

# STAT7016 Assignment 2

Rui Qiu

2017-08-28

## Settings for this assignment

```
set.seed(7016)
library(LearnBayes)
library(ggplot2)
library(RColorBrewer)
source("hdr_2d.r")
```

## Problem 1

Suppose  $y$  has a binomial distribution with parameters  $n$  and  $p$ , and we are interested in the log-odds value  $\theta = \log(p/(1-p))$ . Our prior for  $\theta$  is that  $\theta \sim N(\mu, \sigma^2)$ .

Suppose we are interested in learning about the probability that a special coin lands heads when tossed. *A priori* we believe that the coin is fair, so we assign  $\theta$  a  $N(\mu = 0, \sigma = 0.25)$  prior. We toss the coin  $n = 5$  times and obtain  $y = 5$  heads.

- (a) What is the posterior distribution of  $\theta$ ? (up to a proportionality constant)

## Solution

Our prior  $\log \frac{p}{1-p} = \theta$  follows a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . If we rephrase it, we can have

$$\begin{aligned}\theta &= \theta - \log(1 + e^\theta) - \log 1 + \log(1 + e^\theta) \\ &= \theta - \log(1 + e^\theta) - \log \frac{1}{1 + e^\theta} \\ &= \log \frac{e^\theta}{1 + e^\theta} - \log(1 - \frac{e^\theta}{1 + e^\theta}) \\ &= \frac{e^\theta / (1 + e^\theta)}{1 - e^\theta / (1 + e^\theta)}\end{aligned}$$

Therefore,

$$p = \frac{\exp(\theta)}{1 + \exp(\theta)}$$

The likelihood function of  $\theta$  is proportional to

$$\begin{aligned}
L(\theta) &\propto \prod_{i=1}^n p^{y_i} (1-p)^{n-y_i} \text{ where } y_i = \{1, 2, \dots, n\} \\
&\propto \prod_{i=1}^n \left( \frac{\exp \theta}{1 + \exp \theta} \right)^{y_i} \left( \frac{1}{1 + \exp \theta} \right)^{n-y_i} \\
&\propto \frac{\exp(\theta y)}{(1 + \exp \theta)^n}
\end{aligned}$$

Then the posterior function is

$$\begin{aligned}
p(\theta | y) &\propto p(\theta) \times L(\theta) \\
&\propto \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(\theta - \mu)^2}{2\sigma^2} \right] \cdot \frac{\exp(\theta y)}{(1 + \exp(\theta))^n} \\
&\propto \frac{\exp(\theta y)}{(1 + \exp(\theta))^n} \exp \left[ -\frac{(\theta - \mu)^2}{2\sigma^2} \right]
\end{aligned}$$

- (b) For a univariate roughly symmetric posterior distribution  $p(\theta | y)$ , it can be convenient to approximate it by a normal distribution. A Taylor series expansion of  $\log p(\theta | y)$  centred at the posterior mode  $\hat{\theta}$  gives

$$\log p(\theta | y) = \log p(\hat{\theta} | y) + \frac{1}{2}(\theta - \hat{\theta})^T \left[ \frac{d^2}{d\theta^2} \log p(\theta | y) \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \dots$$

where the linear term in the expansion is zero because the log-posterior density has zero derivative at its mode. The remainder terms of higher order fade in importance relative to the quadratic term when  $\theta$  is close to  $\hat{\theta}$  and  $n$  is large. From the above expansion, we obtain the following normal approximation:

$$p(\theta | y) \approx N(\hat{\theta}, [I(\hat{\theta})]^{-1})$$

where  $I(\theta)$  is the observed information ( $I(\theta) = -\frac{d^2}{d\theta^2} \log p(\theta | y)$ ). The function `laplace` in the `LearnBayes` package in R finds the posterior mode using the function `optim`. The output of `laplace` is a list. The component `mode` gives the value of the posterior mode  $\hat{\theta}$  and the component `var` is the associated variance estimate.

Use the `laplace` function to find a normal approximation to the posterior density of  $\theta$  and compute the posterior probability that the coin is biased towards heads. Note: as inputs into `laplace`, you will need to write a function in R to evaluate the log posterior density and provide an initial guess for the value of  $\hat{\theta}$ .

## Solution

The log posterior density is written as below:

```

# log posterior
log.p <- function(theta,par){
  y <- par[1];
  n <- par[2];
  mu <- par[3];
  sigma <- par[4]
  res <- y*theta-n*log(1+exp(theta))-(theta-mu)^2/(2*sigma^2)
  return(res)
}

```

Our initial guess about  $p = 0.75$ , then correspondingly the initial guess of  $\theta = \log(p/(1-p)) \approx 1.099$ . Then we plug it in the `laplace` function.

```
post.mode <- laplace(log.p,1.099,c(5,5,0,0.25))
```

```
post.mode$mode
```

```
## [1] 0.144995
```

```
post.mode$var
```

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

```
## [1,] 0.05799303
```

So the normal distribution we used to approximate is  $N(\mu = 0.144995, \sigma^2 = 0.05799303)$ .

```
(prob.bias <- 1-pnorm(0,post.mode$mode,sqrt(post.mode$var)))
```

```
## [1] 0.7264446
```

The posterior probability that the coin is biased toward heads is rather high, which is 0.726. If we consider the known 5 tosses are all heads, this is not surprising.

- (c) Using the prior density as a proposal density, design a rejection algorithm for sampling from the posterior distribution. Using simulated draws from your algorithm, approximate the probability that the coin is biased towards heads.

## Solution

Here we simulate 10000 times.

```
# rejection algorithm
reject.sample <- function(nums=10000) {
  theta <- rnorm(nums,0,0.25)
  prob <- exp(5*theta)/(1+exp(theta))^5
  return(theta[runif(nums) < prob])
}
```

```
thetas <- reject.sample()
sum(thetas > 0)/length(thetas)
```

```
## [1] 0.7582418
```

```
mean(reject.sample() > 0) # slightly different answer but generally the same
```

```
## [1] 0.7589041
```

The probability that the coin is biased based on rejection algorithm is slightly higher than previous part. This probability is equal to 0.759.

