

BDA - Assignment 3

Contents

Exercise 1..	1
a)	1
b)	1
c)	2
d)	2
e)	3
f)	4
g)	5
h)	5
i)	6
j)	6

Exercise 1.

a)

Provided that $\text{corr}(\alpha, \beta) = \frac{\text{Covariance}(\alpha, \beta)}{\sqrt{\text{Variance}(\alpha)\text{Variance}(\beta)}}$, the formula of the covariance matrix

$$\begin{bmatrix} \text{Cov}(\alpha, \alpha) & \text{Cov}(\alpha, \beta) \\ \text{Cov}(\beta, \alpha) & \text{Cov}(\beta, \beta) \end{bmatrix}$$

and also $\text{Cov}(\alpha, \alpha) = 2^2, \text{Cov}(\beta, \beta) = 10^2$, we can get **the mean and the covariance of the bivariate distribution**

$$\mu = [0, 10]$$

and

$$\Sigma = \begin{bmatrix} 4 & 10 \\ 10 & 100 \end{bmatrix}$$

.

b)

The code for the function that computes the logarithm of the density of the prior distribution in a) for arbitrary values of α and β goes as follows.

```
p_log_prior<-function(alpha, beta){
return(log(dmvnorm(c(alpha,beta),c(0,10),matrix(c(4,10,10,100),ncol=2,nrow=2))))
}
```

Also we can have a check here using the data from the instruction. The function should return a value very close to the correct answer.

```
# have a test
alpha <- 3
beta <- 9
cat("the correct answer is: -6.296435 \n")

## the correct answer is: -6.296435
cat("the calculated answer is: ",p_log_prior(alpha,beta),"\n")

## the calculated answer is: -6.296435
```

```
cat("the difference is",p_log_prior(alpha,beta)+6.296435,"\n")
```

```
## the difference is 2.959589e-08
```

c)

The code for the function that computes the logarithm of the density of the posterior for arbitrary values of α and β goes as follows.

```
data("bioassay")
p_log_posterior<-function(alpha, beta){
prior<-log(dmvnorm(c(alpha,beta),c(0,10),matrix(c(4,10,10,100),ncol=2,nrow=2)))
likelihood<-bioassaylp(alpha,beta,bioassay$x,bioassay$y,bioassay$n)
return(prior+likelihood)
}
```

Also we can have a check here using the data from the instruction. The function should return a value very close to the correct answer.

```
# have a test
alpha <- 3
beta <- 9
cat("the correct answer is: -15.78798 \n")
```

```
## the correct answer is: -15.78798
```

```
cat("the calculated answer is: ",p_log_posterior(alpha,beta),"\n")
```

```
## the calculated answer is: -15.78798
```

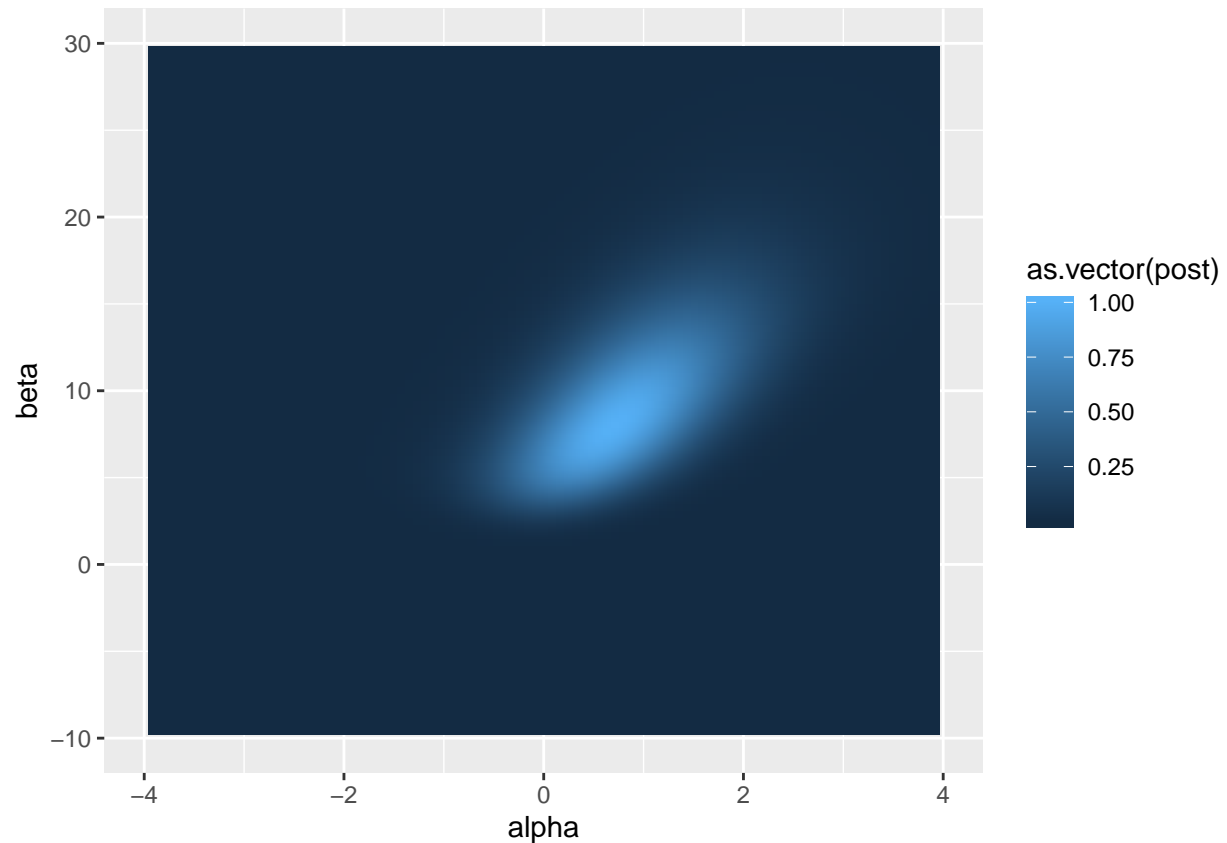
```
cat("the difference is",p_log_posterior(alpha,beta)+15.78798,"\n")
```

```
## the difference is 3.035914e-07
```

