Qbs-assignment 2

梁文宣 R05724080

10/4/2017

Question 1

Import data

```
birth1 <- c(1,0,0,0,1,1,0,1,0,1,0,0,1,1,0,1,1,0,0,0,1,0,0,0,1,0, 0,0,0,1,1,1,
0,1,0,1,1,1,0,1,0 ,1,1,0 ,1,0,0 ,1,1,0 ,1,0,0 ,0,0 ,0,0,0 , 1,1,0,1,0,0,1,0,0,
0,1,0,0,1,1 ,1,1,0 ,1,0,1 ,1,1,1 ,1,0,0 ,1,0 ,1,1,0 , 1,0,1,1,1,0,1,1,1,1)
birth2 <- c(0,1,0,1,0,1,1,1,0,0,1,1,1,1,1,0,0,1,1,1,0,0,1,1,1,0, 1,1,1,0,1,1,
1,0,1,0,0,1,1,1,1 ,0,0,1 ,0,1,1 ,1,1,1 ,1,1,1 ,1,1 ,1,1,1 , 1,1,1,0,1,1,0,1,1,
0,1,1,1,0,0 ,0,0,0 ,0,1,0 ,0,0,1 ,1,0,0 ,1,0 ,0,1,1 , 0,0,0,1,1,1,0,0,0,0)
```

Define the grid

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

Define the prior

```
prior <- rep( 1 , 1000 )
```

Likelihood

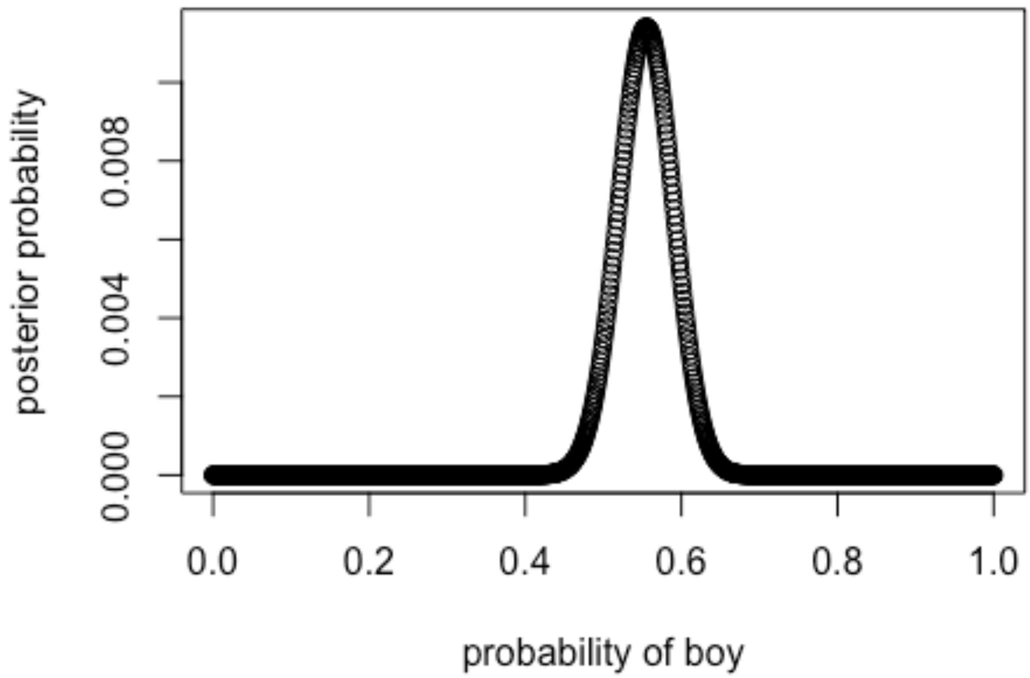
```
likelihood = dbinom(sum(birth1)+sum(birth2),size=length(birth1)+length(birth
2),prob=p_grid)
```

