BDA - Assignment 2

Anonymous

Contents

Introduction 1

Loaded packages 2

Exercise 1) 2

a) 2

Code 2

Results 2

b) 3

Code 3

Results 3

c) 4

Code 4

Result 4

d) 4

Result 4

e) 4

Code 4

Plots 5

Explanation 9

markmyassignment Report 10

Introduction

This assignment is related to chapter 1 and chapter 2 from textbook, mainly focuses on Conjugate Prior Distributions. In this assignment, monitored algae status in 274 sites at Finnish lakes and rivers are discussed. The observation follows a binomial model with parameter π, which follows a Beta(2,10) prior. The detailed derivation will not be included in this assignment.

Loaded packages

Below are packages that are used in the assignment.

```
library(aaltobda)
library(ggplot2)
library(markmyassignment)
```

Exercise 1)

a)

Code

```
data("algae")
algae_test <- c(0, 1, 1, 0, 0, 0)
y <- sum(algae == 1)
n <- length(algae)
cat("Positive observation is ", y)
```

```
## Positive observation is  44
```

```
cat("Total number of observation is", n)
```

```
## Total number of observation is 274
```

```
prior_alpha <- 2
prior_beta <- 10
posterior_alpha <- prior_alpha + y
posterior_beta <- n + prior_beta - y
cat("posterior_alpha is ", posterior_alpha)
```

```
## posterior_alpha is  46
```

```
cat("posterior_beta is ", posterior_beta)
```

```
## posterior_beta is  240
```

Results

(1) The likelihood

$$\begin{aligned} p(y|\pi) &= \binom{n}{y} \pi^y (1-\pi)^{n-y} \\ &= \binom{274}{44} \pi^{44} (1-\pi)^{274-44} \\ &= \binom{274}{44} \pi^{44} (1-\pi)^{230} \end{aligned}$$

(2) The prior

$$\begin{aligned} p(\pi) &= Beta(\alpha, \beta) \\ p(\pi) &= Beta(2, 10) \end{aligned}$$

(3) The resulting posterior

$$
\begin{aligned}
p(\pi|y) &= Beta(\alpha + y, \beta + n - y) \\
p(\pi|y) &= Beta(2 + 44, 10 + 274 - 44) \\
p(\pi|y) &= Beta(46, 240)
\end{aligned}
$$

b)

Code

```r
beta_point_est <- function(prior_alpha, prior_beta, data){
  y <- sum(data == 1)
  n <- length(data)
  est <- (prior_alpha + y)/(prior_alpha + prior_beta + n)
  return(est)
}
beta_interval <- function(prior_alpha, prior_beta, data, prob){
  y <- sum(data == 1)
  n <- length(data)
  posterior_alpha <- prior_alpha + y
  posterior_beta <- n + prior_beta - y
  left_q <- (1-prob)/2
  right_q <- 1-(1-prob)/2
  left_bd <- qbeta(left_q, posterior_alpha, posterior_beta)
  right_bd <- qbeta(right_q, posterior_alpha, posterior_beta)
  result_interval <- c(left_bd, right_bd)
  return(result_interval)
}
beta_point_est(prior_alpha = 2, prior_beta = 10, data = algae)
```

```
## [1] 0.1608392
```

```r
beta_interval(prior_alpha = 2, prior_beta = 10, data = algae, prob = 0.9)
```

```
## [1] 0.1265607 0.1978177
```

Results

$$
\begin{aligned}
E(\pi|y) &= \frac{\alpha + y}{\alpha + \beta + n} \\
&= \frac{2 + 44}{2 + 10 + 274} \\
&= \frac{46}{286} \\
&\approx 0.1608392
\end{aligned}
$$

The point estimate is roughly 0.1608392. The 90% posterior interval is [0.1265607, 0.1978177]

c)

Code

```
beta_low <- function(prior_alpha, prior_beta, data, pi_0){
  y <- sum(data == 1)
  n <- length(data)
  posterior_alpha <- prior_alpha + y
  posterior_beta <- n + prior_beta - y
  prob <- pbeta(pi_0, posterior_alpha, posterior_beta)
  return(prob)
}
beta_low(prior_alpha = 2, prior_beta = 10, data = algae, pi_0 = 0.2)
```

```
## [1] 0.9586136
```

Result

The probability of the proportion of monitoring sites with detectable algae levels π is smaller than $\pi_0 = 0.2$ that is known from historical records is 0.9586136.

d)

Result

The required assumption is that the observation in algae in each lake or river should be independent and identically distributed bernoulli distribution with parameter π, only if the assumption holds, can the binomial model of y hold. Since in real life, different lakes or rivers may have different probability of having detectable blue-green algae levels, and the distribution of algae maybe dependent.