## Problem 2

After a posterior analysis on data from a population of squash plants, it was determined that the total vegetable weight of a given plant could be modelled with the following distribution:

$$p(y \mid \theta, \sigma^2) = 0.31 \times \text{dnorm}(y, \theta, \sigma) + 0.46 \times \text{dnorm}(y, 2\theta, 2\sigma) + 0.23 \times \text{dnorm}(y, 3\theta, 3\sigma)$$

where  $\text{dnorm}(y, \theta, \sigma)$  refers to the density of a normal distribution with mean  $\theta$  and standard deviation  $\sigma$ . You are told that the posterior distributions of the parameters have been calculated as  $1/\sigma^2 \mid \mathbf{y} \sim \text{gamma}(10, 2.5)$ , and  $\theta \mid \sigma^2, \mathbf{y} \sim \text{normal}(4.1, \sigma^2/20)$ .

- (a) Sample at least 5,000  $y$  values from the posterior predictive distribution.
- (b) Form a 75% quantile-based confidence interval for a new value of  $Y$ .
- (c) Form a 75% HPD region for a new  $Y$  as follows:
  - Compute estimates of the posterior density of  $Y$  using the `density` command in R, and then normalise the density values so that they sum to 1.
  - Sort these discrete probabilities in decreasing order.
  - Find the first probability value such that the cumulative sum of the sorted values exceeds 0.75. Your HPD region includes all values of  $y$  which have a discretised probability greater than this cutoff. Describe your HPD region, and compare it to your quantile based region.
- (d) Can you think of a physical justification for the mixture sampling distribution of  $Y$ ?

## Solution

### part (a)

We use Monte-Carlo sampling to sample inverse gamma  $\frac{1}{\sigma^2}$  and  $\theta$  separately then apply the given posterior formula.

Before starting simulation, we have to set up a range of possible values of squash plant weight. Due to simplicity, we assume the upper bound is 20kg while the lower bound is 0 (excluded). Of course we don't want the minimum weight to be zero, which is meaningless.

The simulated 5000  $y$  values are stored in the variable `sample.y`.

```
sqsh.max <- 20
sims <- 100000
sqsh.min <- sqsh.max/sims

yrange <- seq(sqsh.min,sqsh.max,sqsh.min) # range of all possible y values
yden <- c() # generate densities of y's
for (i in 1:sims) {
  inv.sig2 <- rgamma(1,10,2.5)
  sigma <- sqrt(1/inv.sig2)
  theta <- rnorm(1,4.1,sqrt(1/inv.sig2/20))
  yden[i] <- 0.31*dnorm(yrange[i],mean=theta,sd=sigma)+
    0.46*dnorm(yrange[i],mean=2*theta,sd=2*sigma)+
    0.23*dnorm(yrange[i],mean=3*theta,sd=3*sigma)
}

# (a)
index <- sample(1:sims, 5000, prob=yden) # randomly sample 5000 indices
```

```
ys <- yrange[index] # sampled y's
ysden <- yden[index]
```

part (b)

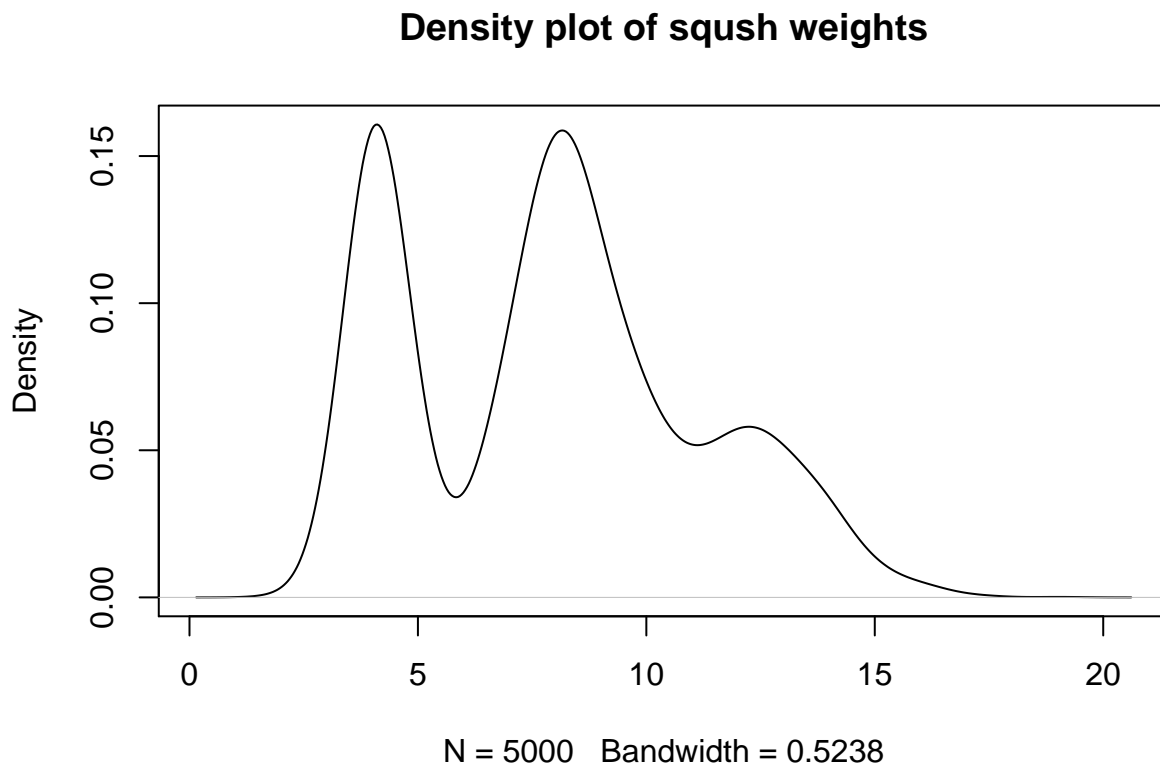
The 75% quantile-based confidence interval for a new new value of  $Y$  is (3.989, 12.193).

```
# (b)
quantile(ys, c(0.125,0.875))
```

```
##      12.5%      87.5%
##  3.989025 12.193475
```

part (c)

```
# (c)
plot(density(ys),type='l',main='Density plot of sqush weights')
```



```
md <- ys[ysden==max(ysden)] # mode
ysden <- ysden/sum(ysden)
ysden <- sort(ysden,decreasing=T)
ct<-min(ysden[cumsum(ysden)< 0.75])
list(hpdr=range(ysden[ysden>=ct]),mode=md)
```

```
## $hpdr
## [1] 0.0001954401 0.0005702946
##
## $mode
## [1] 4.1494
```

According to the plot above, the mixture model has 3 peaks (though there is only one mode). This is because the model consists of three different normal distribution components with different weights.

