

# BDA - Assignment 2

## Contents

Exercise 1.. . . . .	1
a) . . . . .	1
b) . . . . .	1
c) . . . . .	2
d) . . . . .	2

## Exercise 1.

a)

```
library(aaltobda)
data("algae")
```

My model goes as follows. We have our prior distribution  $\pi \sim \text{Beta}(2, 10)$ , the likelihood  $P(y|\pi) \sim \text{binomial}$ , and the resulting posterior  $P(\pi|y) \sim \text{Beta}(\alpha + y; \beta + n - y)$ .  
Now let's define the function *beta\_point\_est* and *beta\_interval* for calculating point estimate and interval estimate of the posterior.

```
beta_point_est<-function(prior_alpha, prior_beta, data){
n_total<-length(data)
n_algae<-sum(data)
return((prior_alpha+n_algae)/(prior_alpha+n_algae+prior_beta+n_total-n_algae))
}

beta_interval<-function(prior_alpha, prior_beta, data,prob){
n_total<-length(data)
n_algae<-sum(data)
interval<-qbeta(c((1-prob)/2,1-(1-prob)/2), prior_alpha+n_algae, prior_beta+n_total-n_algae)
return (interval)
}

beta_point_est(prior_alpha = 2, prior_beta = 10, data = algae)
```

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

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

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

So for the value of the unknown  $\pi$  according to the observations and the prior knowledge, the point estimate is 0.1608392 and the 90% interval estimate is [0.1265607,0.1978177].

b)

Next let's define our function for calculating the probability that the proportion of monitoring sites with detectable algae levels  $\pi$  is smaller than  $\pi_0 = 0.2$  that is known from historical records. This function simply gets the resulting posterior first, and returns the value of its cumulative density function with input equals  $\pi_0$ , that is,  $P(\pi < \pi_0|y)$ .

```
beta_low<-function(prior_alpha, prior_beta, data, pi_0){
n_total<-length(data)
```

```
n_algae<-sum(data)
prob<-pbeta(pi_0,prior_alpha+n_algae, prior_beta+n_total-n_algae)
return(prob)
}
beta_low(prior_alpha = 2, prior_beta = 10, data = algae, pi_0 = 0.2)
```

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

So the probability that the proportion of monitoring sites with detectable algae levels  $\pi$  is smaller than  $\pi_0$  is 0.9586136.

c)

The assumption is that the algae status monitored in different sites are independent with each other and follows identical distribution(in other words same  $\pi$ ).

d)

Here I would like to choose uniform prior and beta prior to compare. For uniform prior, it's simple, we consider that  $\pi$  is uniformly distributed from 0 to 1, which can be described as  $Beta(1,1)$ , then we use our derived function to calculate the point estimation and interval estimation.

```
beta_point_est(prior_alpha = 1, prior_beta = 1, data = algae)
```

```
## [1] 0.1630435
```

```
beta_interval(prior_alpha = 1, prior_beta = 1, data = algae, prob = 0.9)
```

```
## [1] 0.1279681 0.2008987
```

It gives 0.163 as point estimation, which doesn't change too much if we merely use MLE ( $\hat{\theta} = \frac{y}{n} = 0.16$ ).