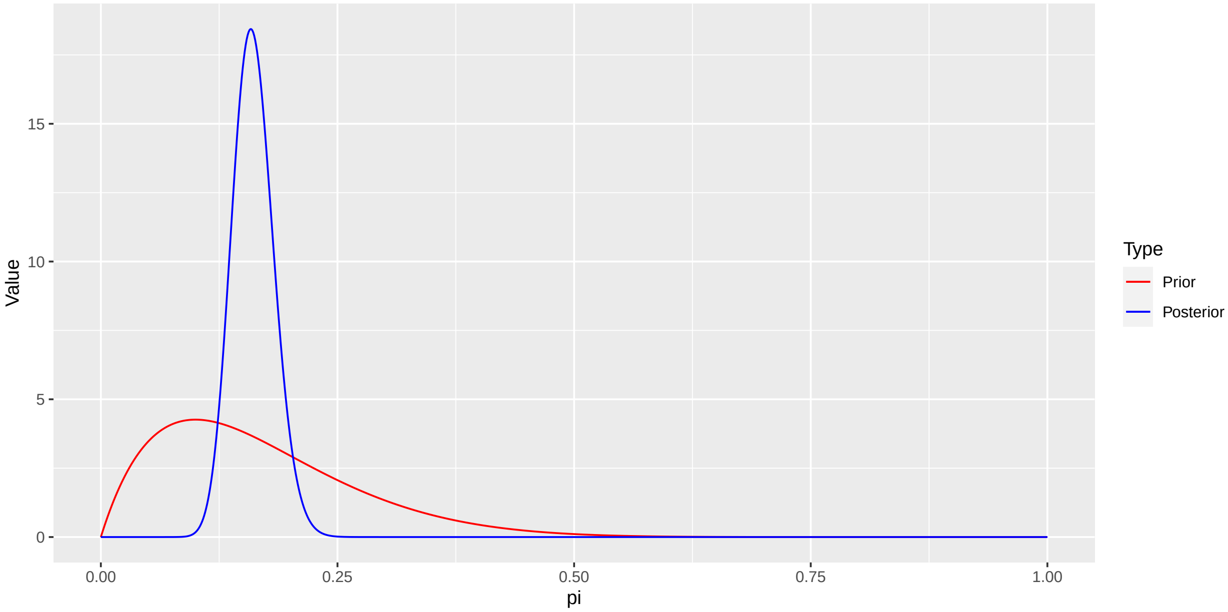
e)

