Stats 551 Homework 1

Yirui Gao

Jan 27

Contents

Question 1. Pratice of Bayes Formula

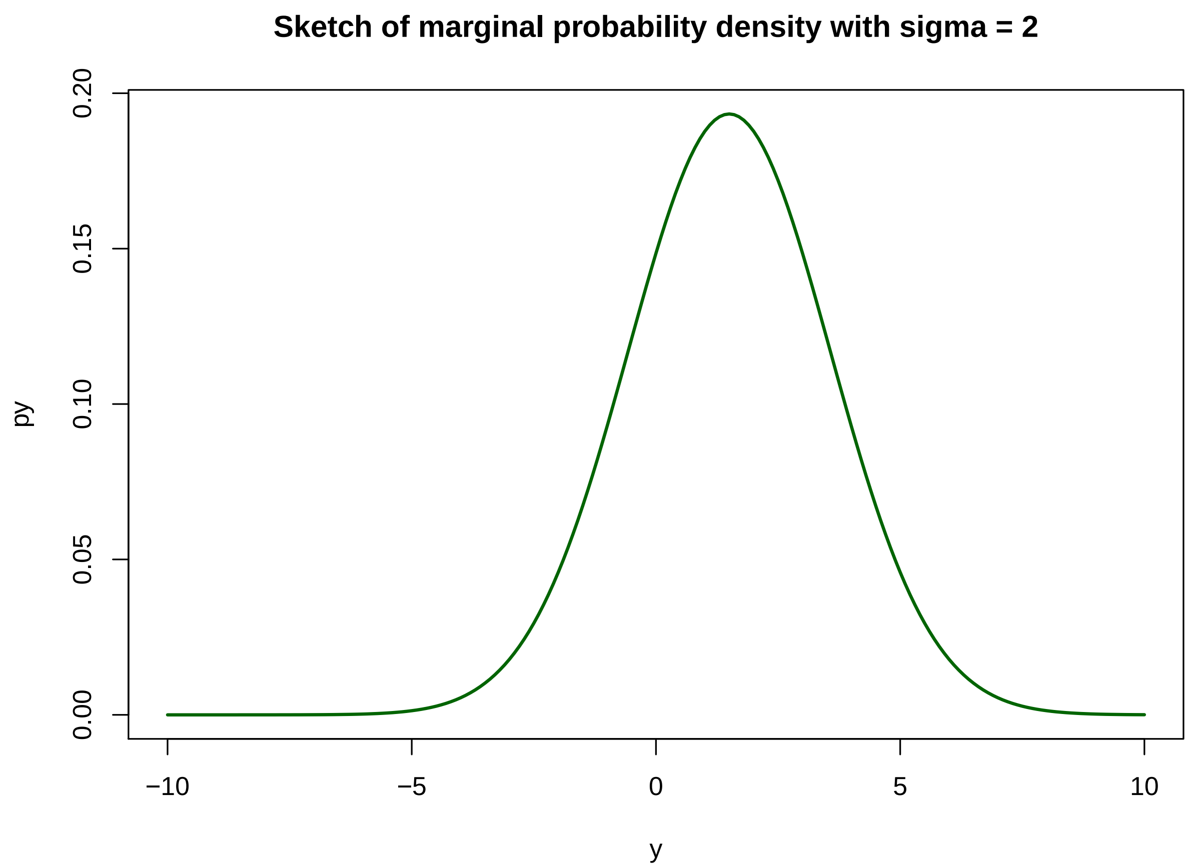
(1) The prior for this problem is $Pr(\theta = 1) = Pr(\theta = 2) = \frac{1}{2}$ and the likelihood is still denoted as $P(y|\theta)$. The marginal probability density for y is:

$$\begin{aligned} p(y) &= p(\theta = 1)p(y|\theta = 1) + p(\theta = 2)p(y|\theta = 2) \\ &= \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(y-\mu_1)^2}{2\sigma^2}) + \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{(y-\mu_2)^2}{2\sigma^2}) \\ &= \frac{1}{2\sigma\sqrt{2\pi}}[\exp(-\frac{(y-\mu_1)^2}{2\sigma^2}) + \exp(-\frac{(y-\mu_2)^2}{2\sigma^2})]. \end{aligned}$$

