BDA - Assignment 2

Anonymous

Contents

Inference for Binomial proportion

We load and use algae status monitored in $N = 274$ sites at Finnish lakes and rivers. The data contains binary measurements of algae presence ('0': no algae, '1': algae present).

```
# To install aaltobda, see the General information in the assignment.
library(aaltobda)
data("algae")
```

a.) We use a Beta distribution as the prior density for the Binomial proportion. Conversely, the posterior distribution is a Beta distribution itself:

$$p(\theta|y) \propto \text{Beta}(\alpha_{prior} + y, \beta_{prior} + N - y)$$

b.) The posterior mean can be calculated as:

$$\alpha_{posterior} = \alpha_{prior} + y$$
$$\beta_{posterior} = \beta_{prior} + N - y$$
$$\theta_{posterior} = \frac{\alpha_{posterior}}{\alpha_{posterior} + \beta_{posterior}}$$

We can calculate the point estimate using the followng function.

```
beta_point_est<-function(prior_alpha,prior_beta,data){
  N<-length(data)
  y<-sum(data)
  posterior_beta<-prior_beta+N-y
  posterior_alpha<-prior_alpha+y
  posterior_alpha/(posterior_alpha+posterior_beta)
}
```

Now, after observing the algae data the point estimate for $\theta_{posterior}$ is:

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

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

```
beta_interval<-function(prior_alpha,prior_beta,data,prob){
  N<-length(data)
  y<-sum(data)
  posterior_beta<-prior_beta+N-y
```

```r
  posterior_alpha<-prior_alpha+y
  prob<-prob+0.05
  qbeta(c(1-prob,prob),posterior_alpha,posterior_beta)
}
```

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

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

c.) The probability of the Binomial proportion being smaller than 0.2can be calculated as:

```r
beta_low<-function(prior_alpha,prior_beta,data,pi_0){
  N<-length(data)
  y<-sum(data)
  posterior_beta<-prior_beta+N-y
  posterior_alpha<-prior_alpha+y
  pbeta(pi_0,posterior_alpha,posterior_beta)
}
```

```r
beta_low(prior_alpha = 2, prior_beta = 10, data = algae,pi_0=0.2)
```

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

d.) All monitoring sites need to be independent each other so the algae levels are conditionally independent. If we take into account the location of the site, some sites could be located far away from the others and this assumption may hold. Otherwise, we would also have to take into account the influence of neighboring sites.

e.) In order to perform snsitivity analysis, we take samples from the posterior distribution $p(\theta|y)$ according to the Beta-Binomial model.

```r
posterior_samples<-function(prior_alpha,prior_beta,data,num_samples){
  N<-length(data)
  y<-sum(data)
  posterior_beta<-prior_beta+N-y
  posterior_alpha<-prior_alpha+y
  rbeta(num_samples,posterior_alpha,posterior_beta)
}
```

```r
library(gridExtra)
library(ggplot2)
library(grid)
library(lattice)
```