Code

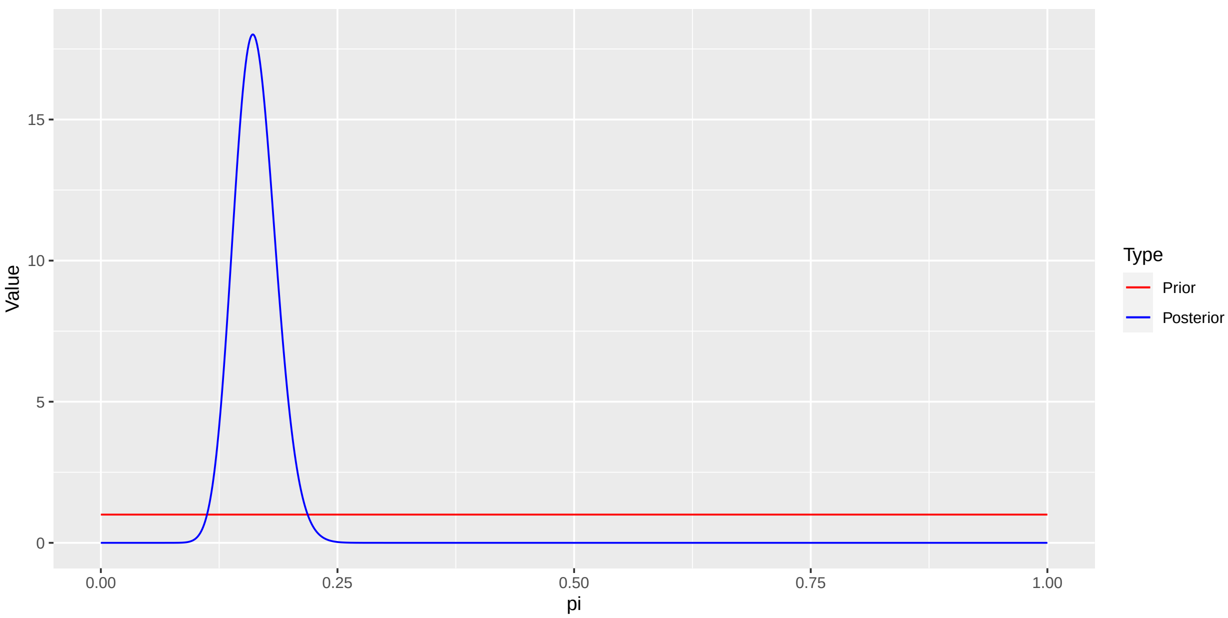
```
df <- data.frame(pi = seq(0, 1.0, 0.001))
prior_sensitivity_analysis <- function(df, prior_alpha, prior_beta, data){
  y <- sum(data == 1)
  n <- length(data)
  posterior_alpha <- prior_alpha + y
  posterior_beta <- n + prior_beta - y
  df$prior <- dbeta(df$pi, prior_alpha, prior_beta)
  df$posterior <- dbeta(df$pi, posterior_alpha, posterior_beta)
  plt<- (ggplot(df)
         +geom_line(aes(x = pi, y = prior, colour="Prior"))
         +geom_line(aes(x = pi, y = posterior, colour="Posterior"))
         +labs(y = "Value")
         +scale_color_manual(name = "Type", values = c("Prior" = "red", "Posterior" = "blue"))
           )
  plot(plt)
}
```

Plots

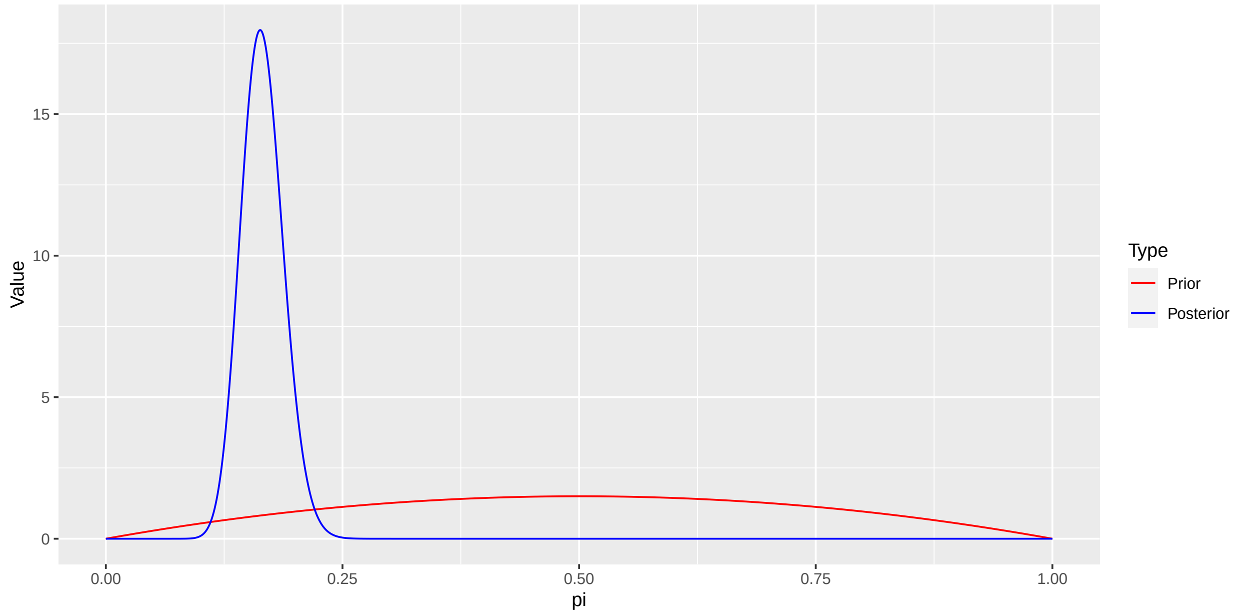
```
prior_sensitivity_analysis(df, 2, 10, algae)
```