By substituting σ with 2 and two means of 1 and 2, the formula gets further simplied:

$$p(y) = \frac{1}{4\sqrt{2\pi}}[\exp(-\frac{(y-1)^2}{8}) + \exp(-\frac{(y-2)^2}{8})].$$

```r
y <- seq(-10, 10, by=0.1)
f <- function(y) {
  result <- (exp(-(y-1)^2/8) + exp(-(y-2)^2/8))/(4*sqrt(2*pi))
  return(result)
}
py <- lapply(y, f)
plot(y, py, col = "dark green", type = 'l', lwd = 2,
     main = "Sketch of marginal probability density with sigma = 2")
```



(2) Use R command and the function from (1), plug in $y = 1$ and we can get $P(y = 1) = 0.1877519$.

```
f(1)
```

```
## [1] 0.1877519
```

When θ equals 1, the mean is 1 and then $p(y = 1|\theta = 1) = \frac{1}{2\sqrt{2\pi}} = 0.1994711$. Based on the Bayes rule, we can get the posterior density

$$
\begin{aligned}
Pr(\theta = 1|y = 1) &= \frac{p(\theta = 1)p(y = 1|\theta = 1)}{p(y = 1)} \\
&= \frac{0.5 * 0.1994711}{0.1877519} \\
&= 0.5312094.
\end{aligned}
$$

(3) Since we have already derived the distribution for the marginal pdf $p(y) = \frac{1}{2\sigma\sqrt{2\pi}}[\exp(-\frac{(y-1)^2}{2\sigma^2}) + \exp(-\frac{(y-2)^2}{2\sigma^2})]$, we can use the Bayes rule to derive correspondingly the posterior distribution for $\theta = 1$ and $\theta = 2$:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} = \frac{exp(-\frac{(y-\theta)^2}{2\sigma^2})}{\exp(-\frac{(y-1)^2}{2\sigma^2}) + \exp(-\frac{(y-2)^2}{2\sigma^2})},$$

$$p(\theta = 1|y) = \frac{exp(-\frac{(y-1)^2}{2\sigma^2})}{\exp(-\frac{(y-1)^2}{2\sigma^2}) + \exp(-\frac{(y-2)^2}{2\sigma^2})} = \frac{1}{1 + exp(\frac{2y-3}{2\sigma^2})},$$

$$p(\theta=2|y) = \frac{exp(-\frac{(y-2)^2}{2\sigma^2})}{\exp(-\frac{(y-1)^2}{2\sigma^2}) + \exp(-\frac{(y-2)^2}{2\sigma^2})} = \frac{1}{1+exp(\frac{-2y+3}{2\sigma^2})}.$$

So we can see that when σ is increased, the posterior probabilities will be more and more close to $\frac{1}{2}$. Especially when σ comes to a very huge number, the posterior probability will be the prior probability that $p(\theta=1|y) = p(\theta=2|y) = p(\theta) = \frac{1}{2}$.

When σ is decreasing, the posterior probabilities will be more and more close to 1. So for $p(\theta=1|y)$, when $y < \frac{3}{2}$, $p(\theta=1|y)$ will appraoch 1 as θ decreases and comes to 1 when $\theta=0$. At this time $p(\theta=2|y)=0$, so all the posterior distribution would incline to $\theta=1$. On the contrary, when $y > \frac{3}{2}$, $p(\theta=2|y)$ will appraoch 1 as θ decreases and comes to 1 when $\theta=0$. At this time $p(\theta=1|y)=0$, so all the posterior distribution would incline to $\theta=2$.

Question 2. Normal distribution with unknown mean

(1) For this problem, the prior distribution for θ is $p(\theta) \sim \mathbb{N}(180, 40^2)$, here denote $\mu_0 = 180$ as the mean and $\tau_0 = 40$ as the standard deviation. To get the posterior distribution $p(\theta|y_1, y_2, ...y_n)$, we need to calculate the likelihood:

$$p(y_1, y_2, ...y_n|\theta) = \prod_{i=1}^{n} \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(y_i-\theta)^2}{2\sigma^2}),$$

where $\sigma = 20$ stated in the problem.