Due to lack of understanding of helper functions in the file `hdr_2d.r`, I didn't successfully draw the density plot of this mixture model with 75% HPD region. However, I believe the HPD region should contain multiple intervals instead of single interval like quantile-based interval in part (b). Moreover, the HPD region should be narrower than quantile-based interval, as it is more precise.

#### **part (d)**

The components of this mixture model could be described as distinct types <sup>1</sup> of squash plants. And the weight of each included normal distribution could represent the ratio of such type among the whole set of squash plants.

---

<sup>1</sup> *Common Types of Squash*, <https://www.realsimple.com/food-recipes/shopping-storing/food/common-types-winter-squash>

### Problem 3

It is a common problem for measurements to be observed in rounded form. For a simple example, suppose we weigh an object five times and measure weights, rounded to the nearest kilogram, of 10, 10, 11, 10, 9. Assume the unrounded measurements are normally distributed with a noninformative prior distribution on the mean  $\mu$  and variance  $\sigma^2$ .

- (a) Give the posterior distribution for  $(\mu, \sigma^2)$  obtained by pretending that the observations are exact unrounded measurements.
- (b) Give the correct posterior distribution (up to a proportionality constant) for  $(\mu, \sigma^2)$  treating the measurements as rounded.
- (c) How do the incorrect and correct posterior distributions differ? Compare means, variances and contour plots.
- (d) Let  $z = (z_1, \dots, z_5)$  be the original, unrounded measurements corresponding to the given observations above. Draw simulations from the posterior distribution of  $z$ . Compute the posterior mean of  $(z_1 - z_3)^2$ .

### Solution

#### part (a)

If our prior is a uniform prior, then the mean of normal distribution with known variance follows

$$\mu \mid \sigma^2, y \sim N(\bar{y}, \frac{\sigma^2}{n}).$$

Then the marginal posterior distribution is

$$\begin{aligned} p(\sigma^2 \mid y) &\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2]\right) d\mu \\ &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} (n-1)s^2\right) \sqrt{2\pi \frac{\sigma^2}{n}} \\ &\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \\ \sigma^2 \mid y &\sim \text{Inv-}\chi^2(n-1, s^2). \end{aligned}$$

```
y <- c(10,10,11,10,9)
(ybar <- mean(y)) # sample mean
```

```
## [1] 10
```

```
(s2 <- var(y)) # sample variance
```

```
## [1] 0.5
```

So  $n = 5$ ,  $\bar{y} = 10$ ,  $s^2 = 0.5$ , thus  $\mu \mid \sigma^2, y \sim N(10, \sigma^2/5)$ ,  $\sigma^2 \mid y \sim \text{Inv-}\chi^2(4, 0.5)$ .

#### part (b)

Also we know the prior is non-informative such that  $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$ .

With this improper prior in mind, we know the joint posterior distribution is proportional to the product of likelihood function and the factor  $\frac{1}{\sigma^2}$ .

$$\begin{aligned}
p(\mu, \sigma^2 | y) &\propto \sigma^{-n-2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right) \\
&\propto \sigma^{-n-2} \exp \left( -\frac{1}{2\sigma^2} \left[ \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2 \right] \right) \\
&\propto \sigma^{-n-2} \exp \left( -\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \mu)^2] \right)
\end{aligned}$$

where  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$  which is the sample variance of  $y_i$ 's.

### part (c)

In order to compare two posterior distributions side by side, we would like to draw contour plots for both. We also changed the scale of  $\sigma^2$  to  $\log(\sigma)$  so that contours can be easily compared in the y-axis direction.

```

post.wo <- function(mu,sd,y) {
  x <- 0
  for (i in 1:length(y)) {
    x <- x + log(dnorm(y[i],mu,sd))
  }
  return(x)
}

post.wi <- function(mu,sd,y) {
  x <- 0
  for (i in 1:length(y)) {
    x <- x + log(pnorm(y[i]+0.5,mu,sd)-pnorm(y[i]-0.5,mu,sd))
  }
  return(x)
}

summary2 <- function(x) {
  return(c(mean(x),sqrt(var(x)),quantile(x,c(0.025,0.25,0.5,0.75,0.975))))
}

nums <- 10000
n <- length(y)
# set up the range of mu
mu.range <- seq(5,15,length.out=250)
# set up the range of log(sigma)
logsigma.range <- seq(-2,2,length.out=250)
# so basically we are setting up a 250-by-250 grid approximation
contours <- c(0.0001,0.001,0.01,seq(0.05,0.95,0.05))

log.dsty.wo <- outer(mu.range, exp(logsigma.range), post.wo, y)
dsty.wo <- exp(log.dsty.wo - max(log.dsty.wo))

sd.wo <- sqrt((n-1)*s2/rchisq(nums,4))
mu.wo <- rnorm(nums,ybar,sd.wo/sqrt(n))

log.dsty.wi <- outer(mu.range, exp(logsigma.range), post.wi, y)
dsty.wi <- exp(log.dsty.wi - max(log.dsty.wi))

```



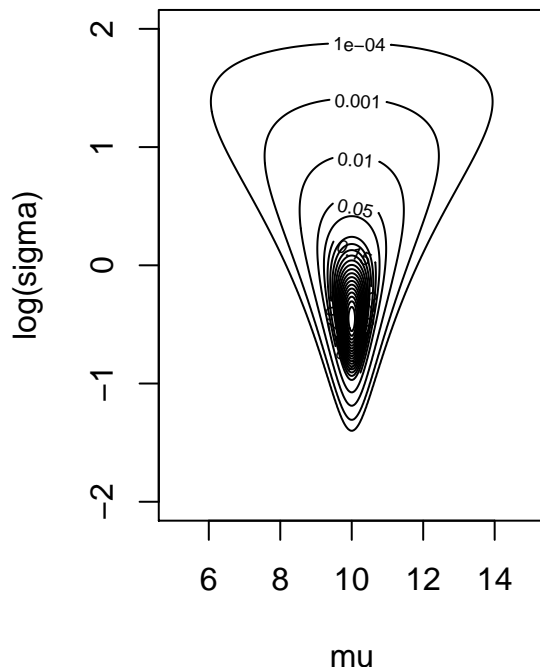
```

mu.dens <- apply(dsty.wi,1,sum)
mu.i <- sample(1:length(mu.range), nums, replace=T, prob=mu.dens) # indices
mu.wi <- mu.range[mu.i]
sd.wi <- rep(0, nums)
for (i in 1:nums) {
  sd.wi[i] <- exp(sample(logsigma.range, 1, prob=dsty.wi[mu.i[i],]))
}

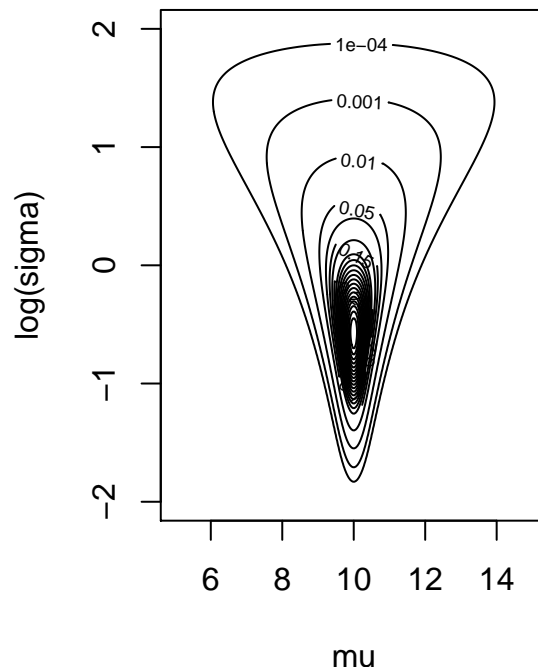
par(mfrow=c(1,2))
contour(mu.range,logsigma.range,dsty.wo,levels=contours,
        xlab="mu",ylab="log(sigma)",main="Post.density W/O rounding")
contour(mu.range,logsigma.range,dsty.wi,levels=contours,
        xlab="mu",ylab="log(sigma)",main="Post.density W rounding")

