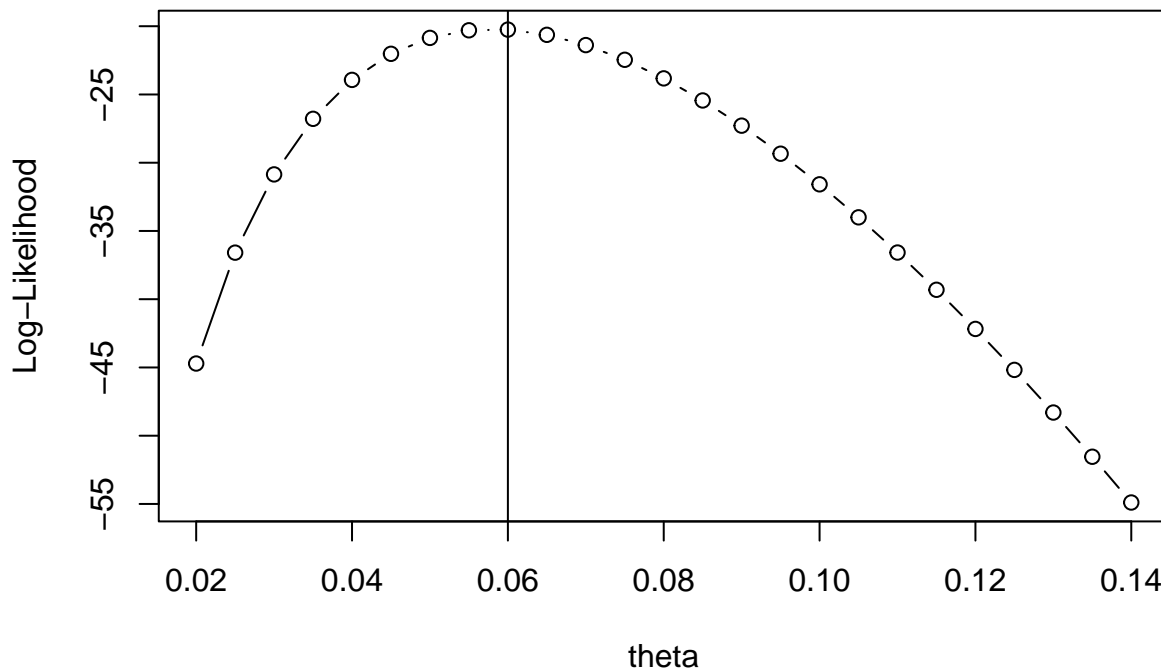
```
head(cbind(theta, bloglike2))
```

```
##      theta bloglike2
## [1,] 0.020 -44.71201
## [2,] 0.025 -36.58811
## [3,] 0.030 -30.85666
## [4,] 0.035 -26.78414
## [5,] 0.040 -23.93284
## [6,] 0.045 -22.02050
```

```
plot(bloglike2 ~ theta, ylab='Log-Likelihood', xlab='theta', type='b')
title(paste('Binomial Log-likelihood for Lab 2, n=', n, sep=''))
```

```
abline(v=0.06)
```