Posterior

```
posterior = likelihood*prior
posterior = posterior / sum(posterior)
```

Visualization

```
plot( p_grid , posterior , type="b" ,
      xlab="probability of boy" , ylab="posterior probability" )
```



Max Parameter

```
max_parameter = p_grid[which(posterior==max(posterior))];max_parameter

## [1] 0.5545546
```

Question 2

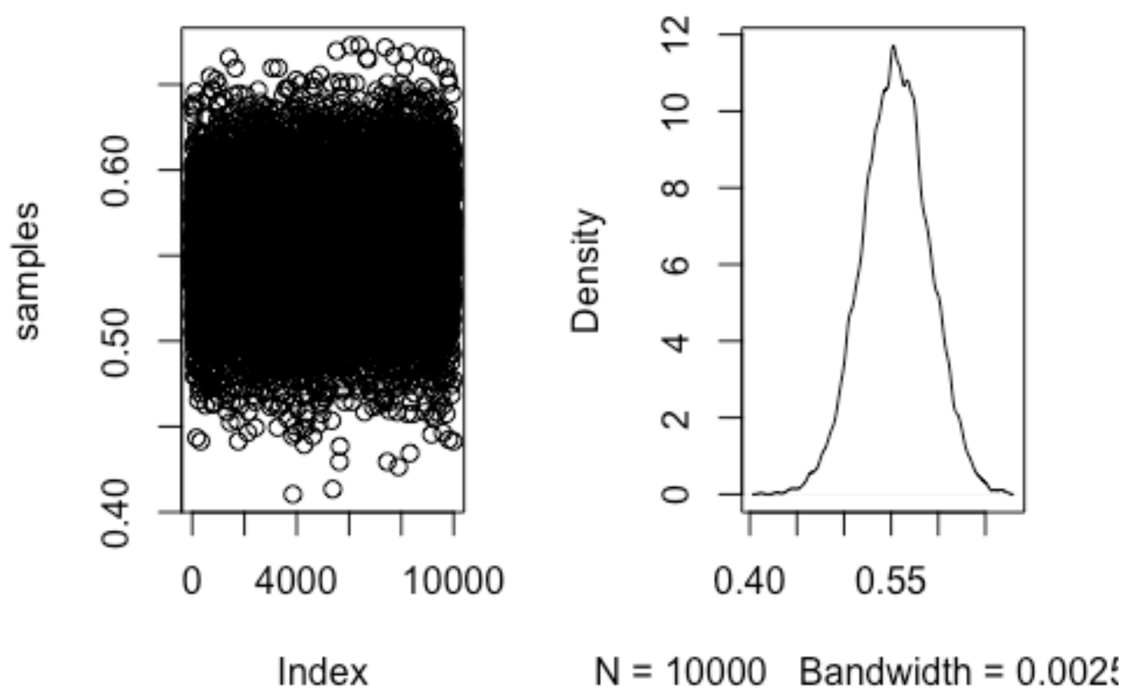
Sample function

```
samples <- sample( p_grid , prob=posterior , size=1e4 , replace=TRUE )
```

Quick view of the sample

```
library(rethinking)

## Loading required package: rstan

## Loading required package: ggplot2

## Loading required package: StanHeaders

## rstan (Version 2.16.2, packaged: 2017-07-03 09:24:58 UTC, GitRev: 2e1f913d
3ca3)
```

```
## For execution on a local, multicore CPU with excess RAM we recommend calli
ng
## rstan_options(auto_write = TRUE)
## options(mc.cores = parallel::detectCores())

## Loading required package: parallel

## rethinking (Version 1.59)

par(mfrow=c(1,2))
plot( samples )
dens( samples )
```