```

**Post.density W/O rounding**



**Post.density W rounding**



```

# Additionally we produce a table for numerical comparison
summary.table <- rbind(summary2(mu.wo),summary2(mu.wi),
                        summary2(mu.wo)-summary2(mu.wi),
                        summary2(sd.wo),summary2(sd.wi),
                        summary2(sd.wo)-summary2(sd.wi))
colnames(summary.table)[1:2] <- c("mean", "sd")
rownames(summary.table) <- c("mu w/o rounding", "mu wi rounding",
                             "mu diff",
                             "sd w/o rounding", "sd wi rounding",
                             "sd diff")

summary.table

```

##		mean	sd	2.5%	25%
##	mu w/o rounding	10.000724637	0.442112707	9.133003826	9.76380393
##	mu wi rounding	9.995032129	0.430510830	9.136546185	9.77911647

```
## mu diff          0.005692509 0.011601877 -0.003542359 -0.01531254
## sd w/o rounding  0.891337620 0.451486155  0.431670794  0.61072243
## sd wi rounding   0.797482829 0.441692610  0.322216976  0.52173139
## sd diff          0.093854791 0.009793544  0.109453818  0.08899104
##                  50%          75%          97.5%
## mu w/o rounding 10.00401240 10.23532688 10.88752821
## mu wi rounding   9.97991968 10.22088353 10.82329317
## mu diff          0.02409273  0.01444335  0.06423503
## sd w/o rounding  0.77808082  1.02507217  2.00785979
## sd wi rounding   0.68556632  0.93026187  1.91669512
## sd diff          0.09251450  0.09481030  0.09116467
```

There is no big difference between our two contour plots, except when we take rounding into accounting, the  $\log(\sigma)$  is possible to reach some lower values than the one without rounding.

The output with some summary statistics also confirms our claim, only the 2.5% quantile for  $\log(\sigma)$  has distinguishable difference.

#### part (d)

We are interested in the conditional probability that  $z_i$  falls into the rounding range to  $y_i$ , i.e.  $[y_i - 0.5, y_i + 0.5)$ .

```
z <- matrix(0,10000,5)
for (i in 1:5) {
  lwbd <- pnorm(y[i]-0.5,ybar,sqrt(s2))
  upbd <- pnorm(y[i]+0.5,ybar,sqrt(s2))
  z[,i] <- qnorm(lwbd + runif(10000)*(upbd-lwbd),ybar,sqrt(s2))
}
mean((z[,1]-z[,3])^2)
```

```
## [1] 0.8660221
```

The posterior mean of  $(z_1 - z_3)^2$  after 10000 simulations is around 0.866.

## Problem 4

In the development of a new drug, forty animals were tested, ten at each of four dose levels. The data are provided in the table below:

Dose $x_i$ (log g/ml)	Number of animals, $n_i$	Number of deaths, $y_i$
-0.86	10	0
-0.30	10	3
-0.05	10	7
0.73	10	8

we assume the following sampling model

$$y_i \mid \theta_i \sim \text{Bin}(n_i, \theta_i)$$

where  $\theta_i$  is the probability of death for animals given dose  $x_i$ . To model the dose-response relationship, we use the logistic model:

$$\text{logit}(\theta_i) = \alpha + \beta x_i$$

where  $\text{logit}(\theta_i) = \log(\theta_i/(1 - \theta_i))$ . Assume a uniform prior distribution for  $(\alpha, \beta)$ ; that is,  $p(\alpha, \beta) \propto 1$

- Assume a normal approximation to the true posterior density  $p(\alpha, \beta \mid y)$ , and sample  $S = 10000$  draws,  $(\alpha^{(1)}, \beta^{(1)}), \dots, (\alpha^{(S)}, \beta^{(S)})$  from the approximate distribution. Call this approximate distribution  $g(\alpha^{(s)}, \beta^{(s)})$ . Then resample  $k = 1000$  samples from the set  $\{(\alpha^{(1)}, \beta^{(1)}), \dots, (\alpha^{(S)}, \beta^{(S)})\}$  without replacement, where the probability of sampling each  $(\alpha^{(s)}, \beta^{(s)})$  is proportional to the importance ratio  $\frac{q(\alpha^{(s)}, \beta^{(s)} \mid y)}{g(\alpha^{(s)}, \beta^{(s)})}$  (where  $q(\alpha, \beta \mid y)$  is the unnormalised density of  $p(\alpha, \beta \mid y)$ ). Draw a scatter plot of your posterior draws.
- Repeat part (a) but sample with replacement. Discuss how your results differ, if at all.