Thus the posterior follows another normal distribution $p(\theta|\mathbf{y}) \propto \mathbb{N}(\mu_n, \tau_n^2)$ as follows:

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto p(\theta)p(\mathbf{y}|\theta) \\ &\propto \exp(-\frac{1}{2}[\frac{\sum_{i=1}^{n}(y_i-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}]) \\ &\propto \exp(-\frac{1}{2\tau_n^2}(\theta-\mu_n)^2). \end{aligned}$$

Given the parameters from the problem, we can derive the weighted average of the prior mean and the observed values as

$$\begin{aligned} \mu_n &= \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\overline{y}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \\ &= \frac{600n+180}{4n+1}, \\ \tau_n^2 &= \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \\ &= \frac{1600}{4n+1}. \end{aligned}$$

So, the posterior distribution for θ is $p(\theta|\mathbf{y}) \propto \mathbb{N}(\frac{600n+180}{4n+1}, \frac{1600}{4n+1})$.

(2) According to the definition of posterior predictive distribution $p(\tilde{y}|\mathbf{y}) = \int_\theta p(\tilde{y}|\theta)p(\theta|\mathbf{y})d\theta$, since we have already derived the posterior distribution for θ, so

$$\begin{aligned} p(\tilde{y}|\mathbf{y}) &= \int_\theta p(\tilde{y}|\theta)p(\theta|\mathbf{y})d\theta \\ &= \int_\theta \mathbb{N}(\theta, 20^2)\mathbb{N}(180, 40^2)d\theta \\ &\sim \mathbb{N}(\theta', \sigma'^2). \end{aligned}$$

Recall that $E(\tilde{y}|\theta) = \theta$, $\text{Var}(\tilde{y}|\theta) = \sigma^2$. Then

$$E(\tilde{y}|\mathbf{y}) = E[\text{Var}(\tilde{y}|\theta, \mathbf{y})|\mathbf{y}] = E(\theta|\mathbf{y}) = \mu_n,$$

$$\text{Var}(\tilde{y}|\mathbf{y}) = E[E(\tilde{y}|\theta,\mathbf{y})|\mathbf{y}] + \text{Var}[E(\tilde{y}|\theta,\mathbf{y})|\mathbf{y} = E(\sigma^2|\mathbf{y}) + \text{Var}(\theta|\mathbf{y}) = \sigma^2 + \tau_n^2.$$

Therefore, after plug in the corresponding values, the posterior predictive distribution for $\tilde{y}$ is:

$$\tilde{y}|\mathbf{y} \sim \mathbb{N}(\frac{600n + 180}{4n + 1}, \frac{1600n + 2000}{4n + 1}).$$

(3) The posterior interval for θ for $n = 10$ and $n = 100$ can be found by implementing the following R code:

```r
mean <- function(n) {
  result <- (600 * n + 180) / (4*n + 1)
  return(result)
}
variance <- function(n){
  result <- 1600 / (4 * n + 1)
  return (result)
}

n1 <- 10
mysamples_1 <- rnorm(10000, mean = mean(n1), sd = sqrt(variance(n1)))
print("The 95% posterior interval for theta when n = 10 is: ")
```

```
## [1] "The 95% posterior interval for theta when n = 10 is: "
```

```r
quantile(mysamples_1, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 138.6119 163.0228
```

```r
n2 <- 100
mysamples_2 <- rnorm(10000, mean = mean(n2), sd = sqrt(variance(n2)))
print("The 95% posterior interval for theta when n = 100 is: ")
```

```
## [1] "The 95% posterior interval for theta when n = 100 is: "
```

```r
quantile(mysamples_2, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 146.1878 154.0075
```

Similarly, the posterior predictive interval for $\tilde{y}$ for $n = 10$ and $n = 100$ can be found by implementing the following R code:

```r
mean_y <- function(n) {
  result <- (600 * n + 180) / (4*n + 1)
  return(result)
}
variance_y <- function(n){
  result <- (1600 * n + 2000) / (4 * n + 1)
  return (result)
}

n1 <- 10
mysamples_1 <- rnorm(10000, mean = mean_y(n1), sd = sqrt(variance_y(n1)))
print("The 95% posterior interval for y_pre when n = 10 is: ")
```

```
## [1] "The 95% posterior interval for y_pre when n = 10 is: "
```

```r
quantile(mysamples_1, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 108.8829 191.2907
```

```r
n2 <- 100
mysamples_2 <- rnorm(10000, mean = mean_y(n2), sd = sqrt(variance_y(n2)))
print("The 95% posterior interval for y_pre when n = 100 is: ")
```

```
## [1] "The 95% posterior interval for y_pre when n = 100 is: "
```

```r
quantile(mysamples_2, c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 111.1402 189.7562
```

Question 3. Nonconjugate single parameter model

(1) Given that the prior is triangular shaped distribution and the likelihood to be the binomial distribution, we also know the distribution of the likelihood $p(y|\theta) = \theta^y(1-\theta)^{n-y}$ since $y|\theta \sim Binomial(500, \theta)$, here $n = 500$.