Binomial Log-likelihood for Lab 2, n=10



6. Now it is your turn. In the R code chunk below, redo question 5 for $n=100$ observations from the same distribution. How does the log-likelihood function compare with that in question 5? Write your comments in the section below the R chunk.

```
set.seed(54321)
n <- 100
bdat3 <- rbinom(n, size=100, prob=.07)
bdat3

##      [1]  6  7  5  5  5 10  3  5  6  6  4  8  3  8 14  8 11  7  7  7  5  3 10  5 11
##     [26]  7  8  9  2 12  8  4 12  4 10  6 11  9 10  7 10  5  8  6 11  7  8 10  4 11
##     [51]  7  9  7  4  7  7  5  6  4  6  6 10  6  7  2  7  4  6  5  8  7  8  7 10  8
##     [76]  3  9  4  4  4  5 11  9  7  5 12  8  8  9  2  7  8  5 10  9  7  8  7  8  7

theta <- seq(0.02, 0.14, by=.005)

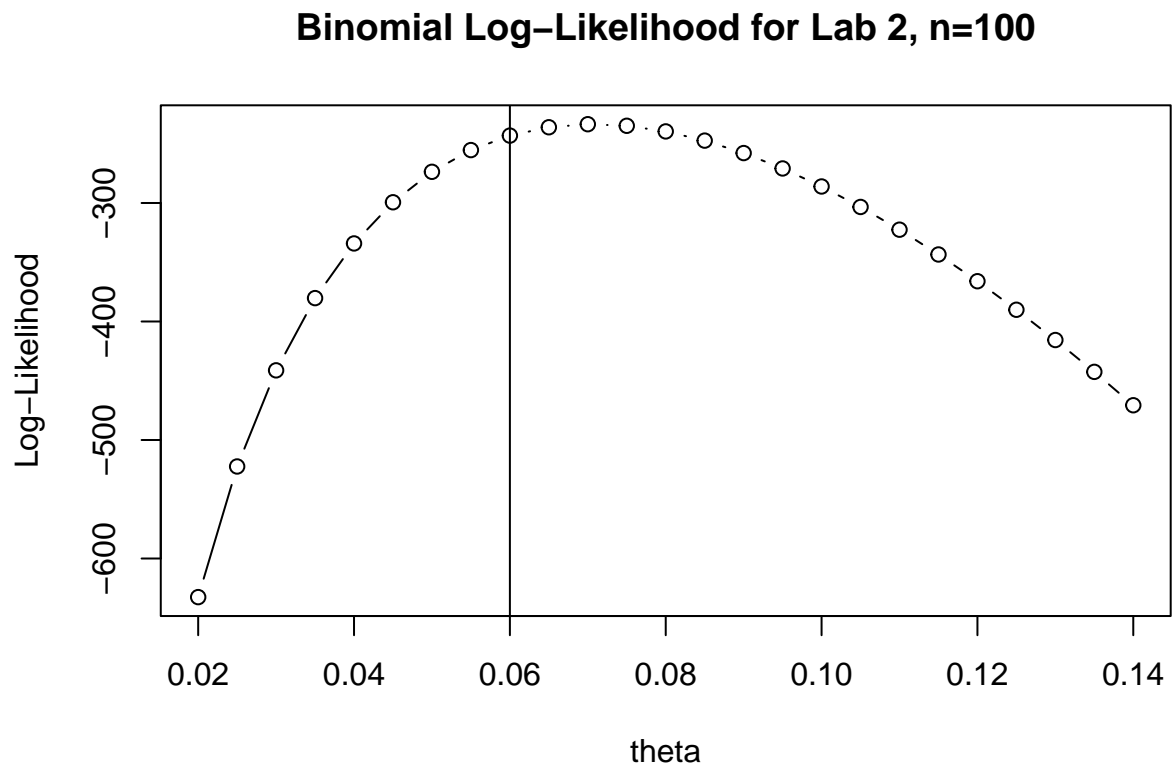
bloglike3 <- 0
for (i in 1:n){
  bloglike3 <- bloglike3 + dbinom(bdat3[i], size=100, prob=theta, log=TRUE)
  #computes the sum of the log of binomial probabilities over the observations
}

head(cbind(theta, bloglike3))

##      theta bloglike3
## [1,] 0.020 -632.5735
```

```
## [2,] 0.025 -522.3457
## [3,] 0.030 -441.2234
## [4,] 0.035 -380.2648
## [5,] 0.040 -334.1335
## [6,] 0.045 -299.3884
```

```
plot(bloglike3 ~ theta, ylab='Log-Likelihood', xlab='theta', type='b')
title(paste('Binomial Log-Likelihood for Lab 2, n=', n, sep=''))
abline(v=0.06)
```



Comments:

The peak has shifted from 0.06. When you are selecting a size more than 40, your distribution is normal; because we are taking more estimates making it more precise.