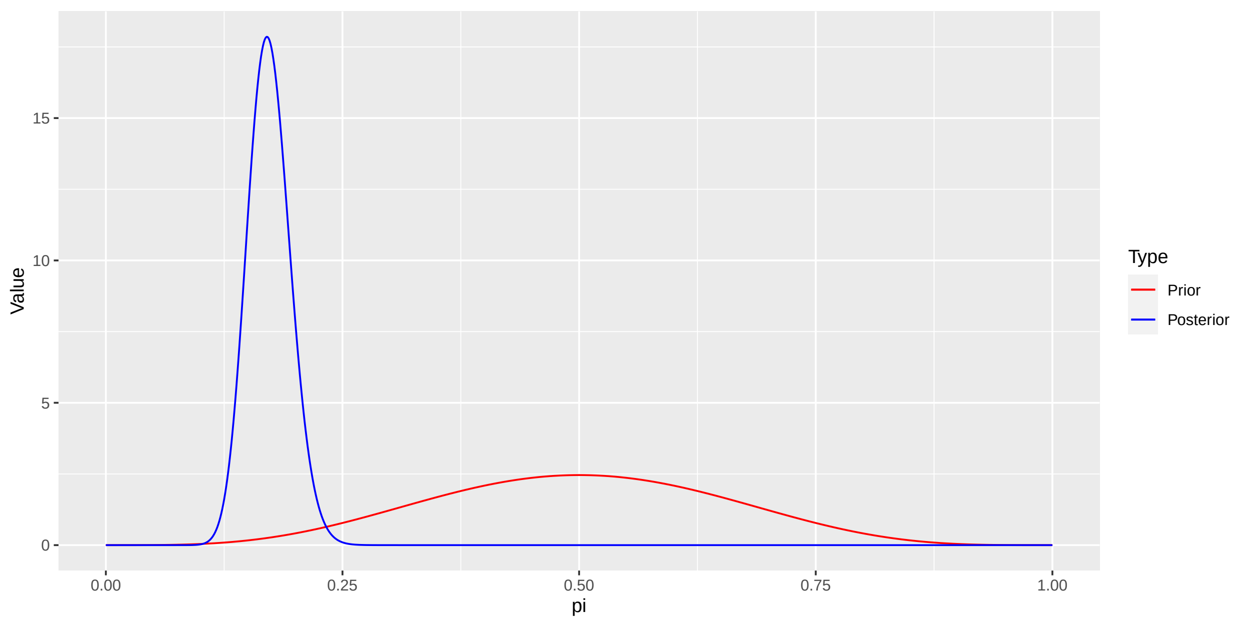

```
prior_sensitivity_analysis(df, 1, 1, algae)
```