## Solution

### part (a)

We first try to fit the data into a generalized linear model and get some basic idea of the approximate distributions of  $\alpha$  and  $\beta$ .

```
x <- c(-0.86, -0.3, -0.05, 0.73)
n <- rep(10, 4)
y <- c(0, 3, 7, 8)
res <- cbind(y, n-y)
res.lm <- glm(res ~ x, family=binomial)
summary(res.lm)

##
## Call:
## glm(formula = res ~ x, family = binomial)
##
## Deviance Residuals:
##      1      2      3      4
## -1.3587 -0.1492  1.3369 -0.9258
##
```

```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.1074      0.4120   0.261  0.79434
## x           2.8414      0.9200   3.089  0.00201 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 20.6085  on 3  degrees of freedom
## Residual deviance:  4.5127  on 2  degrees of freedom
## AIC: 16.192
##
## Number of Fisher Scoring iterations: 4
```

So  $\alpha \sim N(\mu = 0.1074, \sigma = 0.4120)$ ,  $\beta \sim N(\mu = 2.8414, \sigma = 0.9200)$ .

The sampling distribution for each dose value in terms of  $\alpha$  and  $\beta$  is

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

The joint posterior density of  $\alpha$  and  $\beta$  is

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

In the process of defining  $g(\alpha^{(s)}, \beta^{(s)})$ , we used pre-defined function `logisticpost` from `LearnBayes` package by Jim Albert.

```
mu.a <- summary(res.lm)$coef[1,1]
sd.a <- summary(res.lm)$coef[1,2]
mu.b <- summary(res.lm)$coef[2,1]
sd.b <- summary(res.lm)$coef[2,2]
S <- 10000
k <- 1000

g <- function(a, b) {
  par <- cbind(a,b)
  newdata <- cbind(x,n,y)
  post <- c()
  for (i in 1:nrow(par)) {
    post[i] <- logisticpost(par[i,], newdata)
  }
  return(exp(post))
}

q <- function(a, b) {
```

```

product <- 1
for (i in 1:4) {
  product <- product*
    (exp(a+b*x[i])/(1+exp(a+b*x[i])))^y[i]*
    (1/(1+exp(a+b*x[i])))^(n[i]-y[i])
}
return(product)
}

iratio <- function(a, b){
  return(q(a,b)/g(a,b))
}

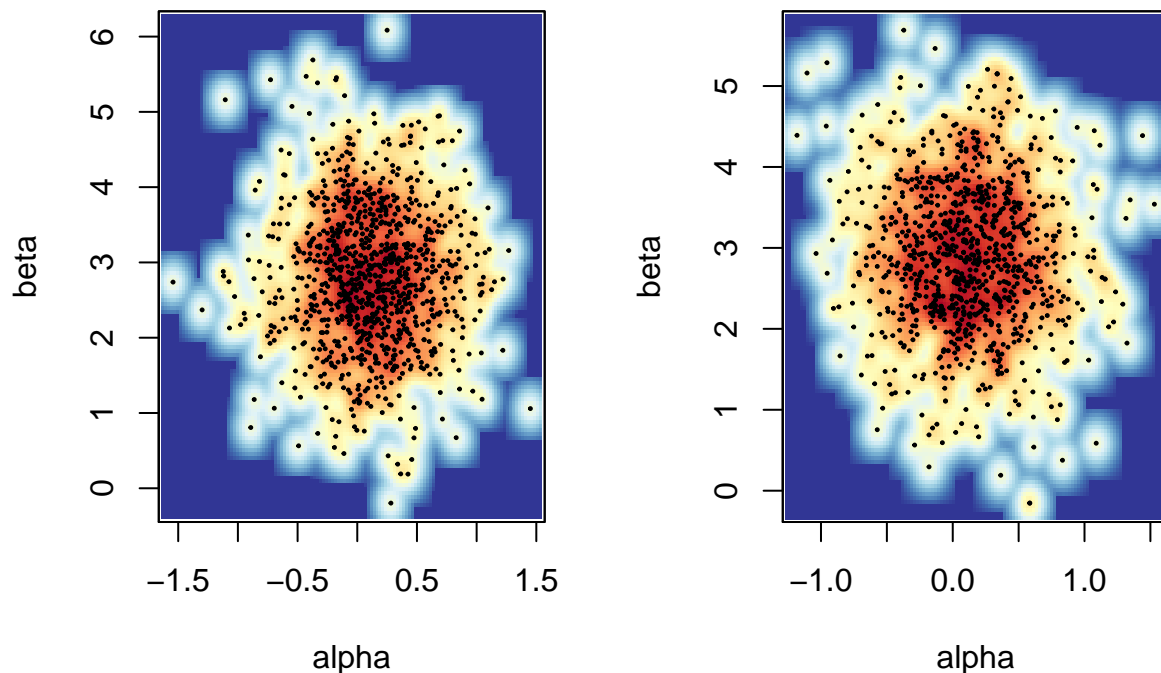
alpha <- rnorm(S, mu.a, sd.a)
beta <- rnorm(S, mu.b, sd.b)
df <- cbind(alpha,beta)
df <- cbind(df, g(alpha,beta), q(alpha,beta), iratio(alpha,beta))
df <- as.data.frame(df)
colnames(df) <- c("alpha","beta","g","q","imp.ratio")

resample <- sample(1:S, k, replace=F, prob=df$iratio)
resample2 <- sample(1:S, k, replace=T, prob=df$iratio)

par(mfrow=c(1,2))

my.cols <- rev(brewer.pal(11, "RdYlBu"))
smoothScatter(alpha[resample],beta[resample], nrpoints=.3*S,
  colramp=colorRampPalette(my.cols), pch=19, cex=.2,xlab='alpha', ylab='beta')
smoothScatter(alpha[resample2],beta[resample2], nrpoints=.3*S,
  colramp=colorRampPalette(my.cols), pch=19, cex=.2,xlab='alpha', ylab='beta')

```



**part (b)**

The scatter plot with replacement is plotted above. By comparing them side by side, we find out that posterior draws with or without replacement seem alike. The scatter plot with replacement seems a little bit wider than the former, but generally we claim there is no big difference. This is probably caused by large sample size of our simulated data.

