Geog 509: Bayesian Data Analysis
Chapter 3 Problem Set

Jesse Piburn

2020-02-03

3E1 - 3E7

```r
library(rethinking)

p_grid <- seq(from = 0,to = 1,length.out = 1000)
prior <- rep( 1 , 1000 )
likelihood <- dbinom(6, size = 9, prob = p_grid)
posterior <- likelihood * prior
posterior <- posterior / sum(posterior)
set.seed(100)
samples <- sample(p_grid, prob = posterior, size=1000, replace=TRUE)
```

```r
mean(samples < 0.2)
```

```
## [1] 0.001
```

```r
mean(samples > 0.8)
```

```
## [1] 0.091
```

```r
mean(samples > 0.2 & samples < 0.8)
```

```
## [1] 0.908
```

```r
quantile(samples, c(0.2, 0.8))
```

```
##        20%        80%
## 0.5163163 0.7427427
```

```r
HPDI(samples, prob = 0.66)
```

```
##     |0.66      0.66|
## 0.5135135 0.7697698
```

```r
PI(samples, prob = 0.66)
```

```
##        17%        83%
## 0.4961562 0.7569269
```

3H1

```r
library(rethinking)
library(tidyverse)
data(homeworkch3)

boys <- sum(birth1, birth2)
n_grid <- 200
```

```r
d <-
tibble(
  prior = 1,
  p_grid = seq(from = 0, to = 1, length.out = n_grid),
  likelihood = dbinom(boys, size = n_grid, prob = p_grid),
  unstd_posterior = likelihood * prior,
  posterior = unstd_posterior / sum(unstd_posterior)
  )

d %>%
  ggplot(aes(x = p_grid, y = posterior)) +
  geom_line() +
  labs(
    subtitle = "Grid Approximated Posterior Distrubtions",
    x = "probability of boy birth",
    y = "posterior probability") +
  theme(panel.grid = element_blank())
```


Grid Approximated Posterior Distrubtions
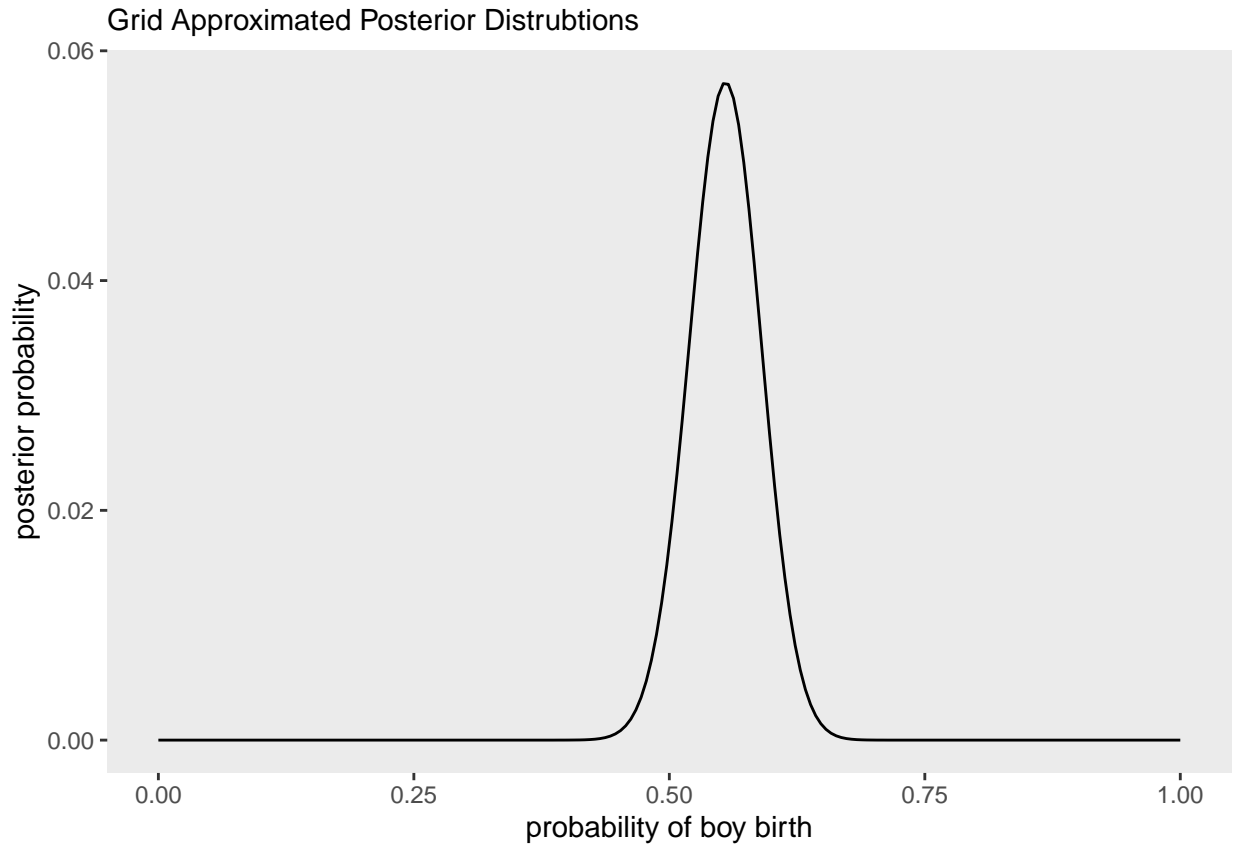
```r
# max posterior
filter(d, posterior == max(posterior))
```

```
## # A tibble: 1 x 5
##    prior p_grid likelihood unstd_posterior posterior
##    <dbl>  <dbl>      <dbl>           <dbl>     <dbl>
## 1      1  0.553     0.0566          0.0566    0.0571
```