For beta prior, we would like to change one of its parameter with another parameter fixed to have a rough view of the prior sensitivity.

```
posterior_evaluation_alpha<-function(prior_beta,data){
n_total<-length(data)
n_algae<-sum(data)
prior_alphas <- seq(1, 20, length = 20)
beta_point_estimations<-list()
beta_interval_estimations<-list()
for (alpha in prior_alphas)
{
  beta_point_estimation<-beta_point_est(alpha, prior_beta,data)
  beta_interval_estimation<-beta_interval(alpha, prior_beta,data,prob=0.9)
  beta_point_estimations=c(beta_point_estimations,beta_point_estimation)
  beta_interval_estimations=c(beta_interval_estimations,beta_interval_estimation)
}
return (list(beta_point_estimations,beta_interval_estimations))
}
```

```
posterior_evaluation_beta<-function(prior_alpha,data){
n_total<-length(data)
n_algae<-sum(data)
prior_betas <- seq(1, 20, length = 20)
beta_point_estimations<-c()
beta_interval_estimations<-c()
for (beta in prior_betas)
```

```

{
  beta_point_estimation<-beta_point_est(prior_alpha,beta,data)
  beta_interval_estimation<-beta_interval(prior_alpha,beta,data,prob=0.9)
  beta_point_estimations=c(beta_point_estimations,beta_point_estimation)
  beta_interval_estimations=c(beta_interval_estimations,beta_interval_estimation)
}
return (list(beta_point_estimations,beta_interval_estimations))
}

estimations1<-posterior_evaluation_alpha(10,algae)
estimations2<-posterior_evaluation_beta(2,algae)

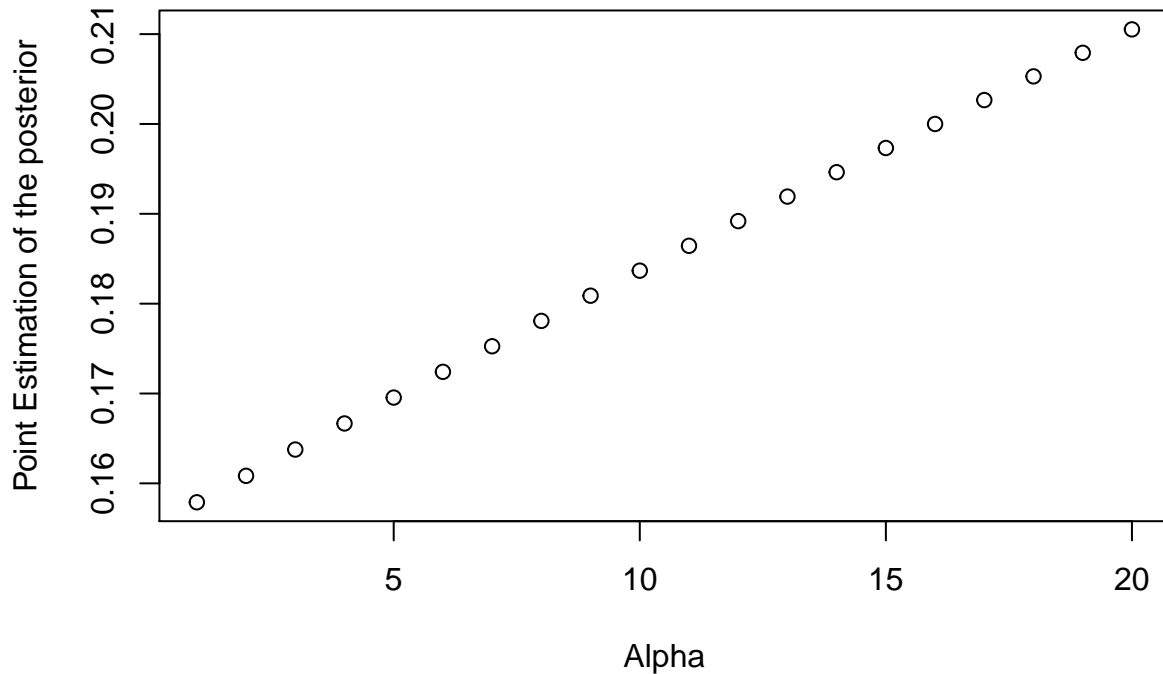
alpha_point_estimation=c()
alpha_interval_estimation_upperbound=c()
alpha_interval_estimation_lowerbound=c()
alpha_interval_estimation_bounds=c()
for (est in estimations1[1])
{alpha_point_estimation=c(alpha_point_estimation,est)}
for (est in estimations1[[2]])
{alpha_interval_estimation_bounds=c(alpha_interval_estimation_bounds,est)}
for (i in seq(1,40,length=40))
{
  if (i%%2==0)
  {
    alpha_interval_estimation_lowerbound=c(alpha_interval_estimation_lowerbound,+
      alpha_interval_estimation_bounds[i])
  }
  else
  {
    alpha_interval_estimation_upperbound=c(alpha_interval_estimation_upperbound,+
      alpha_interval_estimation_bounds[i])
  }
}
beta_point_estimation=c()
beta_interval_estimation_upperbound=c()
beta_interval_estimation_lowerbound=c()
beta_interval_estimation_bounds=c()
for (est in estimations2[1])
{beta_point_estimation=c(beta_point_estimation,est)}
for (est in estimations2[[2]])
{beta_interval_estimation_bounds=c(beta_interval_estimation_bounds,est)}
for (i in seq(1,40,length=40))
{
  if (i%%2==0)
  {
    beta_interval_estimation_lowerbound=c(beta_interval_estimation_lowerbound,+
      beta_interval_estimation_bounds[i])
  }
  else
  {
    beta_interval_estimation_upperbound=c(beta_interval_estimation_upperbound,+
      beta_interval_estimation_bounds[i])
  }
}

```

```

}
prior_alphas <- seq(1, 20, length = 20)
prior_betas <- seq(1, 20, length = 20)
st<-"Point Est of posterior with beta =12 changing alpha"
plot(prior_alphas,alpha_point_estimation,sub=st,xlab="Alpha",ylab="Point Estimation of the posterior")