The 50%, 89%, and 97% highest posterior density intervals. Find the narrowest interval containing mass

```
(percnt_50 = HPDI( samples , prob=0.5 ))

##      |0.5      0.5|
## 0.5265265 0.5735736

(percnt_89 = HPDI( samples , prob=0.89 ))

##     |0.89     0.89|
## 0.5005005 0.6126126
```

```
(percnt_97 = HPDI( samples , prob=0.97 ))

##      |0.97      0.97|
## 0.4754755 0.6286286
```

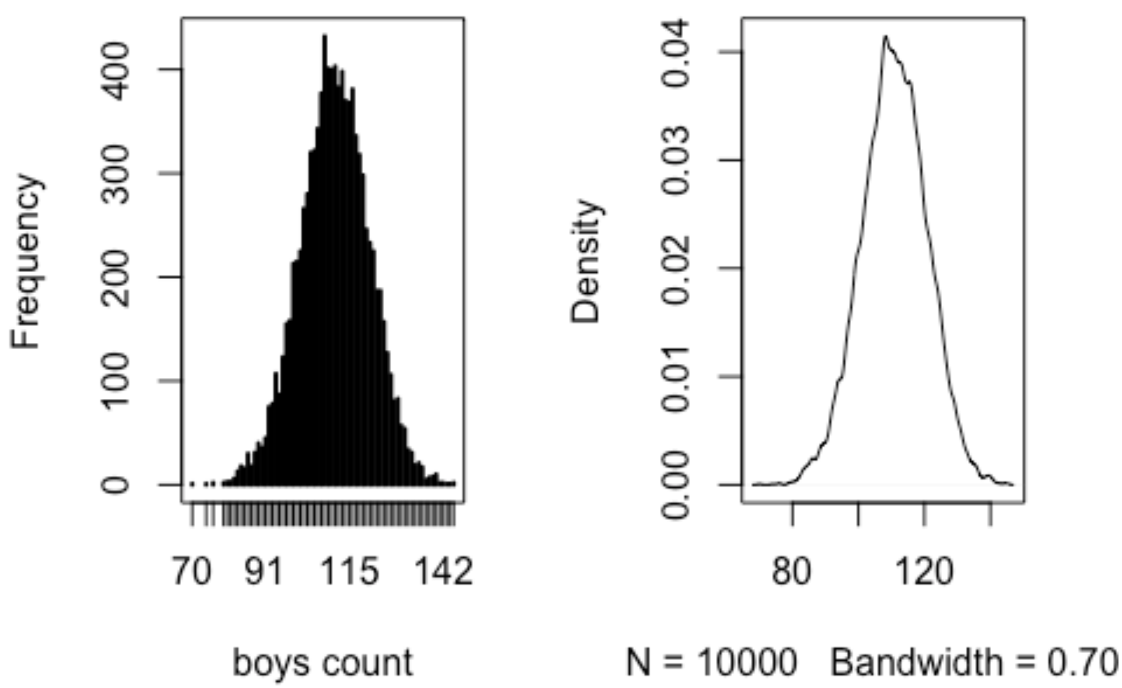
Question 3

Simulation

```
nb = rbinom(1e4, size = 200, prob = samples)
```

Visualization

```
par(mfrow=c(1,2))
simplehist( nb , xlab="boys count" )
dens(nb)
```



Find top 5 most frequent counts of boys

```
tail(sort(table(nb)),5)

## nb
## 113 110 109 111 108
## 398 400 401 403 432
```

Conclusion: The simulation outcomes are pretty close to the actual count (111). The model fits the data well.

