

BDA - Assignment 2

Anonymous

Contents

```
# To install aaltobda, see the General information in the assignment.
library(aaltobda)
library(markmyassignment)
assignment_path <- paste("https://github.com/avehtari/BDA_course_Aalto/", "blob/master/assignments/tests", 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
```

```
data("algae")
algae
```

[illegible]

Algae status is monitored in 274 sites at Finnish lakes and rivers. The observations for the 2008 algae status at each site are presented in file `algae.txt` ('0': no algae, '1': algae present).

π_i : the probability of a monitoring site having detectable blue-green algae levels y_i : the observations in algae

Use a binomial model for the observations y and a Beta(2, 10) prior for binomial model parameter π to formulate a Bayesian model. Here it is not necessary to derive the posterior distribution for π as it has already been done in the book and it suffices to refer to that derivation.

Your task is to make Bayesian inference for binomial model and answer questions based on it.

a)

1)

Likelihood $p(y|\pi) = \text{Bin}(y|n, \pi) = \text{Bin}(44|274, \pi)$

2)

Prior $p(\pi) = \text{Beta}(\alpha, \beta), \text{Beta}(2, 10)$

3)

Posterior $p(\pi|y) = \text{Beta}(\pi|\alpha + y, \beta + n - y), \text{Beta}(46, 240)$

b)

$E(\pi|y)$ and posterior interval

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

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

  a <- posterior_alpha
  b <- posterior_beta

  c(qbeta((1-prob) - (1-prob)/2, a, b), qbeta(prob + (1-prob)/2, a, b))
}
```

c)

```
beta_low <- function(prior_alpha, prior_beta, data, pi_0){
  y <- sum(data)
  n <- length(data)
  posterior_alpha <- prior_alpha + y
  posterior_beta <- prior_beta + n - y
  a <- posterior_alpha
  b <- posterior_beta

  pbeta(pi_0, a, b)
}
```

Checking that everything is correct...

```
mark_my_assignment()
```

```
## v | OK F W S | Context
##
```

```

/ | 0      | beta_point_est()
v | 5      | beta_point_est()
##
/ | 0      | beta_interval()
v | 5      | beta_interval()
##
/ | 0      | beta_low()
v | 5      | beta_low()
##
## == Results =====
## Duration: 0.2 s
##
## OK:      15
## Failed:  0
## Warnings: 0
## Skipped: 0
## Good work!

```

d)

Assumptions are required in order to use this kind of a model with this type of data:

1)

Data is i.i.d

2)

Conjugacy -> The property that the posterior distribution follows the same parametric form as the prior distribution (in this case beta distribution)

3)

Informative prior -> Assumption, that we know something about the phenomenon from before

e)

prior1: $(\alpha, \beta) = 1$ prior2: $(\alpha, \beta) < 10$ prior3: $10 < (\alpha, \beta) < 100$ prior4: $(\alpha, \beta) > 100$

```

n = length(algae)
y = sum(algae)

# Posterior of prior1:
alpha <- 1
beta <- 1
curve(dbeta(x, alpha + y, beta + n - y), from = 0, to = 1, col = 'green')

# Posterior of prior2:
alpha <- 7

```

```

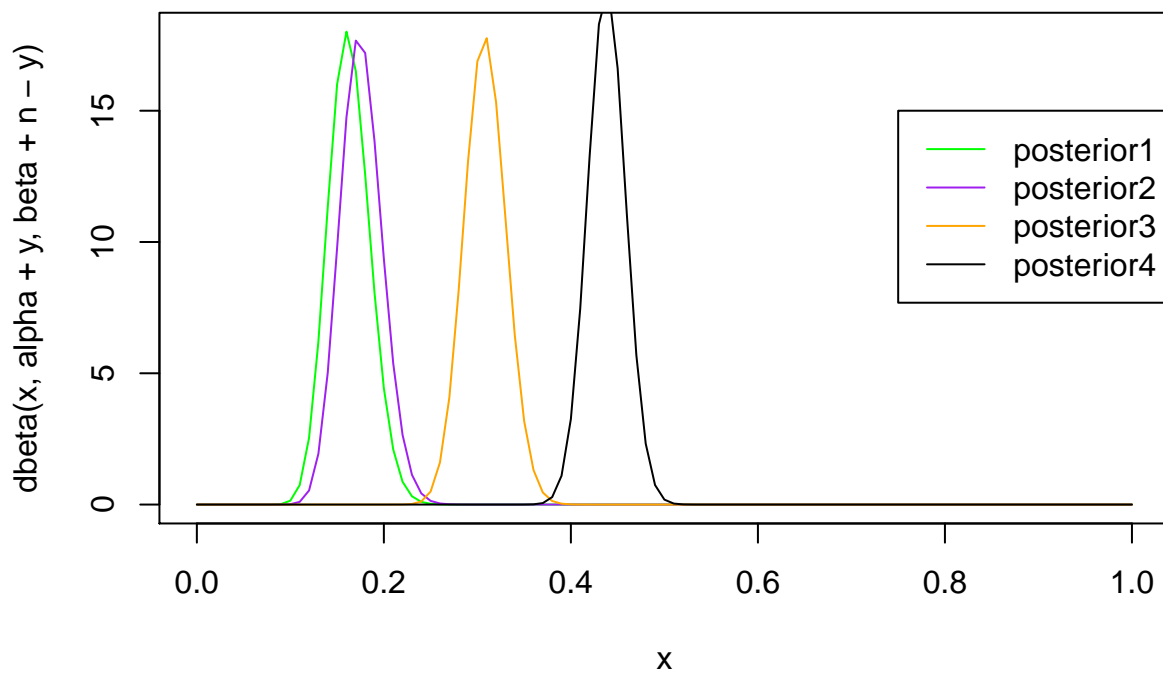
beta <- 9
curve(dbeta(x, alpha + y, beta + n - y), from = 0, to = 1, col = 'purple', add = TRUE)

# Posterior of prior3:
alpha <- 88
beta <- 66
curve(dbeta(x, alpha + y, beta + n - y), from = 0, to = 1, col = 'orange', add = TRUE)

# Posterior of prior4:
alpha <- 222
beta <- 111
curve(dbeta(x, alpha + y, beta + n - y), from = 0, to = 1, col = 'black', add = TRUE)

legend(0.75, 15, legend=c("posterior1", "posterior2", "posterior3", "posterior4"), col=c("green", "purple", "orange", "black"), add = TRUE)

```



From the plotted graph we can see that as α and β grow larger, the expected value of the distribution grows / moves towards 1.