Lab Report in Bayesian Learning

Laboration 1

TDDE07

David Grönberg, Erik Lindgren
Davgr686, Erili040

LINKÖPINGS UNIVERSITET

17-04-2020

Contents

List of Figures

List of Tables

1. Assignments

1.1 Assignment 1

1.1.1 a)

Random numbers are drawn from the posterior $\theta|y\tilde{}Beta(7, 15), y = (y_1, ..., y_n)$ by sampling a growing set of numbers from 1 to 100000 with a step size of 100. Figure 1.1 and 1.2 shows the posterior mean and standard deviation converging to the true values when the sampling size becomes large. The red line represents the true mean and standard deviation.

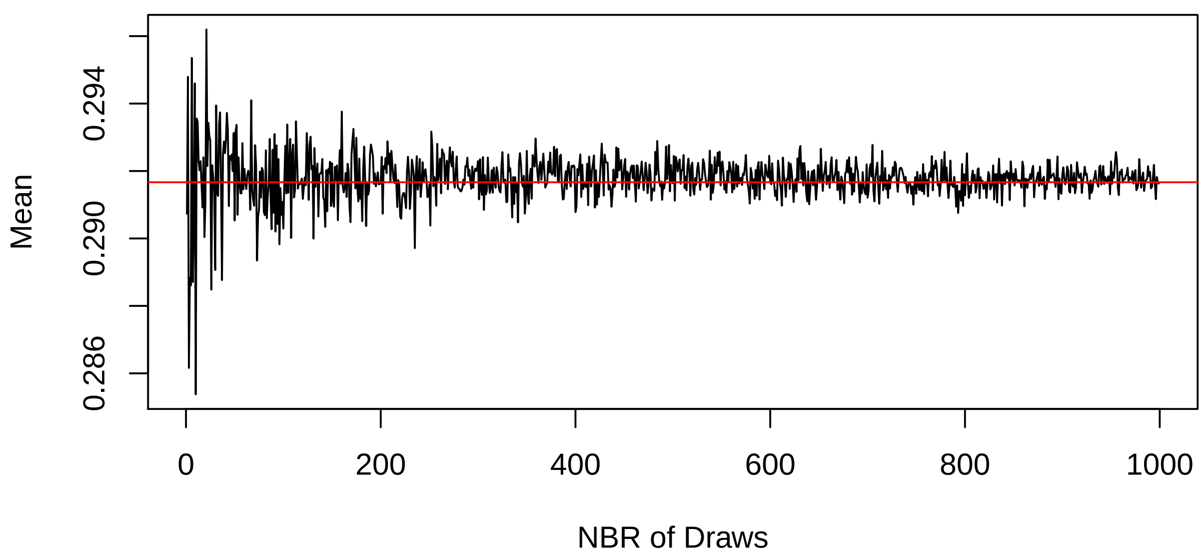

Figure 1.1: Posterior Mean

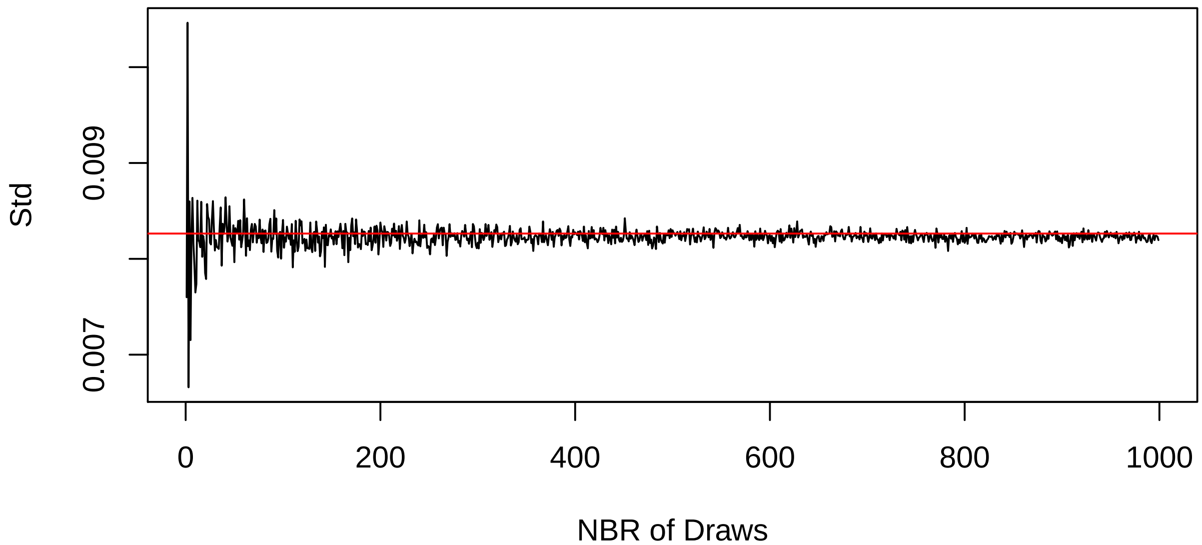

Figure 1.2: Posterior Standard Deviation

1.1.2 b)