To get the unnormalized posterior density function, we can just multiply the likelihood with the triangular prior, with the given $y = 285$:

$$p(\theta|y)_{unnorm} = p(\theta)p(y|\theta) = p(\theta)\theta^y(1-\theta)^{(n-y)},$$

where $p(\theta)$ should be replaced by the given triangle function.

Now use R codes to compute the unnormalized posterior density function on a grid given $m = 100$:

```r
# Define the prior distribution:
triangle.prior <- function(theta){
  if(theta >= 0 && theta < 0.25){
    return (8 * theta)
  }
  else if(theta >= 0.25 && theta <= 1){
    return (8 * (1 - theta) / 3)
  }
  else return (0)
}

# Define the unnormalized posterior distribution:
unnorm.posterior <- function(theta, n, y){
  posterior <- (theta^y) * (1 - theta)^(n - y) * triangle.prior(theta)
  return (posterior)
}

# Grid approximation for unnormalized posteriors:
m <- 100
grid.points <- seq(0, 1, length.out = m)
unnorm.points = unnorm.posterior(grid.points, 500, 285)
unnorm.points
```

```
##   [1]  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
##   [6]  0.000000e+00  0.000000e+00  0.000000e+00 3.693635e-320 1.463372e-306
##  [11] 1.616996e-294 9.816122e-284 5.399247e-274 3.933900e-265 5.099701e-257
##  [16] 1.485950e-249 1.174796e-242 2.939187e-236 2.642853e-230 9.500240e-225
##  [21] 1.493691e-219 1.108959e-214 4.151931e-210 8.297379e-206 9.299171e-202
```

```
##  [26] 6.102668e-198 2.435847e-194 6.114779e-191 9.944877e-188 1.075879e-184
##  [31] 7.926671e-182 4.061797e-179 1.475142e-176 3.861806e-174 7.399450e-172
##  [36] 1.051992e-169 1.123546e-167 9.115150e-166 5.673887e-164 2.734349e-162
##  [41] 1.028484e-160 3.041299e-159 7.116206e-158 1.325139e-156 1.973791e-155
##  [46] 2.362081e-154 2.279866e-153 1.780589e-152 1.128338e-151 5.814164e-151
##  [51] 2.440259e-150 8.352066e-150 2.332670e-149 5.317286e-149 9.889248e-149
##  [56] 1.499381e-148 1.850765e-148 1.856366e-148 1.509360e-148 9.918157e-149
##  [61] 5.248142e-149 2.226710e-149 7.537917e-150 2.024381e-150 4.285086e-151
##  [66] 7.096507e-152 9.118431e-153 9.005263e-154 6.763488e-155 3.817396e-156
##  [71] 1.597594e-157 4.883396e-159 1.071916e-160 1.657545e-162 1.766978e-164
##  [76] 1.267056e-166 5.943327e-169 1.766368e-171 3.206729e-174 3.409217e-177
##  [81] 2.021240e-180 6.311920e-184 9.708170e-188 6.791852e-192 1.964961e-196
##  [86] 2.095067e-201 7.148835e-207 6.551363e-213 1.292198e-219 4.124767e-227
##  [91] 1.463372e-235 3.467711e-245 2.686920e-256 2.390560e-269 4.819196e-285
##  [96] 1.451131e-304  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
```

Next, use grid approximation again, to compute the normalized posterior density function, we need to estimate the normalizing constant denoted as c, where c represents the area under the unnormalized posterior. And here c can be estimated by the sum of the areas of the small rectangles of width $\frac{1}{m}$ and height at the unnormalized posteior ordinates, accoridng to the resources found at http://patricklam.org/teaching/grid_print.pdf:

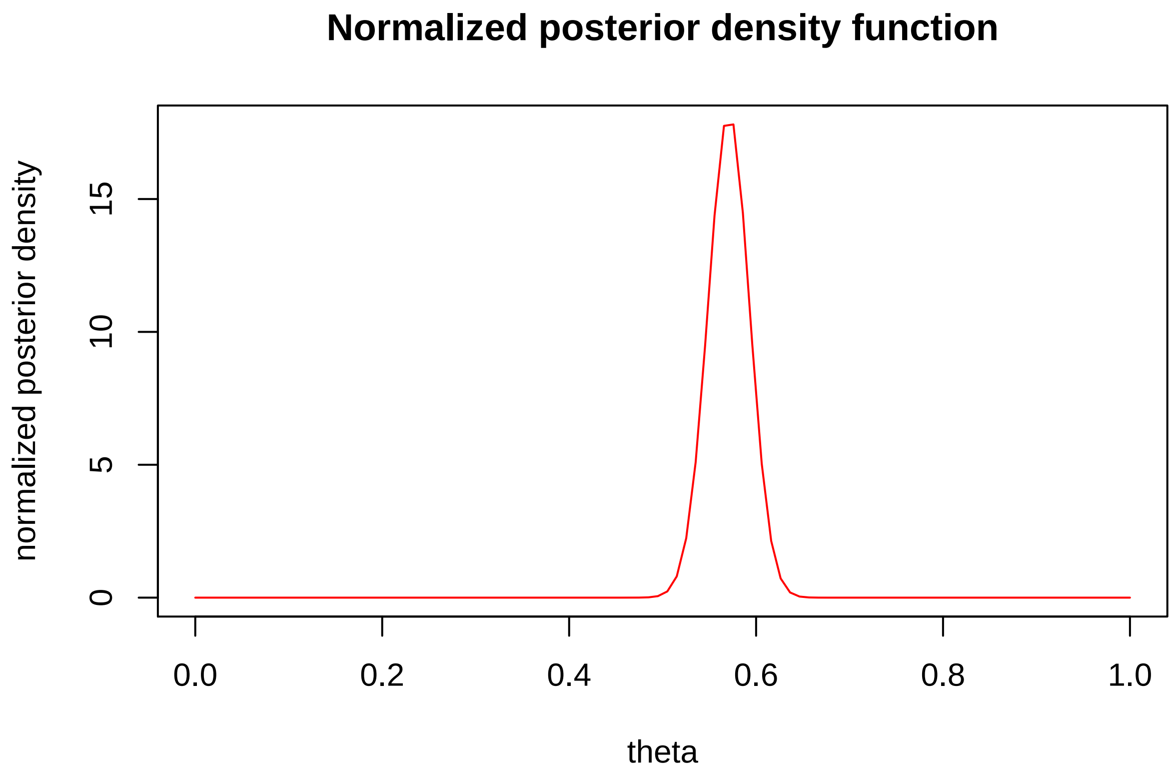
$$c \approx \sum_{i=1}^{m} \frac{1}{m} p(y|\theta_0 + \frac{i}{m}) p(\theta_0 + \frac{i}{m}).$$

To calculate and plot, use the following R code:

```r
# Normalized constant:
c <- sum(unnorm.points / m)
norm.points <- unnorm.points / c
norm.points
```

```
##   [1]  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
##   [6]  0.000000e+00  0.000000e+00  0.000000e+00 3.543095e-171 1.403730e-157
##  [11] 1.551093e-145 9.416052e-135 5.179193e-125 3.773568e-116 4.891855e-108
##  [16] 1.425388e-100  1.126915e-93  2.819396e-87  2.535140e-81  9.113044e-76
##  [21]  1.432814e-70  1.063761e-65  3.982713e-61  7.959207e-57  8.920170e-53
##  [26]  5.853945e-49  2.336570e-45  5.865562e-42  9.539559e-39  1.032030e-35
##  [31]  7.603608e-33  3.896252e-30  1.415021e-27  3.704412e-25  7.097875e-23
##  [36]  1.009117e-20  1.077754e-18  8.743649e-17  5.442639e-15  2.622907e-13
##  [41]  9.865666e-12  2.917347e-10  6.826175e-09  1.271131e-07  1.893346e-06
##  [46]  2.265811e-05  2.186947e-04  1.708019e-03  1.082351e-02  5.577200e-02
##  [51]  2.340803e-01  8.011665e-01  2.237598e+00  5.100572e+00  9.486197e+00
##  [56]  1.438272e+01  1.775334e+01  1.780707e+01  1.447844e+01  9.513928e+00
##  [61]  5.034247e+00  2.135957e+00  7.230698e-01  1.941875e-01  4.110442e-02
##  [66]  6.807279e-03  8.746796e-04  8.638240e-05  6.487832e-06  3.661812e-07
##  [71]  1.532482e-08  4.684366e-10  1.028229e-11  1.589989e-13  1.694962e-15
##  [76]  1.215415e-17  5.701098e-20  1.694377e-22  3.076034e-25  3.270270e-28
##  [81]  1.938862e-31  6.054669e-35  9.312499e-39  6.515040e-43  1.884876e-47
##  [86]  2.009680e-52  6.857474e-58  6.284352e-64  1.239533e-70  3.956656e-78
##  [91]  1.403730e-86  3.326379e-96 2.577411e-107 2.293129e-120 4.622782e-136
##  [96] 1.391988e-155  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
```