3H2

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

HPDI(samples, prob=0.50)

##      |0.5       0.5|
## 0.5226131 0.5678392
HPDI(samples, prob=0.89)

##      |0.89      0.89|
## 0.4974874 0.6080402
HPDI(samples, prob=0.97)

##      |0.97      0.97|
## 0.4824121 0.6331658
```
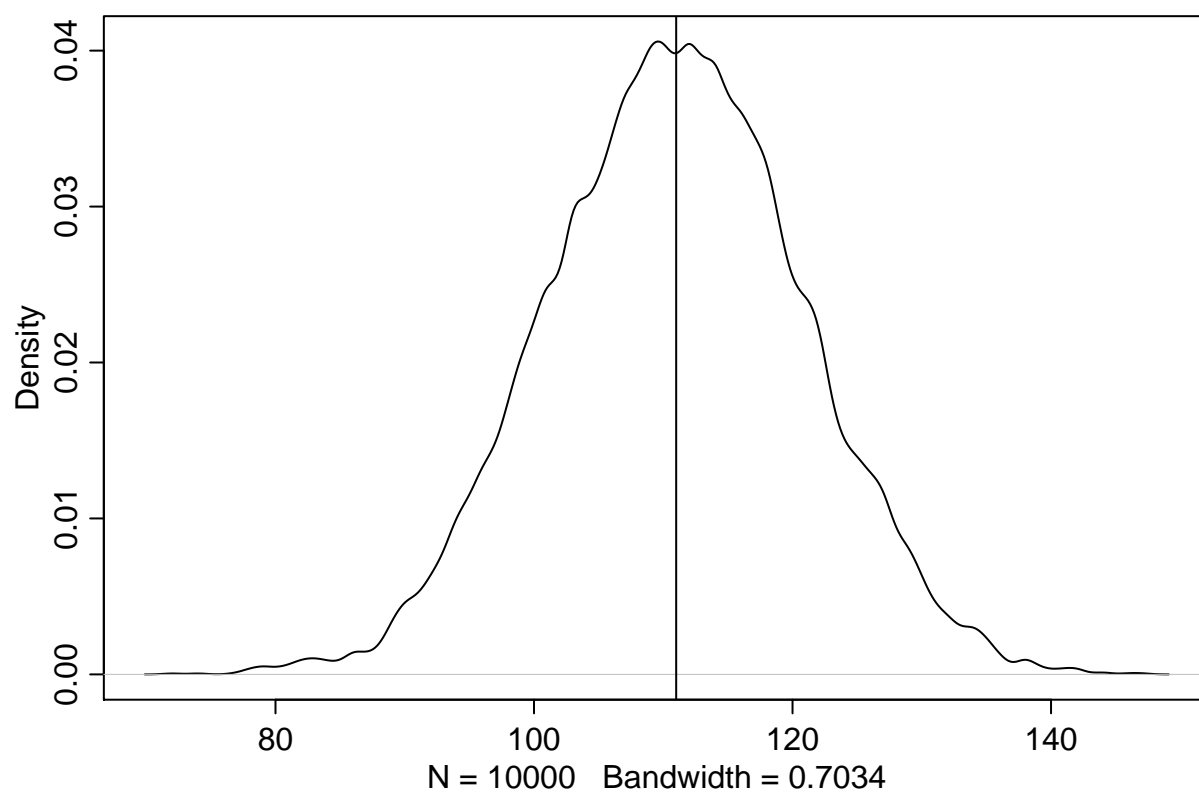
3H3

```
simulation <- rbinom(1e4, size=200, prob=samples)
dens(simulation)
abline(v=111)