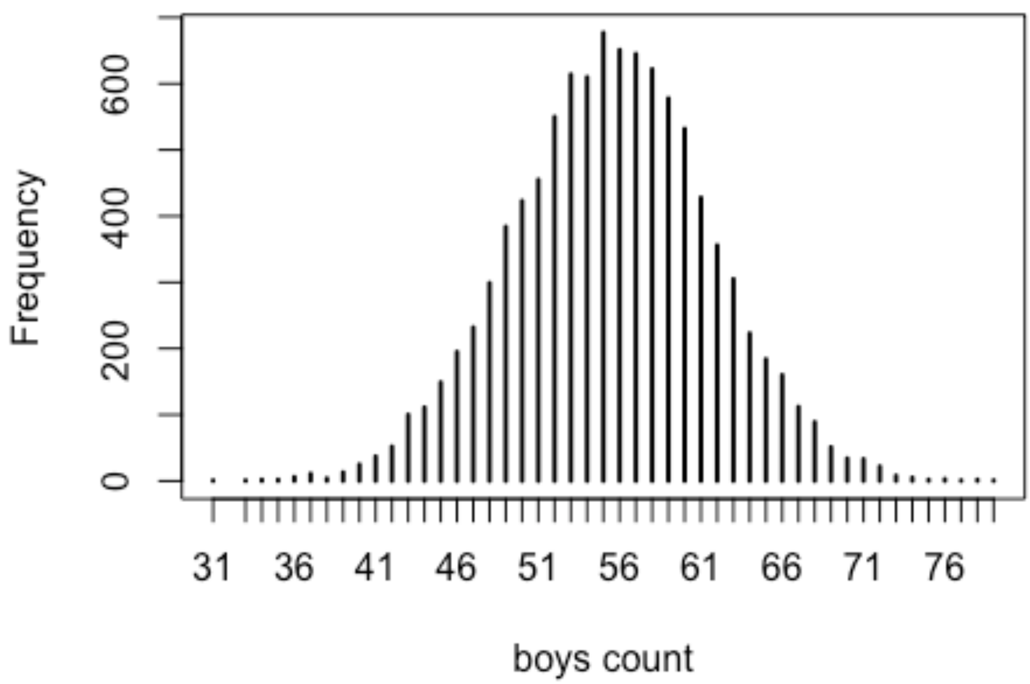
Question 4

Simulation

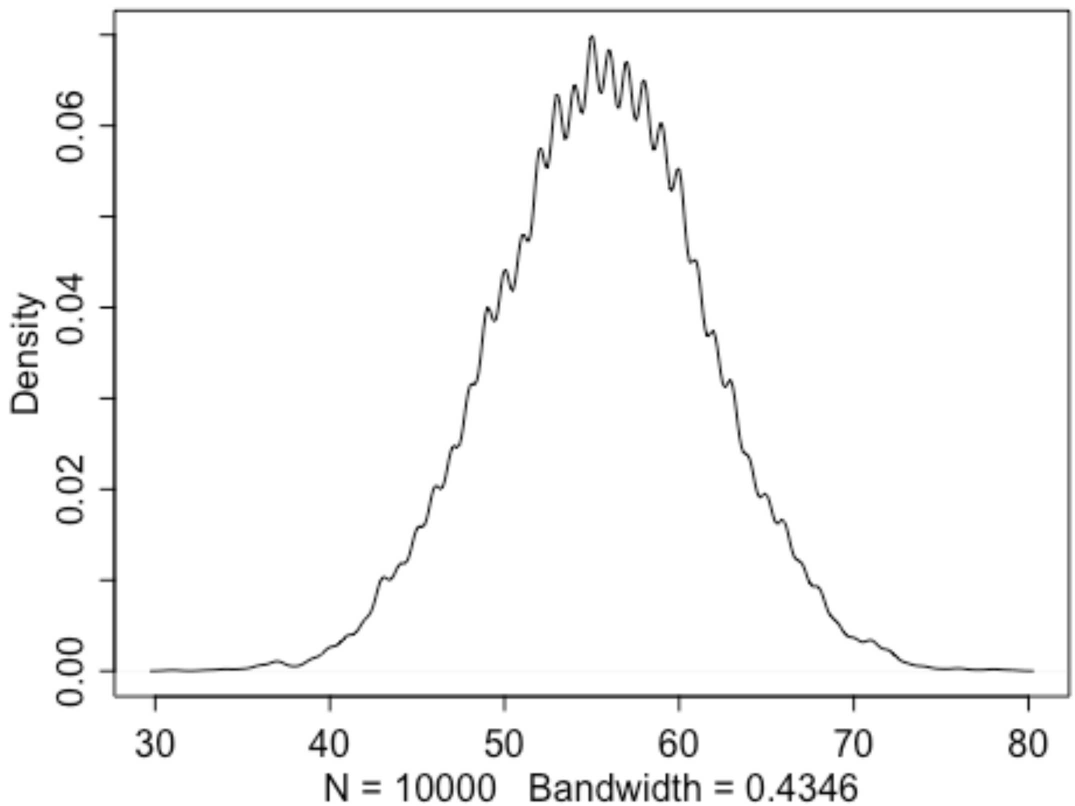
```
nb2 = rbinom(1e4, size = 100, prob = samples)
```

