BDA - Assignment 2

Anonymous

Contents

List of Tables


```
library(markmyassignment)
exercise_path = 'https://github.com/avehtari/BDA_course_Aalto/blob/master/exercises/tests/ex2.yml'
set_assignment(exercise_path)
```

```
## Assignment set:
## ex2: Bayesian Data Analysis: Assignment 2
## The assignment contain the following (3) tasks:
## - beta_point_est
## - beta_interval
## - beta_low
```

```
library(aaltobda)
data('algae')
head(algae)
```

```
## [1] 0 1 1 0 0 0
```

```
algae_test = c(0,1,1,0,0,0)
```

Exercise 1 a

- Likelihood:

=> p(y | π) = π^y * (1 − π)^(n−y)

=> p(44 | π) = π^44 * (1 − π)^(274−44)

=> p(44 | π) = π^44 * (1 − π)^(230)

- Prior:

=>Beta(π| α, β) π^(α-1) * (1- π)^(β-1)

=> Beta(π| 2, 10) π^(2-1) * (1- π)^(10-1)

=> Beta(π| 2, 10) π^(1) * (1- π)^(9)

- Posterior:

=> Beta(π| α + y, β) + n - y)

=> Beta(π| 2 + 44, 10 + 274 - 44)

=> Beta(π| 46, 240)

```
beta_point_est = function(prior_alpha, prior_beta, data) {
  y = sum(data)
  n.minus.y = sum(data < 1)
  cat('y is',y, 'and n minus y is', n.minus.y, '\n')

  mean = sum(prior_alpha, y) / sum(prior_alpha, y, prior_beta, n.minus.y)
  cat('The point estimate with Beta(2,10) prior is:', mean, '\n')
}

beta_point_est(2, 10, algae)
```

```
## y is 44 and n minus y is 230
## The point estimate with Beta(2,10) prior is: 0.1608392
```

```
beta_interval = function(prior_alpha, prior_beta, data, prob_interval) {
  y = sum(data)
  n.minus.y = sum(data < 1)

  an = sum(prior_alpha, y)
  bn = sum(prior_beta, n.minus.y)

  lower_q = qbeta(0.05, an, bn)
  upper_q = qbeta(0.95, an, bn)

  cat('The 90% interval estimate with Beta(2,10) prior is [', lower_q,',', upper_q, ']')
}

beta_interval(2, 10, algae, 0.9)
```

```
## The 90% interval estimate with Beta(2,10) prior is [ 0.1265607 , 0.1978177 ]
```

Exercise 1 b

```r
beta_low = function(prior_alpha, prior_beta, data, pi_0){
  y = sum(data)
  n.minus.y = sum(data < 1)

  an = sum(prior_alpha, y)
  bn = sum(prior_beta, n.minus.y)

  prob = pbeta(pi_0, an, bn)
  cat('The probability that the proportion of monitoring sites with\n detectable algae levels',
      'pi is smaller than pi = 0.2 with Beta(2,10) prior is:', prob,'\n',
      '(',prob*100,'%)')
}

beta_low(2, 10, algae, 0.2)
```

```
## The probability that the proportion of monitoring sites with
##  detectable algae levels pi is smaller than pi = 0.2 with Beta(2,10) prior is: 0.9586136
##  ( 95.86136 %)
```

Exercise 1 c

Assumptions

- We assume that the level of algae has exactly two outcomes: '0': no algae and '1': algae present (exchangeability)
- We assume that the algae levels n are conditionally independent given theta, with the probability of algae present equal to theta for all cases
- We assume theta to have prior uniform distribution [0,1]

Exercise 1 d

```r
beta_sensitivity = function(prior_alpha, prior_beta, data, prob_inteval) {

  cat('This is a function for building a sensitivity table')
  cat('\n')
  cat('####################################################')
  cat('\n')
  prior = prior_alpha / sum(prior_alpha, prior_beta)
  prior2 = prior_alpha + prior_beta
  cat('Prior alpha / (alpha + beta):',prior)
  cat('\n')
  cat('Prior alpha + beta:', prior2)
  cat('\n')

  y = sum(data)
  n.minus.y = sum(data < 1)
  point_est = sum(prior_alpha, y) / sum(prior_alpha,y, prior_beta, n.minus.y)
  cat('The point estimate:', point_est, '\n')
```

```r
  an = sum(prior_alpha, y)
  bn = sum(prior_beta, n.minus.y)
  lower_q = qbeta(0.05, an, bn)
  upper_q = qbeta(0.95, an, bn)
  cat('The 90% interval estimate: [', lower_q,',', upper_q, '] \n')
  cat('################################################### \n')
}

beta_sensitivity(16.6666667, 83.333333,algae, 0.9)
```

```
## This is a function for building a sensitivity table
## ###################################################
## Prior alpha / (alpha + beta): 0.1666667
## Prior alpha + beta: 100
## The point estimate: 0.1622103
## The 90% interval estimate: [ 0.1319532 , 0.1945234 ]
## ###################################################
```

```r
library(knitr)
library(kableExtra)

df = matrix(c(0.5, 2, 0.375, '[ 0.1287564 , 0.6587386 ]',
             0.1666667, 5, 0.160693, '[ 0.1260208 , 0.1981342 ]',
             0.1666667, 10, 0.1607981, '[ 0.1264086 , 0.1979071 ]',
             0.1666667, 20, 0.1609977, '[ 0.1271531 , 0.1974676 ]',
             0.1666667, 100, 0.1622103, '[ 0.1319532 , 0.1945234 ]',
             0.1666667, 200, 0.1631505, '[ 0.1360914 , 0.1918269 ]'), ncol = 4, byrow = TRUE)
colnames(df) = c('alpha / alpha + beta', 'alpha + beta', 'Posterior median of theta',
                 '90% posterior interval for theta')
df = as.table(df)

kable(df, 'latex', booktabs = T, caption = 'Sensitivity analysis', linesep = "") %>%
  kable_styling(latex_options = c('striped', 'hold_position', 'scale_down')) %>%
  column_spec(2:5, bold = T)
```

Table 1: Sensitivity analysis

	alpha / alpha + beta	alpha + beta	Posterior median of theta	90% posterior interval for theta
A	**0.5**	**2**	**0.375**	**[0.1287564 , 0.6587386]**
B	**0.1666667**	**5**	**0.160693**	**[0.1260208 , 0.1981342]**
C	**0.1666667**	**10**	**0.1607981**	**[0.1264086 , 0.1979071]**
D	**0.1666667**	**20**	**0.1609977**	**[0.1271531 , 0.1974676]**
E	**0.1666667**	**100**	**0.1622103**	**[0.1319532 , 0.1945234]**
F	**0.1666667**	**200**	**0.1631505**	**[0.1360914 , 0.1918269]**

```r
cat('Table 1 shows that posterior inferences are not sensitive to prior distribution.\n',
    'This means that the Bayesian analysis is robust to the change of prior.')
```

```
## Table 1 shows that posterior inferences are not sensitive to prior distribution.
##  This means that the Bayesian analysis is robust to the change of prior.
```