Simulation is used to compute the probability $p(\theta > 0.3|y)$. This is done by sampling 10,000 values from the posterior distribution and then calculating the number of occurrences of $\theta > 0.3$ and dividing with the number of samples.

```
sampled <- sample(posterior, 10000, replace=TRUE )
sum(sampled > 0.3) / length(sampled)

## [1] 0.4478
```

The approximated probability is very close to the true probability $p(\theta > 0.3|y) \approx 0.4399$.

1.1.3 c)

Lastly, the posterior distribution of the log-odds are computed by sampling 10,000 values from the posterior distribution. Figure 1.3 shows the posterior log-odds distribution.

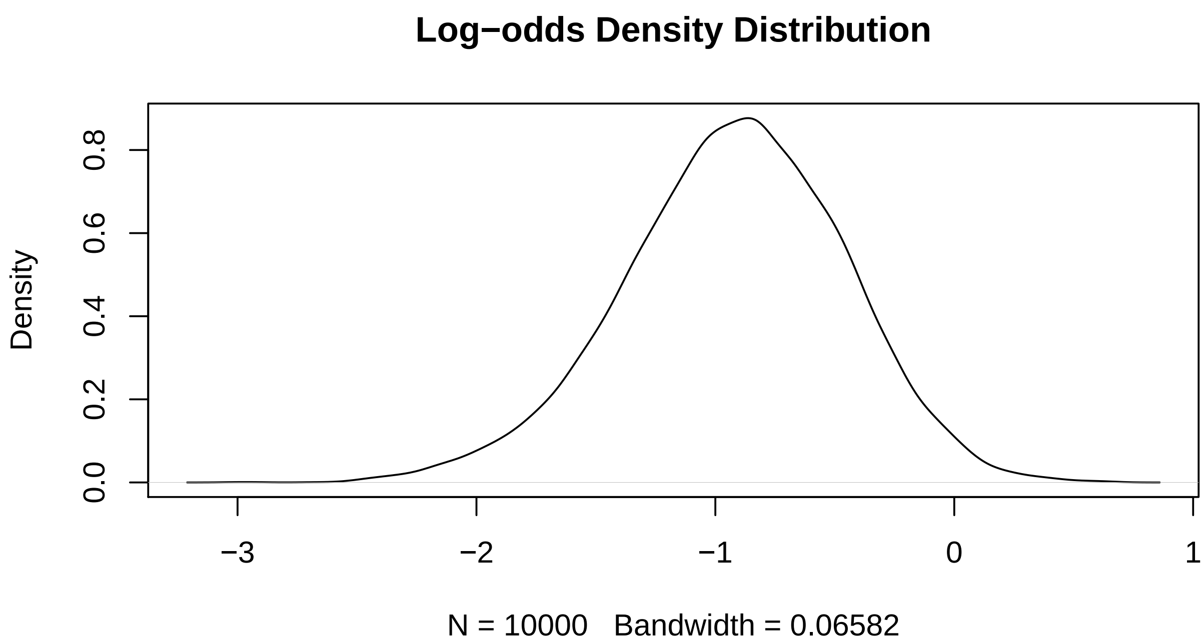

Figure 1.3: Posterior Log-odds Distribution

1.2 Assignment 2

1.2.1 a)