```

3H4

```r
birth1sum <- sum(birth1)
likelihood2 <- dbinom(birth1sum, size=100, prob=d$p_grid)
prior2 <- rep( 1, 100)
unstd.posterior2 <- likelihood2 * prior2
posterior2 <- unstd.posterior2 / sum(unstd.posterior2)
samples2 <- sample(d$p_grid, size=1e4, replace=TRUE, prob=posterior2)

simulation2 <- rbinom(1e4, size=100, prob=samples2)
dens(simulation2)
abline(v=51)
```
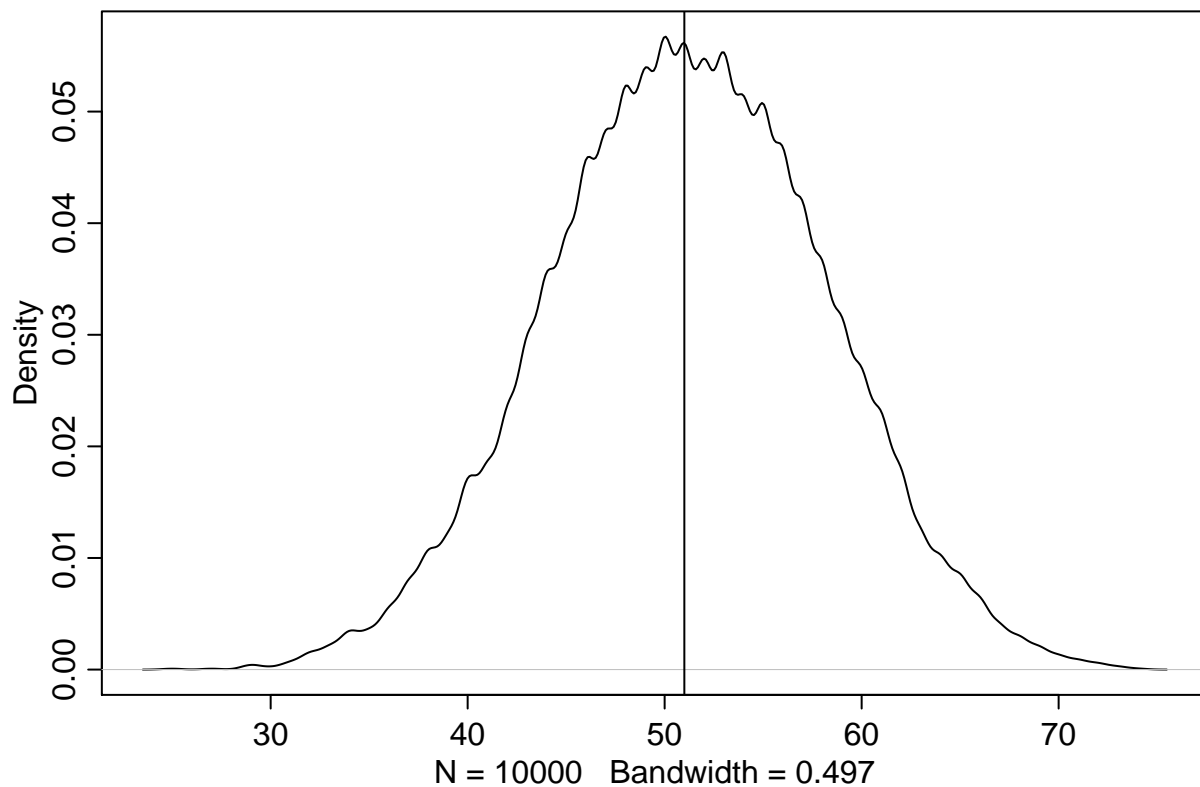