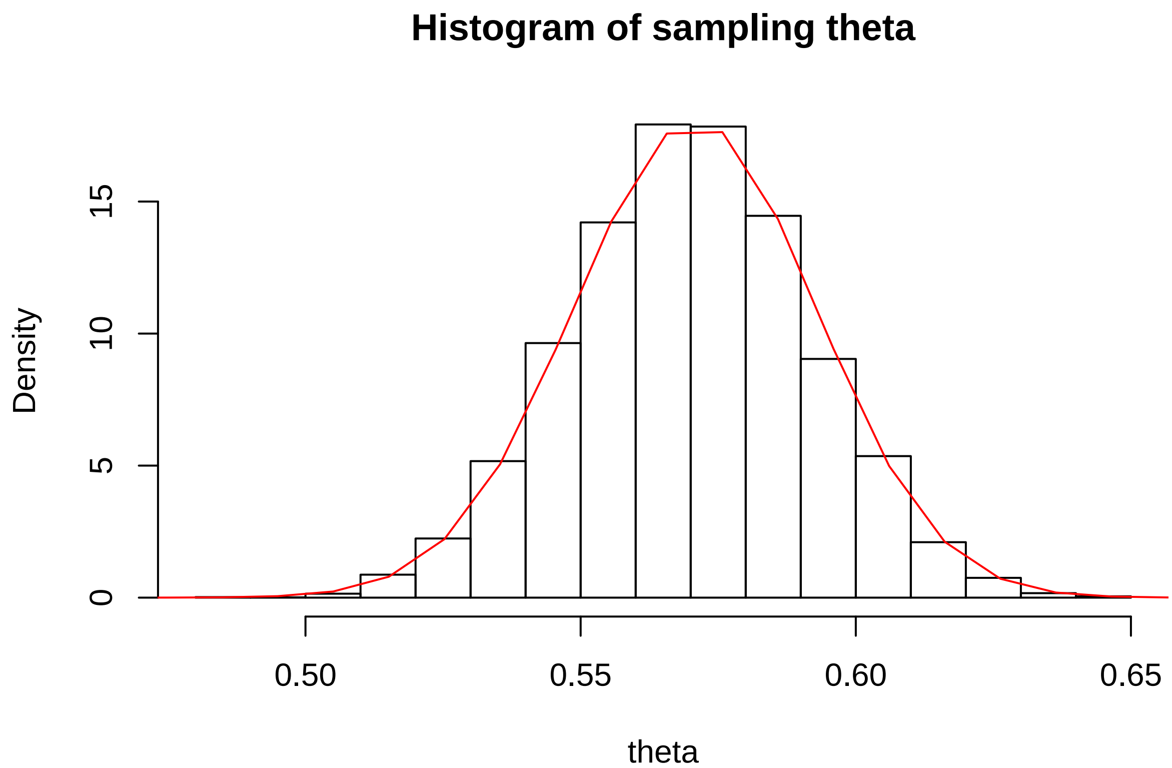
```r
# Plot the normalized posterior density function:
plot(x = grid.points, y = norm.points, type = 'l', col = 'red',
     xlab = 'theta', ylab = 'normalized posterior density',
     main = 'Normalized posterior density function')
```



(2)

```
set.seed(123456789)

N = 10000
samples <- sample(grid.points, size = N, replace = TRUE, prob = norm.points/sum(norm.points))
hist(samples, freq = FALSE, xlab = 'theta', main = 'Histogram of sampling theta')
lines(grid.points, norm.points/(sum(norm.points) * (grid.points[2]-grid.points[1])), col = "red")
```


Histogram of sampling theta
Density
0
5
10
15
0.50
0.55
0.60
0.65
theta