```

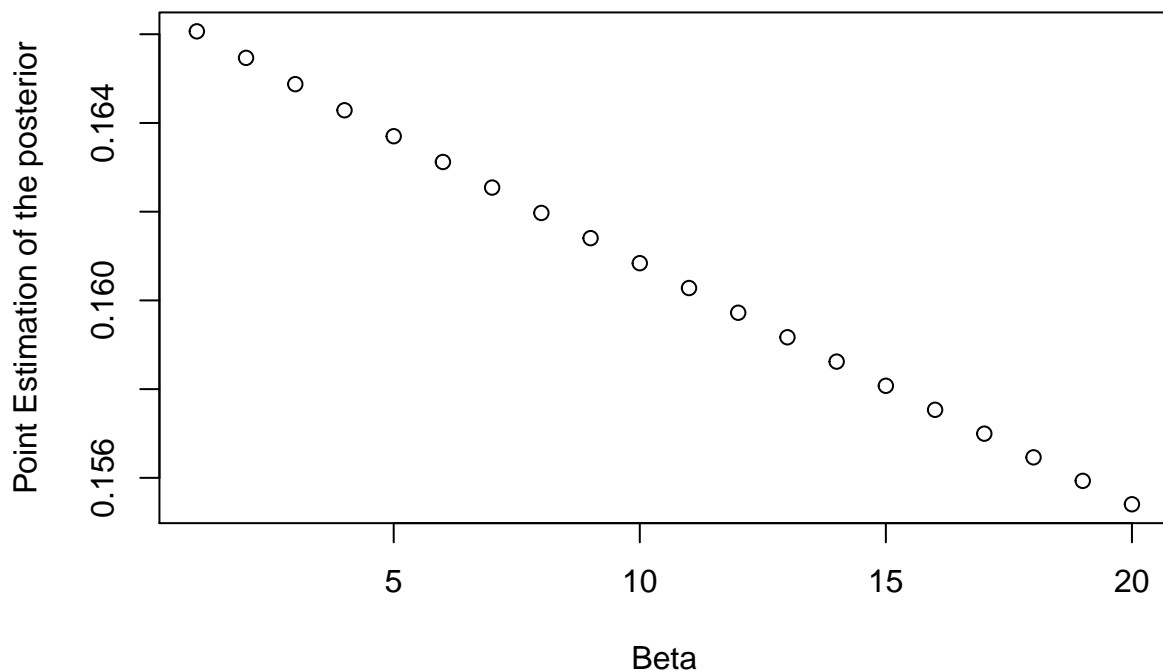


Point Est of posterior with beta =12 changing alpha

```

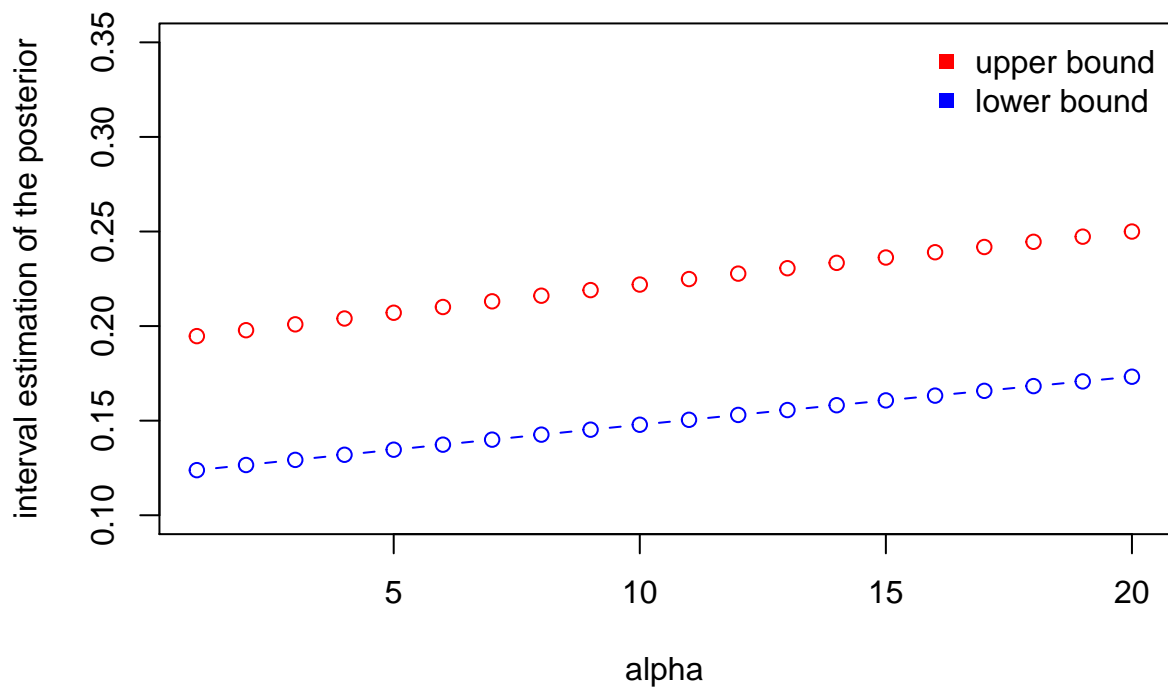
st<-"Point Est of posterior with alpha =12 changing beta"
plot(prior_betas,beta_point_estimation,sub=st,xlab="Beta",ylab="Point Estimation of the posterior")