## Problem 5

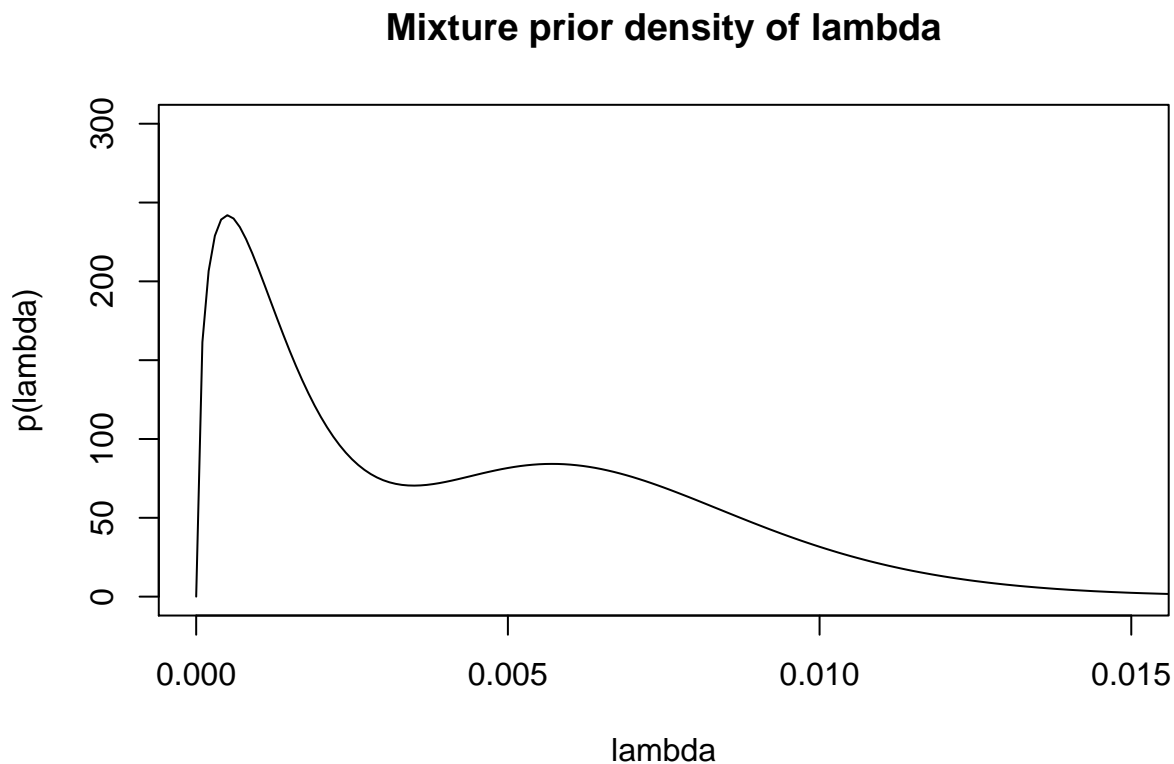
Suppose you are interested in estimating the mortality rate  $\lambda$  (per unit of exposure) of heart transplant surgeries for a particular hospital. To construct your prior, you talk to two experts. The first expert's beliefs about  $\lambda$  are described by a  $\text{gamma}(1.5, 1000)$  distribution and the second expert's beliefs are described by a  $\text{gamma}(7, 1000)$  distribution. You place equal credence in both experts.

- Construct a graph of the prior density of  $\lambda$ .
- Suppose the hospital of interest experiences  $y_{\text{obs}} = 4$  deaths with an exposure of 1767 patients. Compute the posterior distribution of  $\lambda$ .
- Plot the prior and posterior densities of  $\lambda$  on the same graph.
- Find the probability that the mortality rate exceeds 0.005.
- Based on the mixing probabilities, were the data more consistent with the beliefs of the first expert or the beliefs of the second expert? Explain.

## Solution

part (a)

```
lambda<-seq(0,1,0.0001)
lambda.d <- 0.5*dgamma(lambda,1.5,1000)+0.5*dgamma(lambda,7,1000)
plot(lambda,lambda.d,type="l",ylab="p(lambda)", xlab="lambda",
      xlim = c(0, 0.015), ylim = c(0, 300),
      main='Mixture prior density of lambda')
```



part (b)

Here we used built-in function `poisson.gamma.mix` from `LearnBayes`. (It's quite odd that the input is a matrix of parameters of Gamma distribution.)

```
gamma.par <- rbind(c(1.5,1000), c(7,1000))
prob <- c(0.5,0.5)
data <- list(y=4,t=1767)
(pgm <- poisson.gamma.mix(prob,gamma.par,data))
```

```
## $probs
## [1] 0.7597182 0.2402818
##
## $gammapar
##      [,1] [,2]
## [1,]  5.5 2767
## [2,] 11.0 2767
```

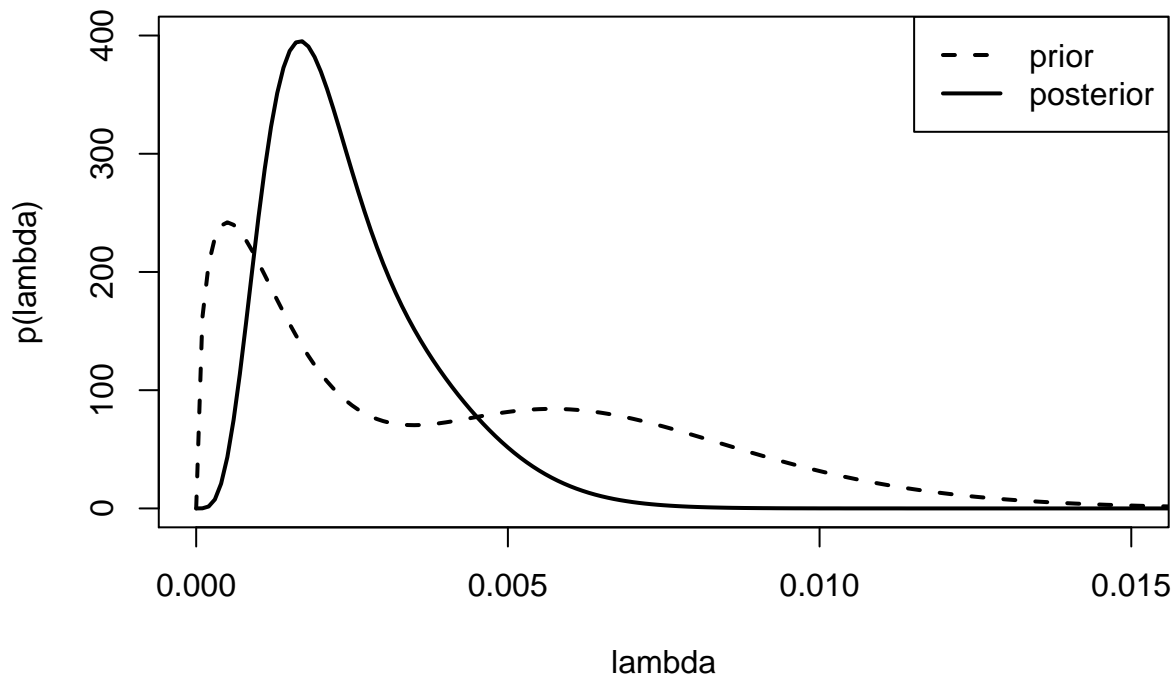
So the posterior distribution of  $\lambda$  is