We first try different values for the α_{prior} parameter:
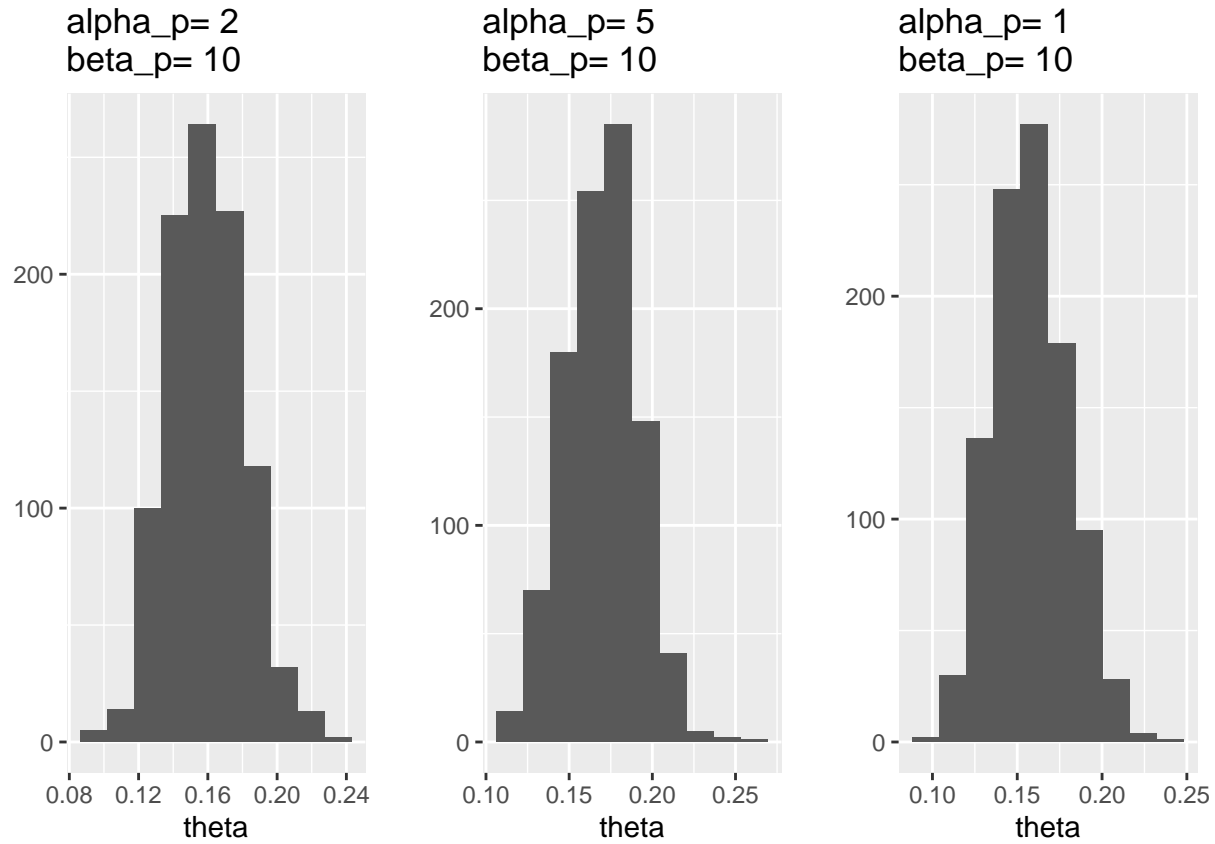
```r
alpha_prior<-2
beta_prior<-10
x1<-posterior_samples(alpha_prior,beta_prior,algae,1000)
p1<- qplot(x1,geom="histogram",bins=10,xlab="theta") + ggtitle(paste("alpha_p=",alpha_prior,"\nbeta_p="

alpha_prior<-5
beta_prior<-10
x2<-posterior_samples(alpha_prior,beta_prior,algae,1000)
p2<- qplot(x2,geom="histogram",bins=10,xlab="theta") + ggtitle(paste("alpha_p=",alpha_prior,"\nbeta_p="

alpha_prior<-1
beta_prior<-10
x3<-posterior_samples(alpha_prior,beta_prior,algae,1000)
```

```
p3<- qplot(x3,geom="histogram",bins=10,xlab="theta") + ggtitle(paste("alpha_p=",alpha_prior,"\nbeta_p="

grid.arrange(p1, p2,p3, nrow = 1)
```


Different α_{prior} parameteres yield different posterior values for the binomial proportions θ. It worth noting that $\alpha_{prior} = 1$ produces a sharp concentration on the posterior mode, while increasing this parameter increases the variance of the posterior distribution.