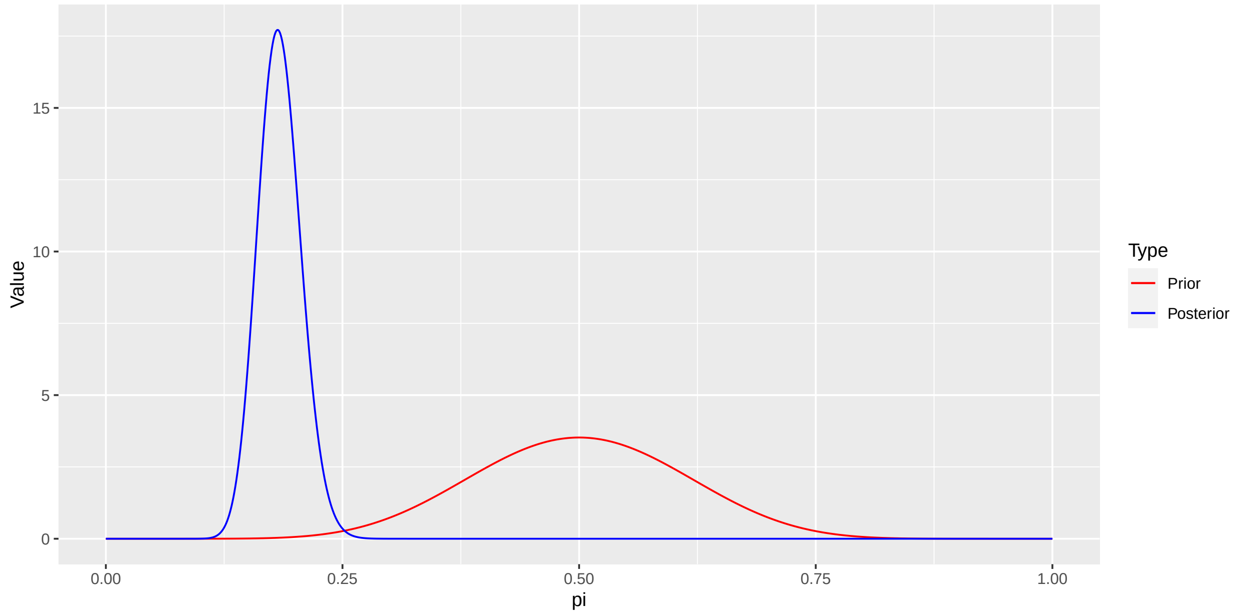
```
prior_sensitivity_analysis(df, 2, 2, algae)
```



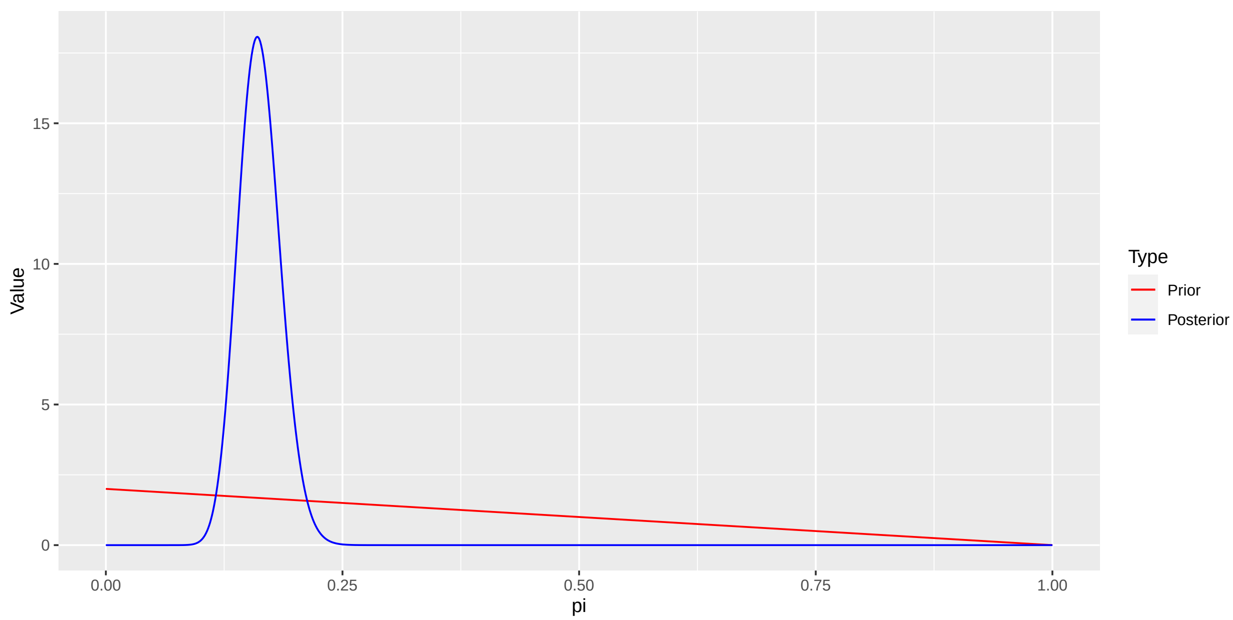

```
prior_sensitivity_analysis(df, 5, 5, algae)
```