$$p(\lambda \mid \text{data}) = 0.76 \cdot \text{gamma}(5.5, 2767) + 0.24 \cdot \text{gamma}(11, 2767)$$

part (c)

```
plot(lambda,lambda.d,type="l",ylab="p(lambda)", xlab="lambda",
      xlim = c(0, 0.015), ylim = c(0, 400),lty=2,lwd=2,
      main='Posterior and prior densities of lambda')
lines(lambda, pgm[[1]][1]*dgamma(lambda,pgm[[2]][1,1],pgm[[2]][1,2])+
      pgm[[1]][2]*dgamma(lambda,pgm[[2]][2,1],pgm[[2]][2,2]),lty=1,lwd=2)
legend("topright", c("prior","posterior"), lty=c(2,1),lwd=c(2,2))
```

### Posterior and prior densities of lambda





#### part (d)

This is computed by simulation.

```
sims <- 10000
expert1 <- rbinom(sims, 1, pgm[[1]][1]); expert2 <- 1 - expert1
gamma1 <- rgamma(sims, pgm[[2]][1,1], pgm[[2]][1,2])
gamma2 <- rgamma(sims, pgm[[2]][2,1], pgm[[2]][2,2])
gamma.mix <- expert1*gamma1+expert2*gamma2
sum(gamma.mix>0.005)/sims

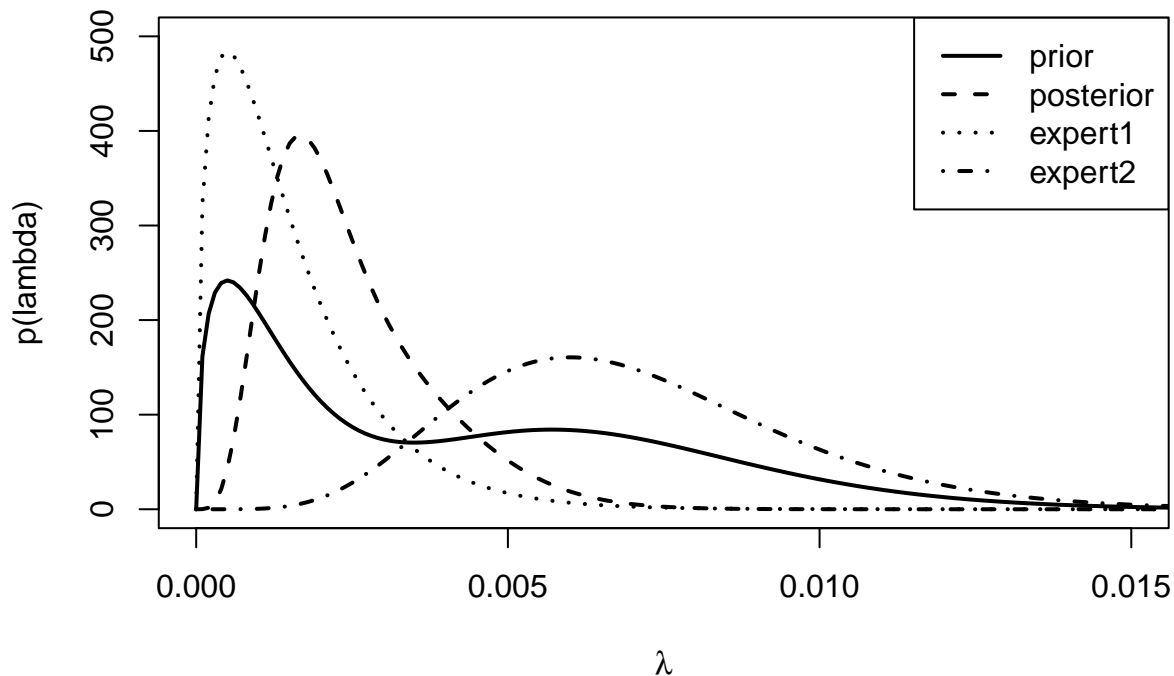
## [1] 0.0449
```

The probability that the mortality rate exceeds 0.005 is 4.49%.

#### part (e)

```
plot(lambda,lambda.d,type="l",ylab="p(lambda)", xlab=expression(lambda),
     xlim = c(0, 0.015), ylim = c(0, 500), lty=1, lwd=2,
     main='Comparison among mixed prior, separate priors and posterior')
lines(lambda, pgm[[1]][1]*dgamma(lambda,pgm[[2]][1,1],pgm[[2]][1,2])+
      pgm[[1]][2]*dgamma(lambda,pgm[[2]][2,1],pgm[[2]][2,2]), lty=2,
      lwd=2)
lines(lambda, dgamma(lambda,1.5,1000), lty=3, lwd=2)
lines(lambda, dgamma(lambda,7,1000), lty=4, lwd=2)
legend("topright", c("prior","posterior","expert1","expert2"), lty=1:4,lwd=rep(2,4))
```

### Comparison among mixed prior, separate priors and posterior



According to the plot with all four of our densities, the curve of Expert 1 seems closer to posterior than our Expert 2. On the other hand, the weights in our mixed model suggest that the model Expert 1 suggests has a larger weight (almost 3 times of the weight of Expert 2's model). Therefore, the data is more consistent with the first expert.

## Problem 6

Given observations  $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} N(\theta, \sigma^2)$  and using the conjugate prior distribution for  $\theta$  and  $\sigma^2$  (that is,  $\frac{1}{\sigma^2} \sim \text{Gamma}(\nu_0/2, \sigma_0^2 \nu_0/2)$  and  $\theta \mid \sigma^2 \sim N(\mu_0, \sigma^2/\kappa_0)$ ), derive the formula for  $p(\theta \mid y_1, \dots, y_n)$ , the marginal distribution of  $\theta$ , conditional on the data but marginal over  $\sigma^2$ . Check your work by comparing your formula to a Monte Carlo estimate of the marginal distribution, using some values of  $Y_1, \dots, Y_n, \mu_0, \sigma_0^2, \nu_0$  and  $\kappa_0$  that you choose.

