BDA - Assignment 4

Anonymous

```
library(ggplot2)
theme_set(theme_minimal())
library(aaltobda)
library(mvtnorm)
library(gridExtra)

data("bioassay")
data <- bioassay
```

Problem 1: Bioassay model and importance sampling

a)

$$p(\alpha,\beta) \sim \mathcal{N}\{\mu, \mathbf{\Sigma}\} = (2\pi)^{\frac{-k}{2}} det(\Sigma)^{\frac{-1}{2}} exp\Big(\frac{-1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\Big)$$

Where

$$k = 2$$

and

$$\mu_\alpha = 0$$

,

$$\sigma_\alpha = 2$$

,

$$\mu_\beta = 10$$

,

$$\sigma_\beta = 10$$

and

$$\rho = corr(\alpha,\beta) = 0.5$$

. Therefore,

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

and

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

.

b)

dmvnorm is used to calculate log prior for bivariate normal distribution.

```
mu_alpha <- 0
s_alpha <- 2
mu_beta <- 10
s_beta <- 10
rho <- 0.5


p_log_prior <- function(alpha, beta){
  x = cbind(alpha, beta)
  d <- dmvnorm(x, mean=c(mu_alpha, mu_beta),
               sigma = matrix(c(s_alpha^2, rho*s_alpha*s_beta, rho*s_alpha*s_beta, s_beta^2 ), ncol=2))
  return(log(d))
}
```

c)

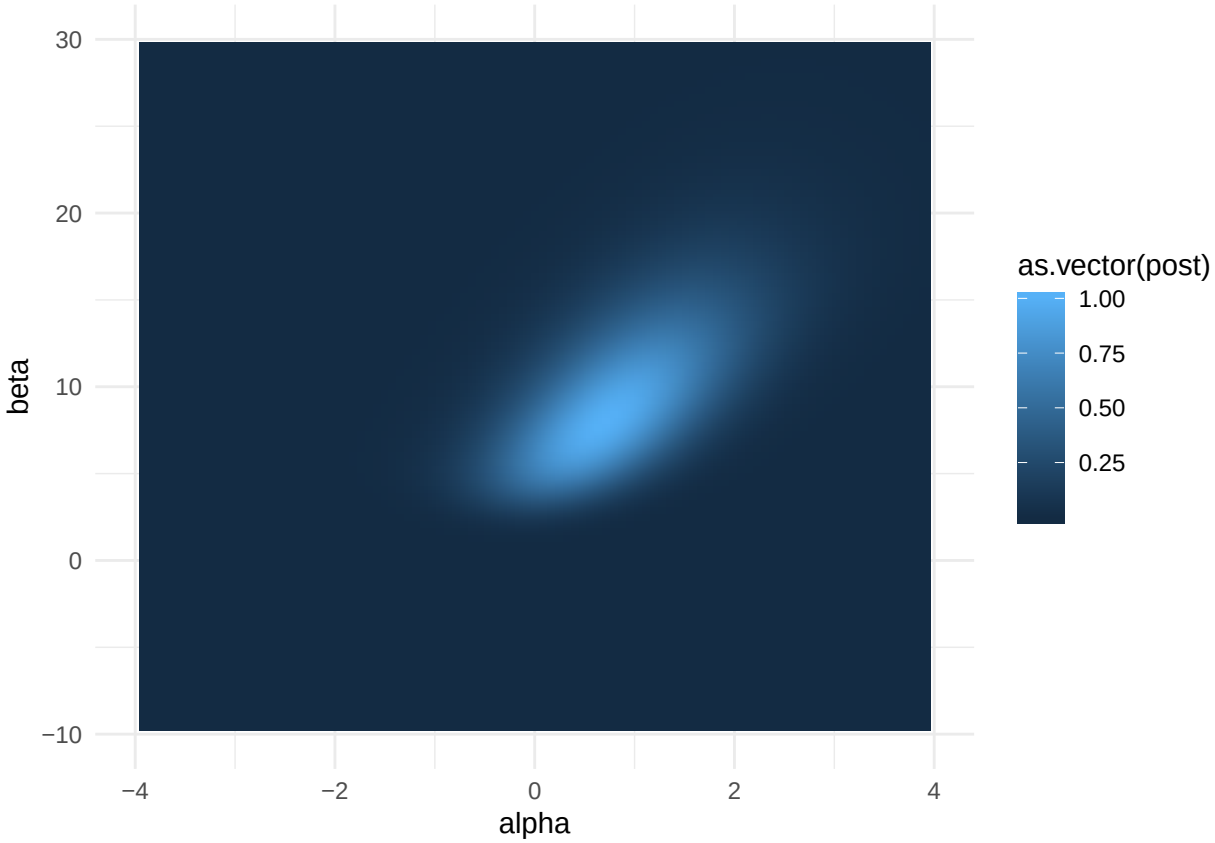
$$p(\alpha, \beta | y, n, x) \propto p(\alpha, \beta) \prod_{i=1}^{k} p(y_i | \alpha, \beta, n_i, x_i)$$

$$p(y_i | \alpha, \beta, n_i, x_i) = [\text{logit}^{-1}(\alpha + \beta x_i)]^{y_i} [1 - \text{logit}^{-1}(\alpha + \beta x_i)]^{n_i - y_i}$$

```
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)
}
```

d)

```
bioassay_posterior_density_plot(alpha_limits = c(-4, 4),
                                beta_limits = c(-10, 30))
```



e)

1)

```
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)
}
```

2)

For the computation of normalized importance weights, the non-logarithmic importance weights were needed, so the exponent of the log-importanceweights was taken first. Each (non-logarithmic)importance weight was normalizedby division with the sum of all non-logarithmic importance weights.

```
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)

```
nr <- 5000
r <- rmvnorm(nr, mean=c(mu_alpha, mu_beta),
             sigma = matrix(c(s_alpha^2, rho*s_alpha*s_beta,
                              rho*s_alpha*s_beta, s_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:

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

```
## [1]  0.953 10.388
```

g)

Using equation (10.4), the effective sample size will be calculated.