The following sample was gathered by asking 10 randomly selected persons about their income(tkr) : 44, 25, 45, 52, 30, 63, 19, 50, 34 and 67. We are tasked to simulate 10,000 draws from the posterior of σ^2 while assuming $\mu = 3.7$. The posterior for θ^2 is the $Inv - \chi^2(n, \tau^2)$ distribution, where

$$\tau^2 = \frac{\sum_{i=1}^{n}(log(y_i) - \mu)^2}{n}$$

After simulating the draws, the theoretical $Inv - \chi^2(n, \tau^2)$ distribution (red line) was plotted together with an histogram over the density of the draws, seen in figure 1.4. From figure 1.4 we can see that for large number of draws, the simulated distribution is close to the theoretical distribution.

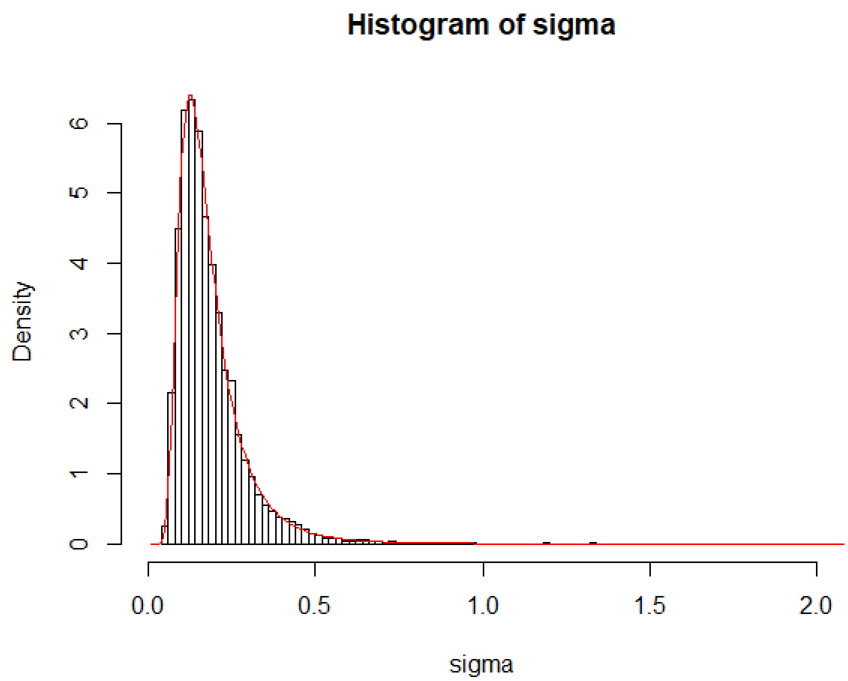

Figure 1.4: Simulated draws from posterior of σ in comparison with the theoretical $Inv - \chi^2(n, \tau^2)$ distribution(red dotted line).

1.2.2 b)

The next task is to calculate the posterior distribution of the Gini coefficient for the current data set. When the data follows a $logN(\mu, \sigma^2)$ distribution, the Gini coefficient G can be obtained by:

$$G = 2\phi(\frac{\sigma}{\sqrt{2}}) - 1$$

The final posterior distribution of the Gini coefficient can be showed in figure 1.5:

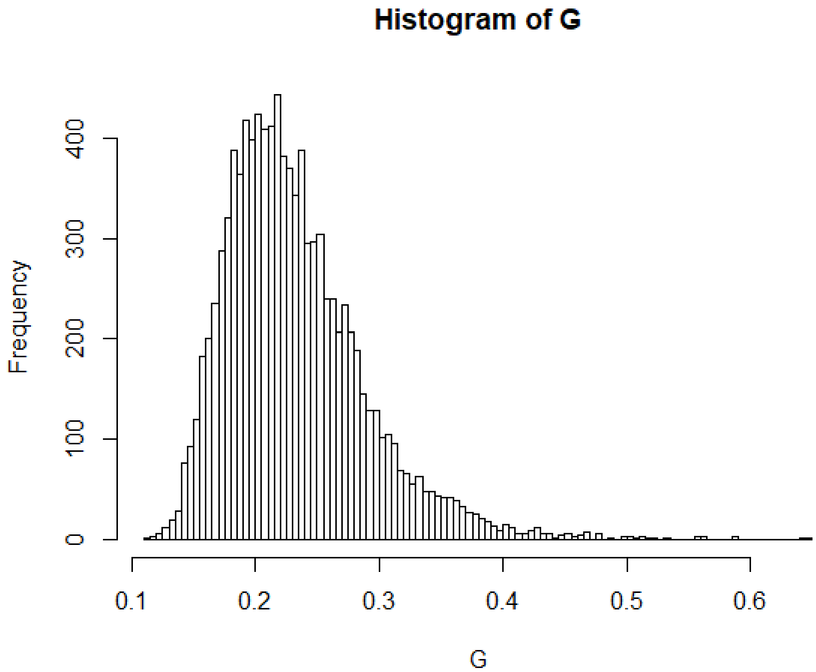

Figure 1.5: Posterior distribution of the Gini coefficient

1.2.3 c)

Lastly a 90% equal tail interval was calculated from the posterior distribution of the Gini coefficient calculated previously in (b). The interval spanned between approximately 0.161 and 0.339.

In addition to calculating the tail interval, the 90% Highest density interval(HDI) was calculated using the density function in r in combination with the hdi function. The interval from the hdi function spanned between 0.147 and 0.317. From this we can see that the two calculated intervals are very similar to each other.

1.3 Assignment 3

We are given 10 wind direction observations that are assumed to be independent observations following the von Mises distribution.

$$p(y|\mu,\kappa) = \frac{\exp[k * cos(y-\mu)]}{2\pi I_0(\kappa)}, \pi \leq y \leq \pi$$

μ is given to be 2.39. $\kappa\tilde{}Exponential(\lambda = 1)$ a priori. The posterior distribution of κ is obtained by:

$$p(\kappa|y_1,,,y_n) \propto \frac{\exp\left(\kappa\sum_{i=1}^{n}\cos(y_i-\mu)\right)}{(2\pi I_0(\kappa))^{-n}}\exp(-\kappa)$$

Figure 1.6 shows the posterior distribution of κ (over κ values ranging between 0 and 10) represented by the black line and the approximate mode is represented by the red line.