## Solution

We have the following priors

$$\begin{aligned} \frac{1}{\sigma^2} &\sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\sigma_0^2 \nu_0}{2}\right) \\ \theta \mid \sigma^2 &\sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \end{aligned}$$

Then the joint prior density is

$$p(\theta, \sigma^2) \propto \sigma^{-1} (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2} [\nu_0 \sigma_0^2 + \kappa_0 (\mu_0 - \theta)^2]\right)$$

Then the joint posterior density is the product of joint prior density and normal likelihood, which is

$$\begin{aligned} p(\theta, \sigma^2 \mid y) &\propto \sigma^{-1} (\sigma^2)^{-(\nu_0/2+1)} \exp\left(-\frac{1}{2\sigma^2} [\nu_0 \sigma_0^2 + \kappa_0 (\mu_0 - \theta)^2]\right) \\ &\quad \times (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} [(n-1)s^2 + n(\bar{y} - \theta)^2]\right) \\ &\propto \sigma^{-1} (\sigma^2)^{-(\nu_0/2+1+n/2)} \exp\left(-\frac{1}{2\sigma^2} [\nu_0 \sigma_0^2 + \kappa_0 (\mu_0 - \theta)^2 + (n-1)s^2 + n(\bar{y} - \theta)^2]\right) \\ &= \sigma^{-1} (\sigma^2)^{-(\nu_n/2+1)} \exp\left(-\frac{1}{2\sigma^2} [\nu_n \sigma_n^2 + \kappa_n (\mu_n - \theta)^2]\right) \text{ where} \\ \mu_n &= \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y} \\ \kappa_n &= \kappa_0 + n \\ \nu_n &= \nu_0 + n \\ \nu_n \sigma_n^2 &= \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_n + n} (\bar{y} - \mu_0)^2. \end{aligned}$$

Next we integrate the joint posterior density with respect to  $\sigma^2$ .

$$\begin{aligned} p(\theta \mid y) &= \int_0^\infty \sigma^{-1} (\sigma^2)^{-(\nu_n/2+1)} \exp\left(-\frac{1}{2\sigma^2} [\nu_n \sigma_n^2 + \kappa_n (\mu_n - \theta)^2]\right) d\sigma^2 \\ &= \left(1 + \frac{\kappa_n (\theta - \mu_n)^2}{\nu_n \sigma_n^2}\right)^{-(\nu_n+1)/2} \\ &= t_{\nu_n}(\theta \mid \mu_n, \sigma_n^2/\kappa_n) \end{aligned}$$

```

n <- 100
y <- rnorm(n,0,1)
ybar <- mean(y)

mu0 <- 0.3 # set
sig0 <- 1.2 # set
nu0 <- 1 # set
k0 <- 1 # set

kn <- k0+n
nun <- nu0+n
mun <- k0/(k0+n)*mu0 + n/(k0+n)*ybar
sign <- sqrt((nu0*sig0^2 + (n-1)*var(y) + k0*n/(kn+n)*(ybar-mu0)^2)/nun)

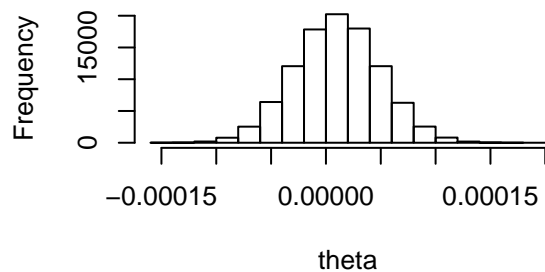
sims <- 100000
inv.sigma2 <- rgamma(1,nu0/2,sig0^2*nu0/2)
sigma2 <- 1/inv.sigma2
theta <- rnorm(sims,mu0,sqrt(sigma2/k0))
theta <- theta/sum(theta)

par(mfrow=c(2,2))
hist(theta,main='Histogram of theta (analytic)')
plot(density(theta),col='red',main='Density plot of theta (simulated)')

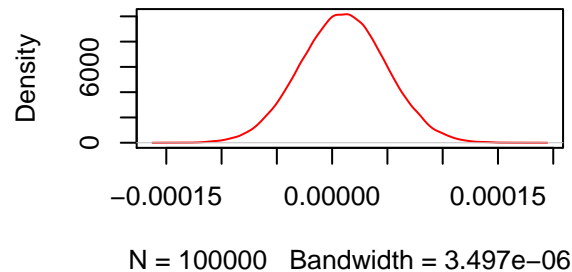
tt <- (rt(sims,df=nun)-mun)*sign/sqrt(kn)
tt <- tt/sum(tt)
hist(tt,main='Histogram of theta (simulated)')
plot(density(tt),col='blue',main='Density plot of theta (simulated)')

```

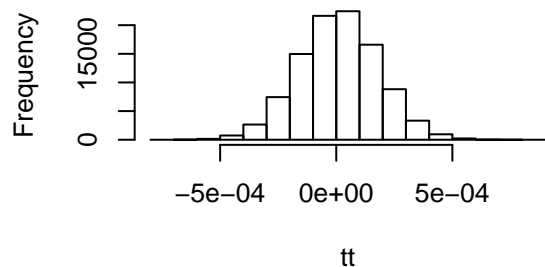
**Histogram of theta (analytic)**



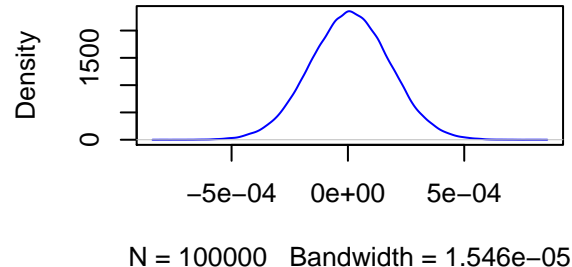
**Density plot of theta (simulated)**



**Histogram of theta (simulated)**



**Density plot of theta (simulated)**



By simulation, we conclude that the marginal posterior distributions of  $\theta$  (conditional on the data but marginal over  $\sigma^2$ ) computed analytically and simulated look alike. Though not with the same scale, the two should be proportional to each other.

## References

- *Bayesian Data Analysis*, Second Edition (Chapman & Hall/CRC Texts in Statistical Science) (29 July 2003) by Andrew Gelman, John B. Carlin, Hal S. Stern, Donald B. Rubin
- *Bayesian Computation with R (Use R)* (15 May 2009) by Jim Albert
- *LearnBayes: Functions for Learning Bayesian Inference*, <https://cran.r-project.org/web/packages/LearnBayes/index.html>
- *Scatterplot with contour/heat overlay*, <https://stats.stackexchange.com/questions/31726/>
- *RColorBrewer: ColorBrewer Palettes*, <https://cran.r-project.org/web/packages/RColorBrewer/index.html>