Now, we try different values for the β_{prior} parameter:

```
alpha_prior<-2
beta_prior<-1
x4<-posterior_samples(alpha_prior,beta_prior,algae,1000)
p4<- qplot(x4,geom="histogram",bins=10,xlab="theta") + ggtitle(paste("alpha_p=",alpha_prior,"\nbeta_p="

alpha_prior<-2
beta_prior<-5
x5<-posterior_samples(alpha_prior,beta_prior,algae,1000)
p5<- qplot(x5,geom="histogram",bins=10,xlab="theta") + ggtitle(paste("alpha_p=",alpha_prior,"\nbeta_p="

alpha_prior<-2
beta_prior<-15
x6<-posterior_samples(alpha_prior,beta_prior,algae,1000)
p6<- qplot(x6,geom="histogram",bins=10,xlab="theta") + ggtitle(paste("alpha_p=",alpha_prior,"\nbeta_p="

grid.arrange(p4, p5,p6, nrow = 1)
```
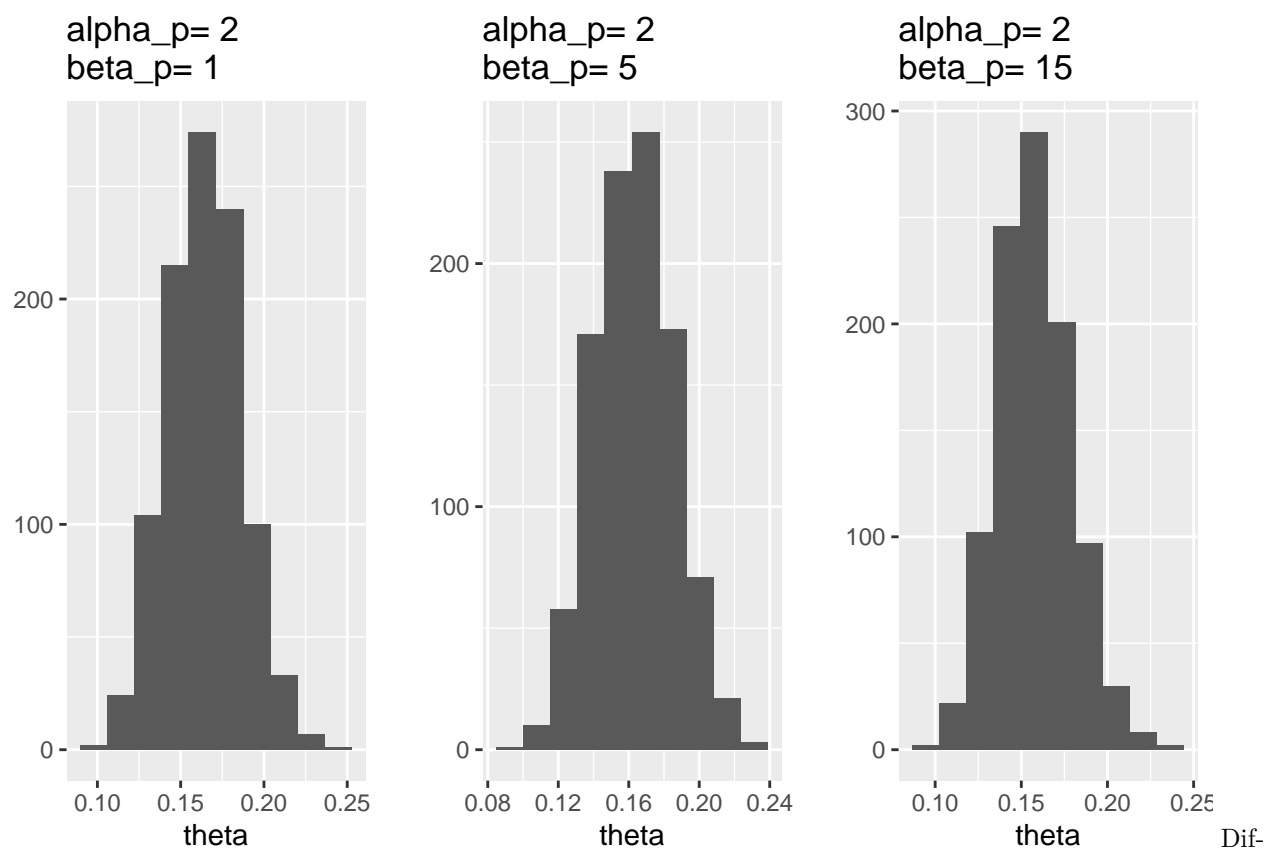
Different β_{prior} parameteres also yield different posterior values for the binomial proportions θ. Increasing this parameter modifies the shape of the posterior distribution.