```r
simulation3 <- rbinom(1e4, size=100, prob=samples)
dens(simulation3)
abline(v=51)
```

N = 10000   Bandwidth = 0.4314
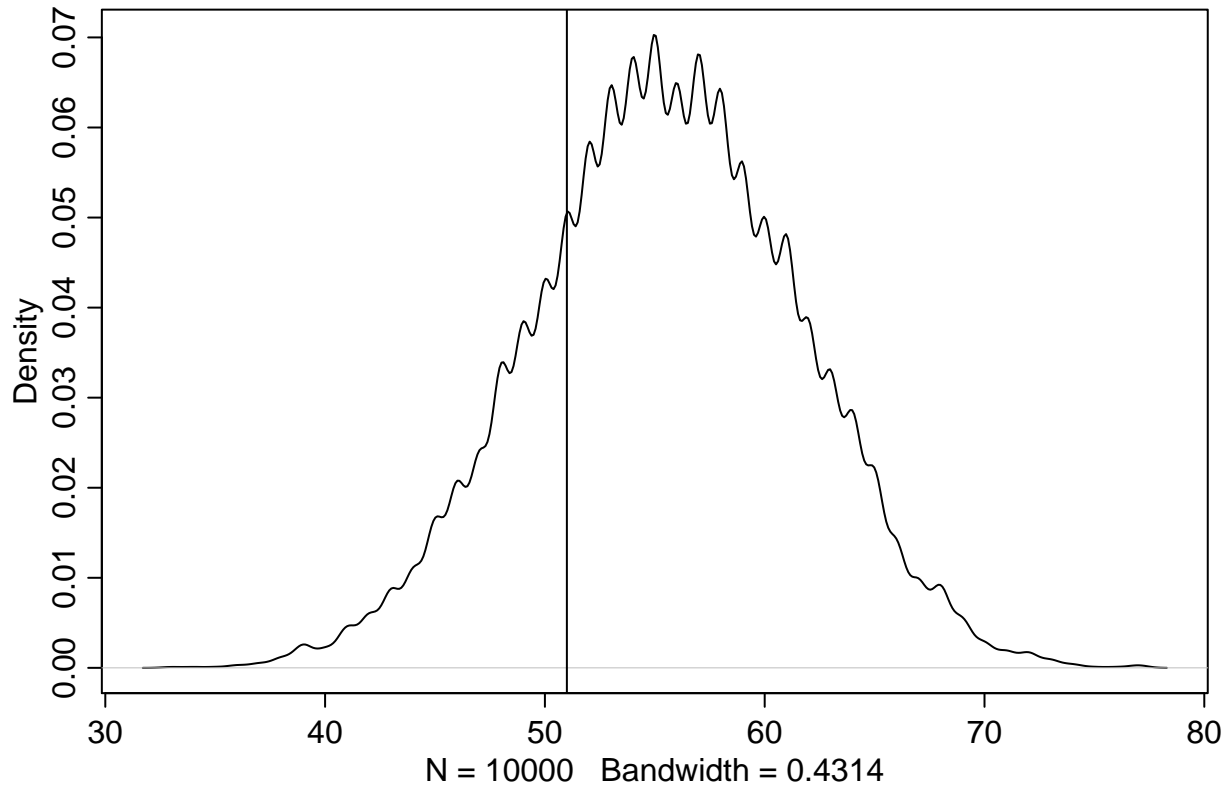
3H5

```
firstgirls <- 100 - sum(birth1)
boysaftgirls <- birth2[birth1==0]
simulation4 <- rbinom(1e4, size=firstgirls, prob=samples)
dens(simulation4)
abline(v=sum(boysaftgirls))
```

N = 10000   Bandwidth = 0.2661