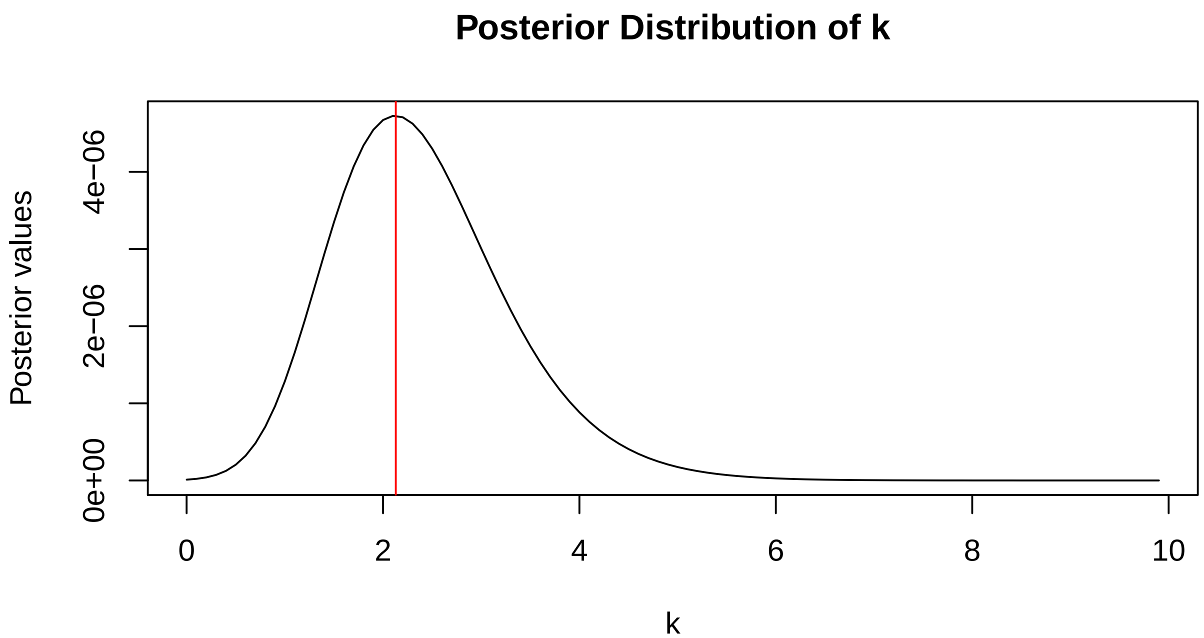

Figure 1.6: Posterior Distribution of κ

2. Code Appendix

2.1 Assignment 1

```r
s <- 5
n <- 20
f <- n - s
a0 <- 2
b0 <- 2
set.seed(12345)

posterior <- rbeta(100000, a0 + s, b0 + f)

means <- c()
vars <- c()
indx <- 0
for (i in seq(1,100000, 100))
{
  post <- sample( posterior, i, replace=TRUE )
  means[indx] <- mean(post)
  vars[indx] <- sd(post)
  indx <- indx + 1

}

truemean <- (a0 + s) / (a0 + s + b0 + f)
truevar <- (a0 + s) * (b0 + f) / ((b0 + f + a0 + s)**2 * (b0 + f + a0 + s + 1))

plot(means, type="l")
abline(h=truemean, col = 'red')
plot(vars**2, type="l")
abline(h=truevar, col = 'red')

## B
post <- sample( posterior, 10000, replace=TRUE )
sum(post > 0.3) / length(post)

prob_theta_bigger_than_03 <- pbeta(0.3, a0 + s, b0 + f, lower.tail = FALSE)

## C
phis <- density(log(post/(1-post)))
plot(phis)
```

2.2 Assignment 2

```r
data = c(44,25,45,52,30,63,19,50,34,67)
```

```
u = 3.7
n = 10
t = (sum((log(data)-u)^2))/n

#a
sigma = t * n / rchisq(10000,df=n)

interval = seq(0,3,0.01)
invchisq = dinvchisq(interval, df=10, t)

hist(sigma,100, xlim=c(0,2), freq = FALSE)
lines(interval, invchisq, col="red", xlim=c(0,2))
#b
z = sqrt(sigma / 2)
G = 2*pnorm(z)-1

hist(G)


#c

tail_interval = quantile(g,probs = c(0.05,0.95) )
tail_interval

dens = density(g)
plot(dens)

hdi_interval = hdi(dens, credMass=0.9)
hdi_interval[1:2]
```

2.3 Assignment 3

```
set.seed(12345)
Y <- c(-2.44,2.14,2.54,1.83,2.02,2.33,-2.79,2.23,2.07,2.02)

u <- 2.39

k <- seq(0.001,10,0.1)

posterior <- function(k) {
  return (exp(k * sum(cos(Y - u)))/((2 * pi * besselI(k, 0))**length(Y)) * dexp(k))
}

post <- sapply(k, posterior)

plot(k, post, type='l')
abline(v=2.15, col="red")
```