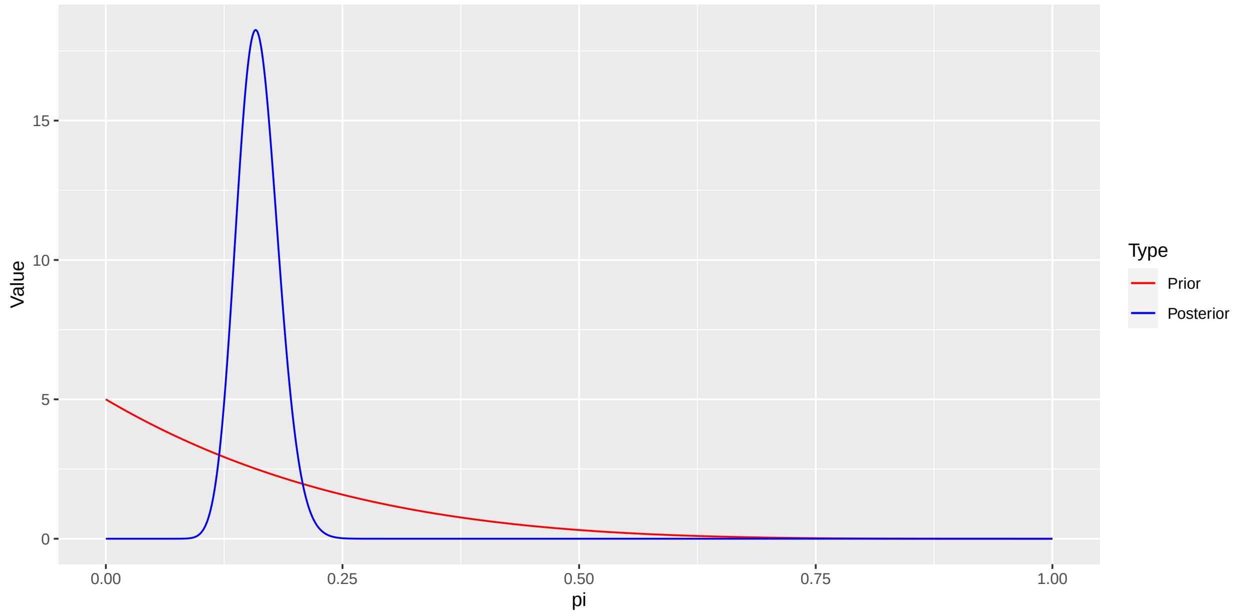
```r
prior_sensitivity_analysis(df, 10, 10, algae)
```