Visualization

```
simplehist( nb2 , xlab="boys count" )
```




```
dens(nb2)
```



Find top 5 most frequent counts of boys

```
tail(sort(table(nb2)),5)

## nb2
##  53  58  57  56  55
## 614 622 645 651 677
```

The actual outcome of first born baby as boy

```
table(birth1==1)

##
## FALSE  TRUE
##    49    51
```

Conclusion: The new simulation outcomes are not as close as it did before. The actual count is 51, while the simulated counts are around 53~57, which is a bit higher.

Question 5

How many first born babies are girls?

```
table(birth1 ==0)
```

```
##
## FALSE  TRUE
##    51    49
```

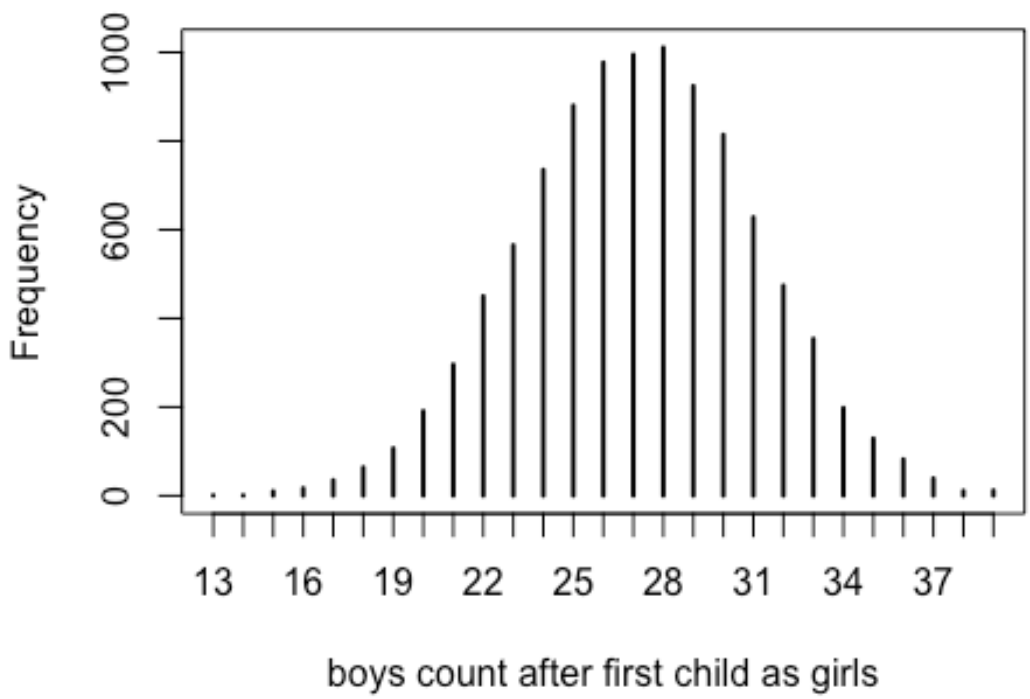
49 babies in actual data are girls

Simulate 49 babies

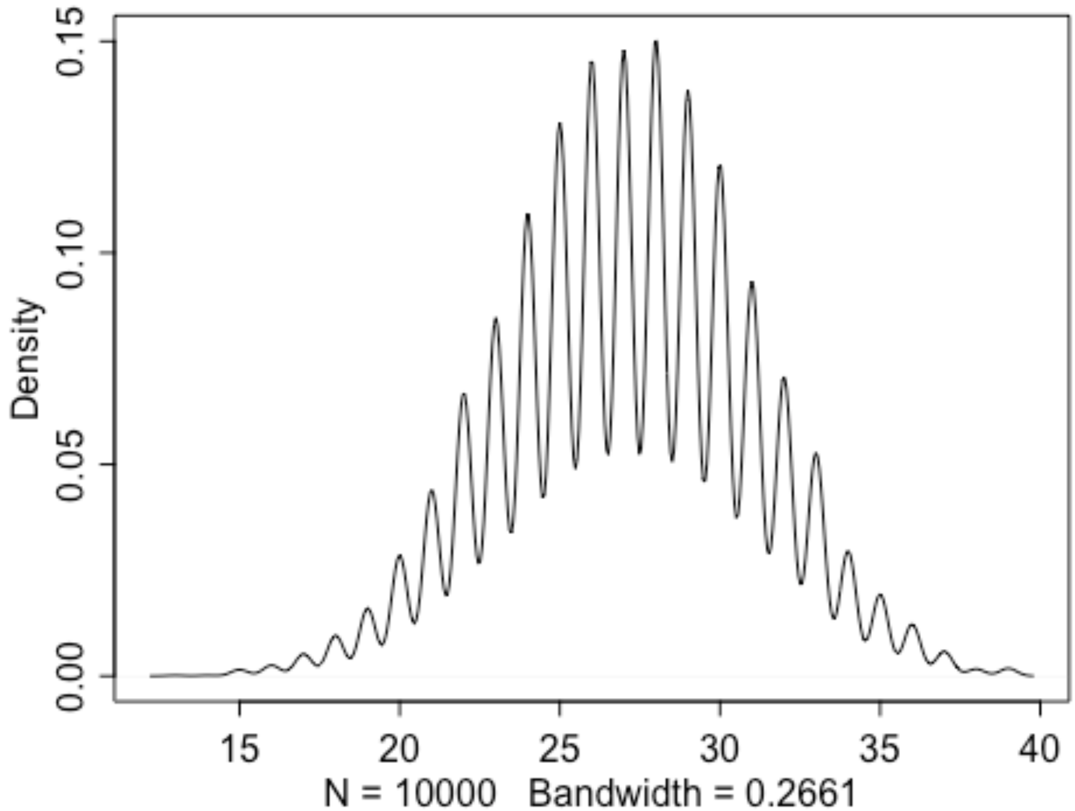
```
nb3 = rbinom(1e4, size = 49, prob = samples)
```

Visualization

```
simplehist( nb3 , xlab="boys count after first child as girls" )
```




```
dens(nb3)
```



Find top 5 most frequent counts of boys who has an older sister out of 49 samples by model

```
tail(sort(table(nb3)),5)

## nb3
##   25   29   26   27   28
##  880  924  977  995 1011
```

The actual outcome of second born boy who has an older sister

```
babies = as.data.frame(cbind(birth1,birth2))
length(which(babies$birth1 ==0 & babies$birth2==1))

## [1] 39
```

39 girls out of 49 have a little brother.

Conclusion: The new simulation outcomes are not accurate at all. The actual count is 39, while the simulated counts are around 25~29, which is pretty low and underestimated. Perheps those parents who have a girl already really wish to have a boy. Thus they may use informal therapy to increase the possibility of getting a boy. Therefore, the independent hypothesis of gender of giving birth no longer exist. That's why the actual data is not as predicted.