d)

The figure is shown below.

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



e)

The code for the functions that computes the importance ratios and the normalized weights for each sample goes as follows.

```
log_importance_weights<-function(alpha, beta){
  likelihoods=c()
  for (i in seq(1,length(alpha),length=length(alpha)))
  {
    likelihoods=c(likelihoods,bioassaylp(alpha[i],beta[i],bioassay$x,bioassay$y,bioassay$n))
  }
  return(likelihoods)
}

normalized_importance_weights<-function(alpha, beta)
{
  weights<-log_importance_weights(alpha,beta)
  weights<-exp(weights)
  nor_weights<-weights/sum(weights)
  return(nor_weights)
}
```

Also we can have a check here using the data from the instruction. The function should return a value very close to the correct answer.

```
#have a test
alpha <- c(1.896, -3.6, 0.374, 0.964, -3.123, -1.581)
```

```

beta <- c(24.76, 20.04, 6.15, 18.65, 8.16, 17.4)
cat("Here's the test case:")

## Here's the test case:
cat("the correct answer for the log weights is: -8.95 -23.47 -6.02 -8.13 -16.61 -14.57 \n")

## the correct answer for the log weights is: -8.95 -23.47 -6.02 -8.13 -16.61 -14.57
cat("the calculated answer for the log weights is: ",round(log_importance_weights(alpha,beta),2),'\n')

## the calculated answer for the log weights is: -8.95 -23.47 -6.02 -8.13 -16.61 -14.57
cat("the difference is ",mean(log_importance_weights(alpha,beta)-c(-8.95,-23.47,-6.02,-8.13,-16.61,-14.57)),'\n')

## the difference is -0.0006765508
cat("the correct answer for the normalized weights is: 0.045 0.000 0.852 0.103 0.000 0.000 \n")

## the correct answer for the normalized weights is: 0.045 0.000 0.852 0.103 0.000 0.000
cat("the calculated answer for the normalized weights is: ",round(normalized_importance_weights(alpha,beta),2),'\n')

## the calculated answer for the normalized weights is: 0.045 0 0.852 0.103 0 0
cat("the difference is ",mean(normalized_importance_weights(alpha,beta)-c(0.045,0.000,0.852,0.103,0.000,0.000)),'\n')

## the difference is -2.100573e-18
cat("test case ends")

```

test case ends

Now that all the functions defined above work correctly. Now we can forget about the test case and draw some real samples from the prior.

```

samples<-rmvnorm(10000,c(0,10),matrix(c(4,10,10,100),ncol=2,nrow=2)) #draw some samples
alpha_samples<-samples[,1]
beta_samples<-samples[,2]
cat("real case:10000 samples drawn")

```

real case:10000 samples drawn

f)

```

posterior_mean<-function(alpha, beta)
{
  weights<-normalized_importance_weights(alpha,beta)
  alpha_mean<-sum(weights*alpha)/sum(weights)
  beta_mean<-sum(weights*beta)/sum(weights)
  return(c(alpha_mean,beta_mean))
}

cat("the posterior mean using importance sampling and draws from e) is: ",
    posterior_mean(alpha_samples,beta_samples))