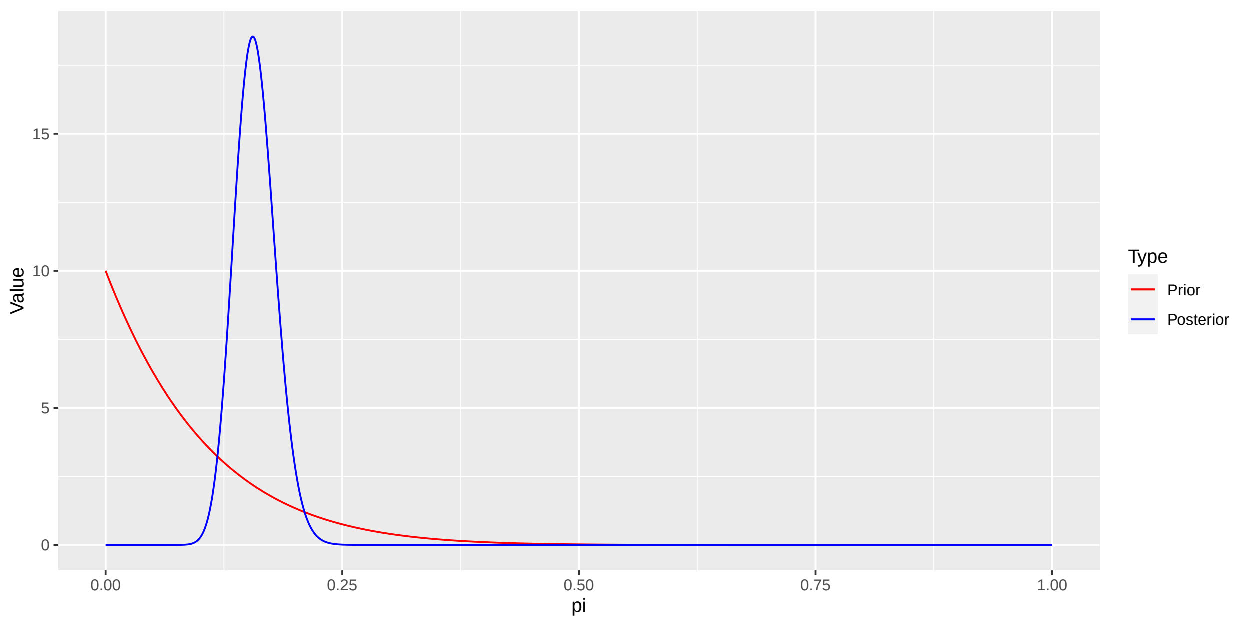

```r
prior_sensitivity_analysis(df, 1, 2, algae)
```



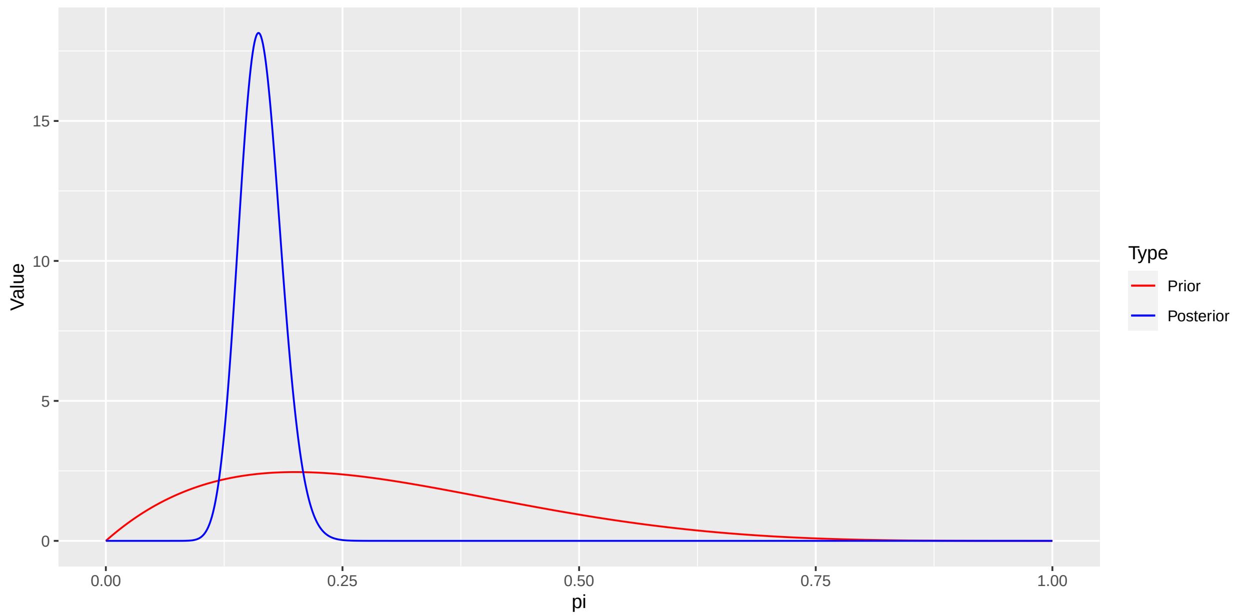
```
prior_sensitivity_analysis(df, 1, 5, algae)
```