```



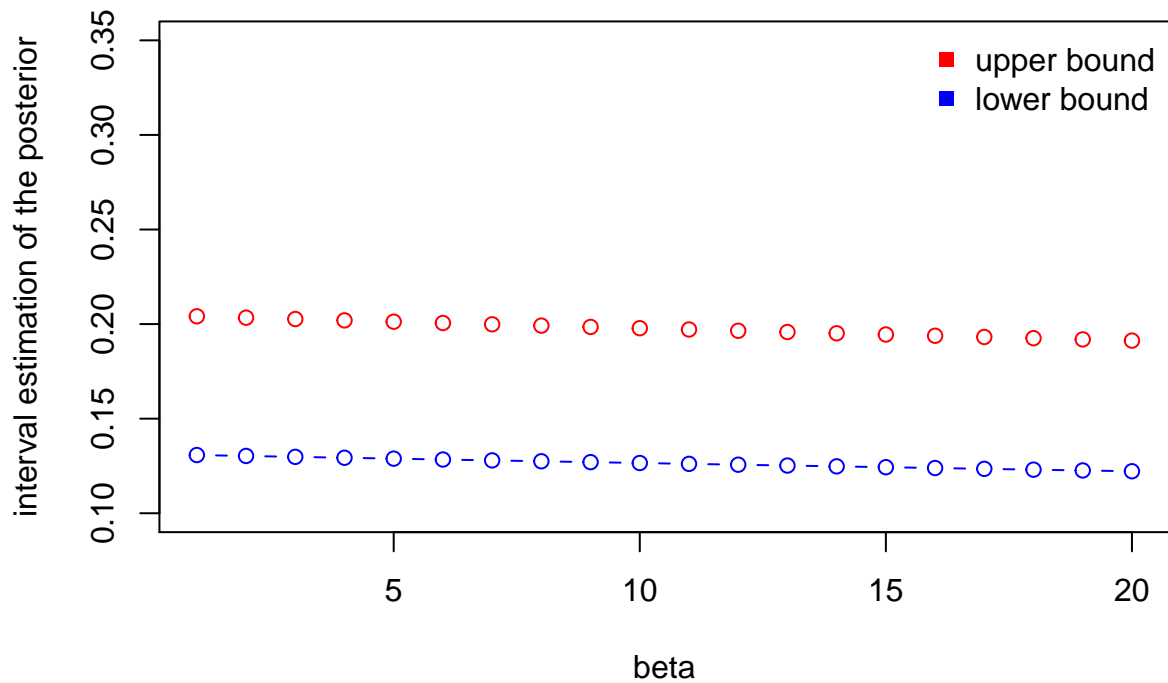
Point Est of posterior with alpha =12 changing beta

```
x_1<-"alpha"
y_1<-"interval estimation of the posterior"
st<-"Interval Est of posterior with beta =12 changing alpha"
plot(prior_alphas,alpha_interval_estimation_lowerbound,ylim=c(0.10,0.35),col=2,sub=st,xlab=x_1,ylab=y_1)
legend("topright",pch=c(15,15),legend=c("upper bound","lower bound"),col=c(2,4),bty="n")
lines(prior_alphas,alpha_interval_estimation_upperbound,col=4,type="b")
```



Interval Est of posterior with beta =12 changing alpha

```
x_1<-"beta"
y_1<-"interval estimation of the posterior"
st<-"Interval Est of posterior with beta =12 changing alpha"
plot(prior_betas,beta_interval_estimation_lowerbound,ylim=c(0.10,0.35),col=2,xlab=x_1,sub=st,ylab=y_1)
legend("topright",pch=c(15,15),legend=c("upper bound","lower bound"),col=c(2,4),bty="n")
lines(prior_betas,beta_interval_estimation_upperbound,col=4,type="b")
```



Interval Est of posterior with beta =12 changing alpha

In conclusion, using uniform prior gives us the estimation of  $\pi = 0.16$ . And also we can see from the plots that using beta prior and changing one of its parameter with another parameter fixed gives us results without oscillating much from the results using MLE or using uniform prior. Consequently we can say that the prior sensitivity of the posterior distribution is low.