```

the posterior mean using importance sampling and draws from e) is: 0.9721849 10.44593

g)

```
S_eff<-function(alpha, beta)
{
weights<-normalized_importance_weights(alpha,beta)
return(1/sum(weights^2))
}
cat("the effective sample size is: ",S_eff(alpha_samples,beta_samples))
```

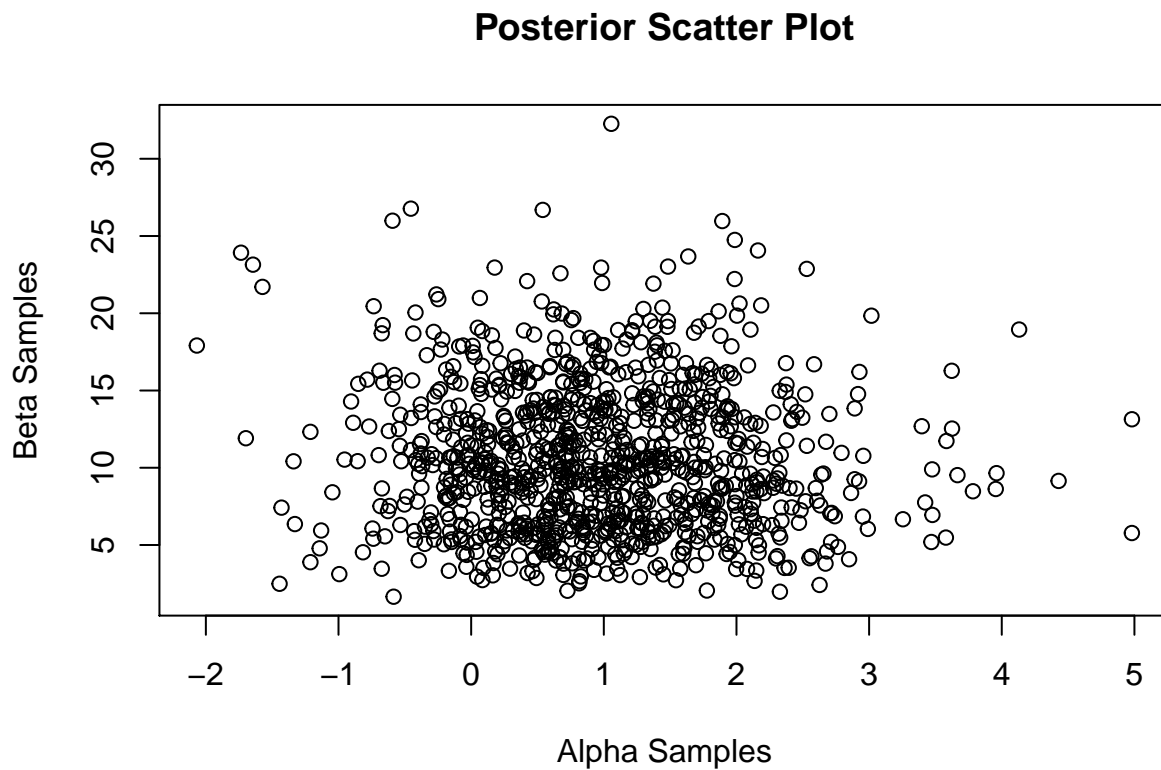
```
## the effective sample size is: 2716.087
```

h)

The scatterplot of the obtained posterior sample is shown below.

```
samples<-rmvnorm(10000,c(0,10),matrix(c(4,10,10,100),ncol=2,nrow=2))
alpha<-samples[,1]
beta<-samples[,2]
weights<-normalized_importance_weights(alpha,beta)
alpha_resample<-sample(alpha, 1000, replace = FALSE, prob = weights)
beta_resample<-sample(beta, 1000, replace = FALSE, prob = weights)

plot(alpha_resample,beta_resample,'p', xlab="Alpha Samples", ylab="Beta Samples",
      main="Posterior Scatter Plot")
```



i)

```
cat("the estimate for p(beta > 0 | x,n,y) (probability that the drug is harmful) is:",length(which(beta  
## the estimate for p(beta > 0 | x,n,y) (probability that the drug is harmful) is: 1
```

j)

The histogram of the draws from the posterior distribution of the LD50 conditional on Beta > 0 is shown below

```
#delete points with beta<0  
alpha_resample_conditional<-c()  
beta_resample_conditional<-c()  
for (i in seq(1,1000,len=1000))  
{  
  if (beta_resample[i]>0)  
  {  
    alpha_resample_conditional<-c(alpha_resample_conditional,alpha_resample[i])  
    beta_resample_conditional<-c(beta_resample_conditional,beta_resample[i])  
  }  
}  
hist(-alpha_resample_conditional/beta_resample_conditional,xlab="LD50",main="histogram of LD50 condition
```