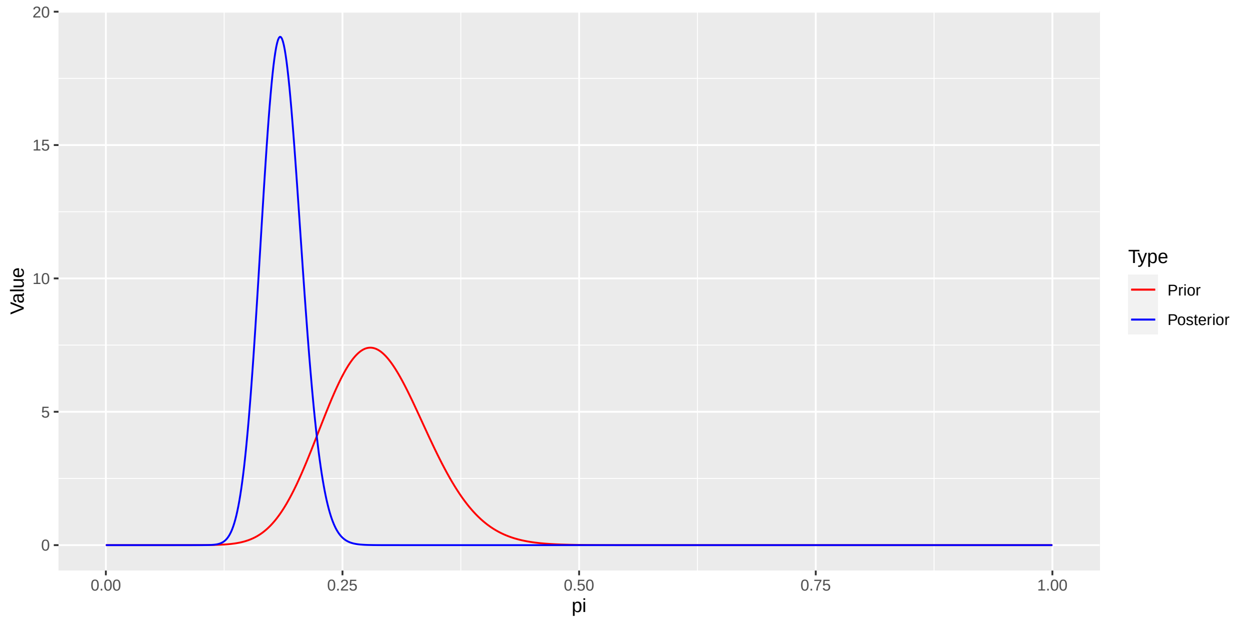

```
prior_sensitivity_analysis(df, 1, 10, algae)
```



```r
prior_sensitivity_analysis(df, 2, 5, algae)
```




```r
prior_sensitivity_analysis(df, 20, 50, algae)
```



Explanation

From the above figures we can see that no matter how the α and β of prior change, the posterior is very robust, always keep the similar shape and value.

markmyassignment Report

```r
assignment_path <-
  paste("https://github.com/avehtari/BDA_course_Aalto/",
        "blob/master/assignments/tests/assignment2.yml", sep="")
set_assignment(assignment_path)
```

```
## Assignment set:
## assignment2: Bayesian Data Analysis: Assignment 2
## The assignment contain the following (3) tasks:
## - beta_point_est
## - beta_interval
## - beta_low
```

```r
# To check your code/functions, just run mark_my_assignment()
mark_my_assignment()
```

```
## v | F W S  OK | Context
##
## / |         0 | task-1-subtask-1-tests
## / |         0 | beta_point_est()
## v |         5 | beta_point_est()
##
## / |         0 | task-2-subtask-1-tests
## / |         0 | beta_interval()
## v |         5 | beta_interval()
##
## / |         0 | task-3-subtask-1-tests
## / |         0 | beta_low()
## v |         5 | beta_low()
##
## == Results =================================================================
## Duration: 0.2 s
##
## [ FAIL 0 | WARN 0 | SKIP 0 | PASS 15 ]
## Yay! All done!
```