BDA - Assignment 4

Anonymous

8/3/2020

Q - Bioassay model and importance sampling

a)

Constructing a bivariate normal distribution as a prior distribution for (α,β):

$$p(\alpha, \beta \mid y) \propto p(\alpha, \beta)p(\alpha, \beta \mid y)$$

Given is:
$\alpha \sim \mathcal{N}(0, 2^2)$
$\beta \sim \mathcal{N}(10, 10^2)$
$corr(\alpha, \beta) = 0.5$

Specifying the values given above:
$k = 2$ and $\mu_\alpha = 0$, $\sigma_\alpha = 2$, $\mu_\beta = 10$, $\sigma_\beta = 10$ and $\rho = corr(\alpha, \beta) = 0.5$.

From this we can find μ:

$$\mu = \begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix} = \begin{pmatrix} 0 \\ 10 \end{pmatrix}$$

The covariance matrix is:

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} 4 & 10 \\ 10 & 100 \end{pmatrix}$$

.

b)

Implementing a function "p_log_prior" for computing logarithm of the density of the prior distribution of α β. We use the dmvnorm from the aaltobda package:

```
mu_alpha <- 0
sd_alpha <- 2
mu_beta <- 10
sd_beta <- 10
rho <- 0.5 #correlation
p_log_prior <- function(alpha, beta){
  x = cbind(alpha, beta)
  d <- dmvnorm(x, mean=c(mu_alpha, mu_beta),
               sigma = matrix(c(sd_alpha^2, rho*sd_alpha*sd_beta, rho*sd_alpha*sd_beta, sd_beta^2 ), nc
  return(log(d))
}
```

c)

Implementing a function "p_log_posterior" for computing logarithm of the density of the posterior for arbitrary values of α β and data x, y and n. Applied on dataset 'Bioassay'

The Bioassay data provided describes the outcome of toxicity tests in animals:

```
head(data)
```

```
##        x n y
## 1 -0.86 5 0
## 2 -0.30 5 1
## 3 -0.05 5 3
## 4  0.73 5 5
```

Where
n = number of animals for each observation x = dose level for each observation (log g/ml)
y = number of deaths for each observation

```
p_log_posterior <- function(alpha, beta, x=bioassay$x, y=bioassay$y, n=bioassay$n){
  p_log_likelihood <- bioassaylp(alpha, beta, x, y, n)
  post <- p_log_prior(alpha, beta) + p_log_likelihood
  return(post)
}
alpha <- 3
beta <- 9
p_log_posterior(alpha, beta, x = bioassay$x, y = bioassay$y, n = bioassay$n)
```

```
## [1] -15.78798
```

d)

Plotting the posterior density in a grid of points using "bioassay_posterior_density_plot" from aaltobda package:

```
# The plot is not really working for me with the functions
# Some bug I can't find
#bioassay_posterior_density_plot(alpha_limits = c(-4, 4),
#                                beta_limits = c(-10, 30))
```

e)

Two functions for 1) computing the log importance ratios.. and for 2) exponentiating the log importance ratios and normalizing them to sum to one

function 1:

```r
log_importance_weights <- function(alpha, beta){
  S <- length(alpha)
  w <- c()
  for(i in 1:S){
    w[i] <-  p_log_posterior(alpha[i], beta[i], x=bioassay$x, y=bioassay$y, n=bioassay$n) - p_log_prior
}
  return(w)
}
```

function 2:

We divide the importance weights with the sum of importance weights, for normalization:

```r
normalized_importance_weights <- function(alpha, beta){
  log_w <- log_importance_weights(alpha, beta)
  exp_w <- exp(log_w)
  return(exp_w/sum(exp_w))
}
```

f)

Sample draws of α and β from the prior distribution

```r
nr <- 2000
r <- rmvnorm(nr, mean=c(mu_alpha, mu_beta),
             sigma = matrix(c(sd_alpha^2, rho*sd_alpha*sd_beta,
                              rho*sd_alpha*sd_beta, sd_beta^2 ), ncol=2))
alpha <- r[, 1]
beta <- r[, 2]
posterior_mean <- function(alpha, beta){
  ab <- cbind(alpha, beta)
  colnames(ab) <- NULL
  post <- colSums( ab * normalized_importance_weights(alpha, beta) )
  return(post)
}
```

Posterior mean of α and β using importance sampling is:

```r
round(posterior_mean(alpha, beta), digit=3)
```

```
## [1]  0.954 10.480
```

g)

Using importance ratio to calculate the effective sample size:

```r
S_eff <- function(alpha, beta){
  s_eff <- 1/sum(normalized_importance_weights(alpha, beta)^2)
  return (s_eff)
}
```

The effective sample size, S_eff, is:

```
S_eff(alpha, beta)
```

```
## [1] 520.0979
```

h)