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

The effective sample size is:

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

```
## [1] 1391.643
```

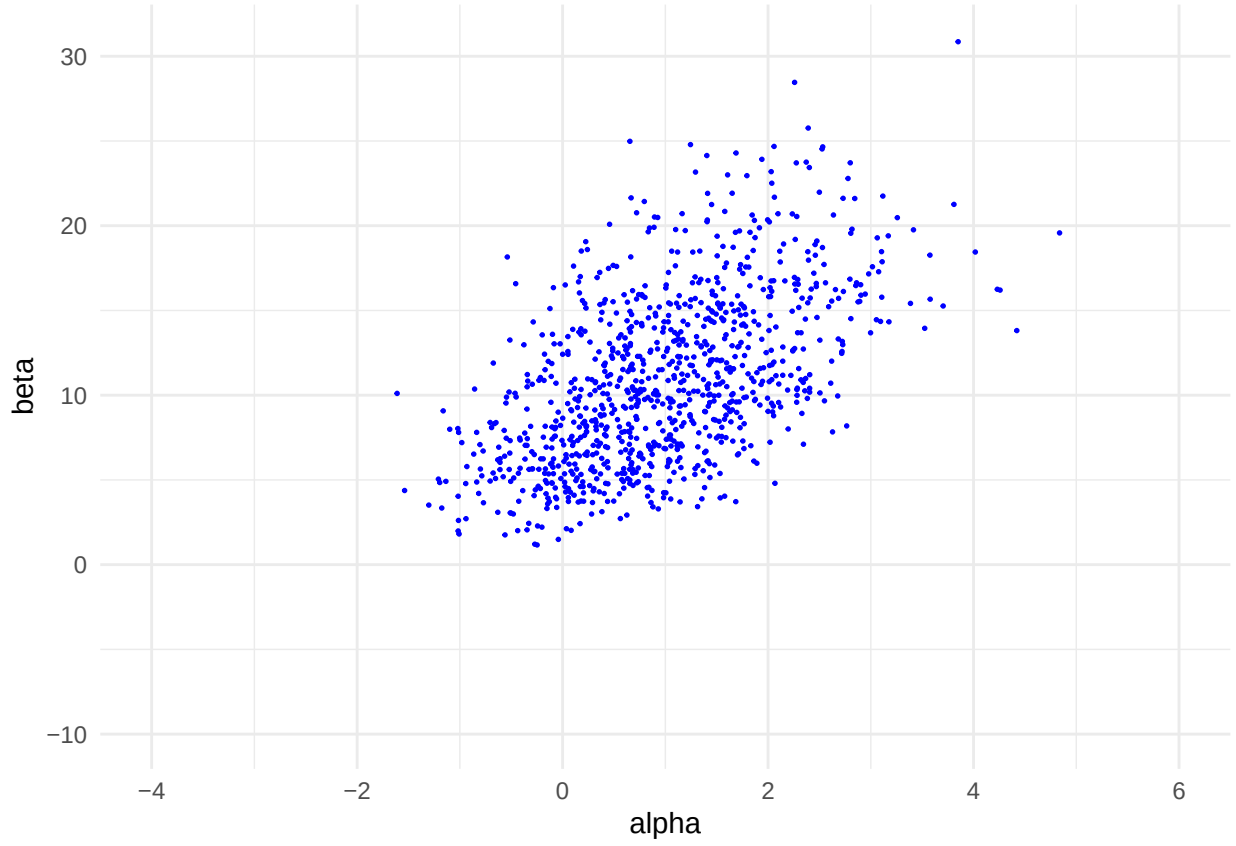
h)

```
nr <- 5000
cR <- rmvnorm(nr, mean=c(mu_alpha, mu_beta),
             sigma = matrix(c(s_alpha^2, rho*s_alpha*s_beta,
                              rho*s_alpha*s_beta, s_beta^2 ), ncol=2))
cA <- cR[, 1]
cB <- cR[, 2]
```

```r
nsamp <- 1000
samp_indices <- sample(length(cA), size = nsamp, replace = FALSE, prob = normalized_importance_weights(
```

```r
samp_A <- cA[samp_indices[1:nsamp]]
samp_B <- cB[samp_indices[1:nsamp]]
xl <- c(-4, 6)
yl <- c(-10, 31)

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



i)

```r
bpi <- samp_B > 0
samp_ld50 <- -samp_A[bpi]/samp_B[bpi]
p_positive_beta <- length(bpi)/nsamp
```

Drug is harmful with the probability of

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

```
p_positive_beta
```

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

The value is approximately hundred percent. If the sample size was increased some negative $\beta < 0$ outlier values could have been found.

j)

As described on page 77 of the course book, the lethal dosage of 50% (LD50) is given by $-\alpha/\beta$. So, we can simply calculate it for each of the posterior samples (for which $\beta > 0$, so all of them) and plot thehistogram

Histogram of LD50 is shown in the following figure.

```
ggplot() +
  geom_histogram(aes(samp_ld50), binwidth = 0.04,
                 fill = 'steelblue', color = 'black') +
  coord_cartesian(xlim = c(-0.8, 0.8)) +
  labs(x = 'LD50 = -alpha/beta')
```