BDA - Assignment 6

Anonymous

Contents

Load packages

```
library(aaltobda)
library("rstan")
```

```
## Loading required package: StanHeaders

## Loading required package: ggplot2

## rstan (Version 2.21.2, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)
SEED <- 48927 # set random seed for reproducability

data("bioassay")
```

Exercise 1

The code in Stan for the Bioassay model has been included in this report. However, for executing the exercises, it was written in an additional file called "bioassay_model.stan". As data, the function receives the number of experiments, the vector containing the doses for each observation, the vector containing the deaths for each observation and, finally, the vector containing the number of animals for each observation. The vector theta contains alpha and beta, the two parameters of the multinomial distribution that defines the prior. Finally, the model was defined as:

$$y_i | \theta_i \propto Bin(\theta_i, n_i)$$

The vector p was defined as following:

$$logit(\theta_i) = log(\frac{\theta_i}{1 - \theta_i}) = \alpha + \beta x_i$$

The code for the model is the following:

```
// Code for the Bioassay model
data {
  int N;
  vector[N] x; // doses for each observation
  int n[N]; // animals for each observation
  int y[N]; // deaths for each observation
}
parameters {
  vector[2] theta;
}
model {
  // model block creates the log density to be sampled
  vector[2] nu;
  matrix[2,2] Sigma;
  vector[N] p;
  nu[1] = 0;
  nu[2] = 10;
  Sigma[1,1] = 4;
```

```
  Sigma[1,2] = 10;
  Sigma[2,1] = 10;
  Sigma[2,2] = 100;
  p = inv_logit(theta[1] + theta[2] * x);

  theta ~ multi_normal(nu, Sigma);
  for (i in 1:N) {

    y[i] ~ binomial_logit(n[i], p[i]);
  }
  // y ~ binomial_logit(n, theta);
}
```

The data for calling the model is:

```
bioassay_data <- list(
  N = 4,
  x = c(bioassay$x),
  n = array(bioassay$n),
  y = array(bioassay$y)
)
```

Where N is the number of experiments that were carried out.

The way the function is called is:

```
fit1 <- stan(
  file = "bioassay_model.stan",  # Stan program
  data = bioassay_data,     # named list of data
  seed = SEED               # no progress shown
  )
```

The result is transformed into a data frame.

```
draws <- as.data.frame(fit1)
```

Exercise 2

R is a metric that involves the variance between two parameters, B and W:

- B measures the between sequence variance, meaning how different is each sequence from the others.
- W measures the within sequence variance, meaning how different parameters within the same sequence are.

The expression for R, according to BDA3 is:

$$`R_{metric} = \sqrt[2]{(\frac{`var^+(\phi|y)}{W})}$$

For a greater number of samples, the previous metric tends to 1, meaning that the variability of the two parameters previously described is really small. It is an analogous concept to the clusterization metrics, in which inter-cluster and intra-cluster similarity must be compared in order to decide if a clusterization was succesful. Nonetheless, in this exercise, the built-in function in Stan is used for computing the metric.

```
R_alpha <-Rhat(draws$`theta[1]`)
R_alpha
```

```
## [1] 1.000776
```

```
R_beta <- Rhat(draws$`theta[2]`)
R_beta
```

```
## [1] 0.9998027
```

In the given case, the metric for α is '$R_{metric,alpha} = 1.0007756 < 1.05$ and the one for β is '$R_{metric,beta} = 0.9998027 < 1.05$. In both cases, the value is close to 1 and, according to the Stan documentation, when having a lower metric than 1.05, then, the sample can be accepted, which is the current situation.

Exercise 3

The scatter plot of the draws is plotted in this exercise. The result is the expected one, very similar to the Figure 3.3b included in the book BDA3.

```
plot(draws$`theta[1]`, draws$`theta[2]`, xlab="alpha", ylab="beta", col="blue",
     main="Scatter plot alpha-beta")
```


Scatter plot alpha–beta
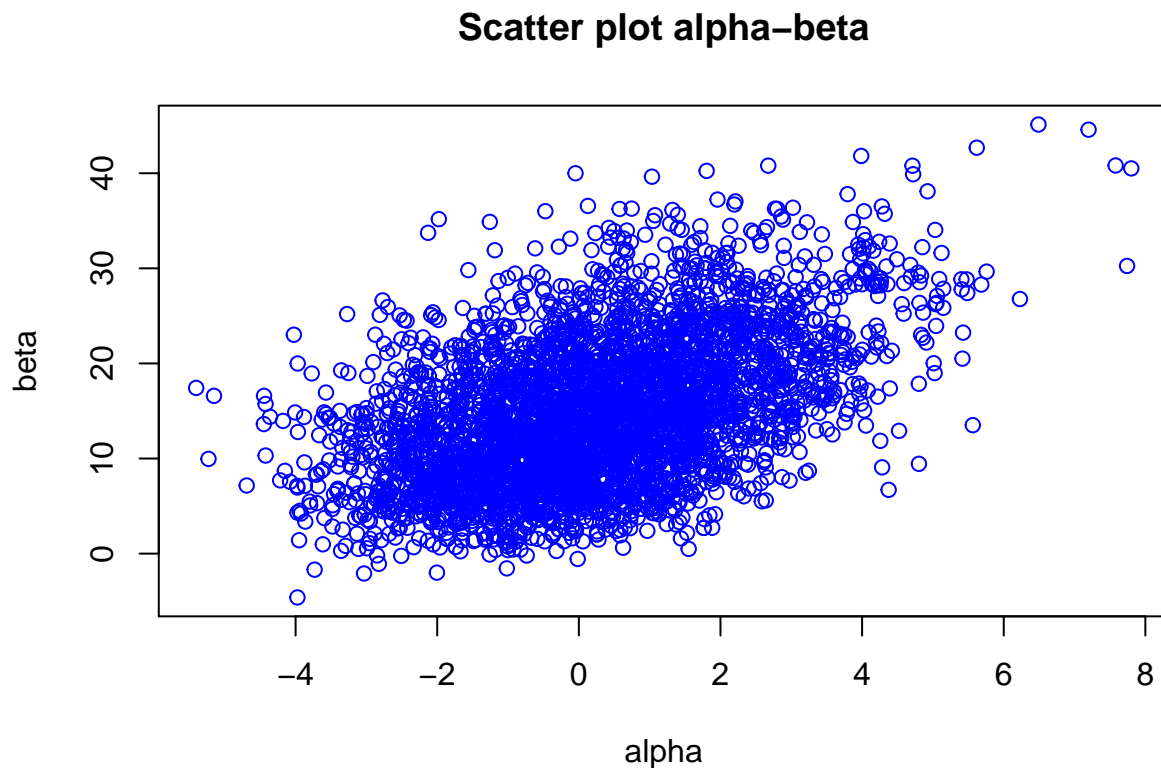
Exercise 4

- Operating system: Linux Ubuntu 18.04
- Programming environment used: R
- Interface used: RStan
- Installation problems: I had to add some libraries to my Linux environment.

- I installed it locally.
- I think there should be more documentation. The Stan examples provided are really useful, but the reference manual for the functions is not enoughly detailed in my opinion.