Using importance ratio to obtain a posterior sample..:

```
nr <- 2000
c_R <- rmvnorm(nr, mean=c(mu_alpha, mu_beta),
               sigma = matrix(c(sd_alpha^2, rho*sd_alpha*sd_beta,
                                rho*sd_alpha*sd_beta, sd_beta^2 ), ncol=2))
c_A <- c_R[, 1]
c_B <- c_R[, 2]
n_s <- 1000
s_indices <- sample(length(c_A), size = n_s, replace = FALSE,
                     prob = normalized_importance_weights(c_A, c_B) )
s_A <- c_A[s_indices[1:n_s]]
s_B <- c_B[s_indices[1:n_s]]
xl <- c(-4, 6)
yl <- c(-10, 31)
```

Scatterplot of the obtained posterior sample

```
ggplot(data = data.frame(s_A, s_B)) +
  geom_point(aes(s_A, s_B), color = 'blue', size = 0.3) +
  coord_cartesian(xlim = xl, ylim = yl) +
  labs(x = 'alpha', y = 'beta')
```
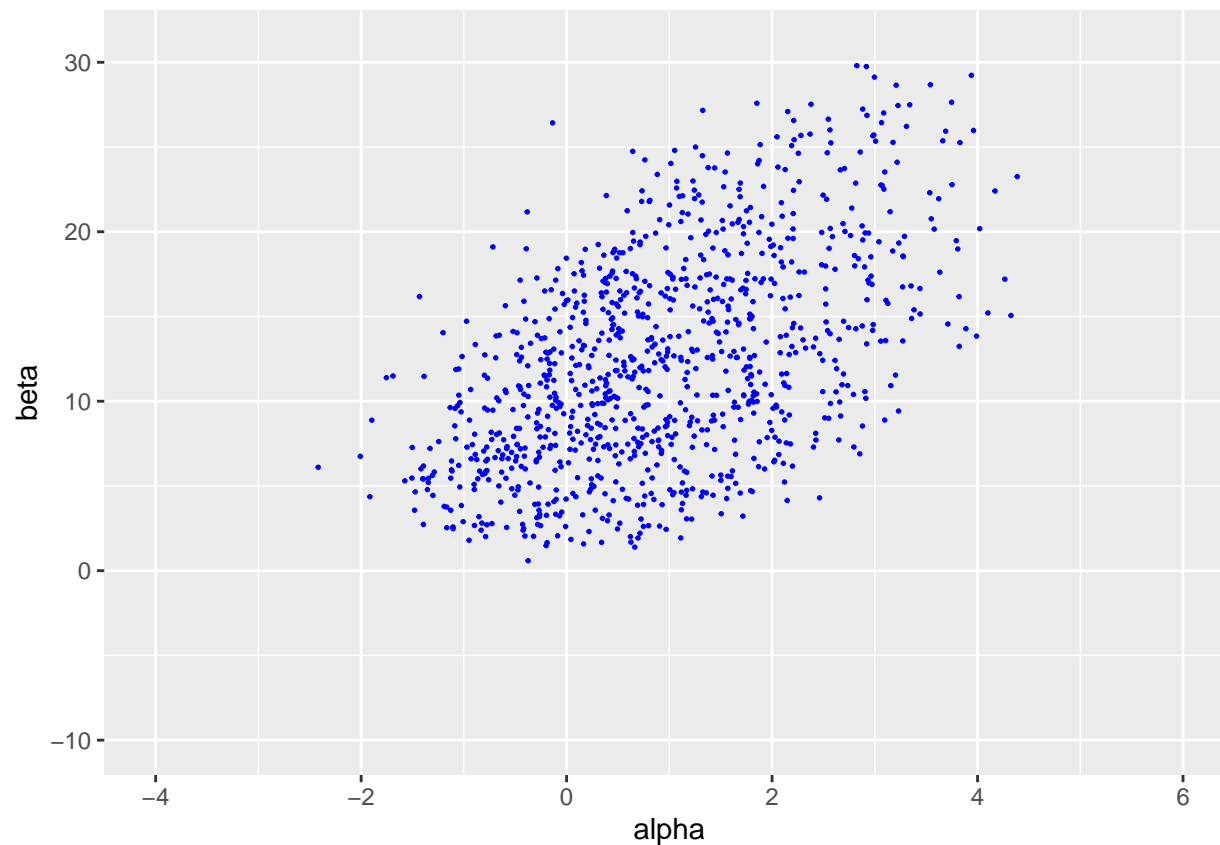
i)

Investigating if the drug is harmful by using the importance resampling without replacement:

```
bpi <- s_B > 0
s_ld50 <- -s_A[bpi]/s_B[bpi]
p_positive_beta <- length(bpi)/n_s
```

Harmful drug is given by:

$$p(\beta > 0|n, x, y)$$

```
p_positive_beta
```

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

j)

Using the posterior sample obtained via importance resampling, draw a histogram of the draws from the posterior distribution.

LD50 = Lethal dosage of $50\% = -\alpha/\beta$.
We calculate this for all of the posterior samplesand.

Histogram of LD50 is shown in the following figure.

5

```
ggplot() +
  geom_histogram(aes(s_ld50), binwidth = 0.05,
                 fill = 'darkgreen', color = 'black') +
  coord_cartesian(xlim = c(-0.8, 0.8)) +
  labs(x = 'LD50')
```