NAME: Kaifa Lu UFID: 5205-0501
STA6348 – Bayesian Analysis for Machine Learning and Uncertainty Quantification

Problem 1: Finding the support of the posterior distribution.

Problem 1:

1.1 $f(y|\theta) = N(\theta, 1)$, $p(\theta) \propto 1$ and $\Theta = R$

The posterior distribution is

$$p(\theta|y) \propto p(y|\theta) \cdot p(\theta)$$

$$\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \cdot 1}} \exp\left\{-\frac{(Y_i-\theta)^2}{2}\right\} \cdot 1$$

$$\propto (2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^{n}(Y_i-\theta)^2}{2}\right\}$$

$$\propto (2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{n\theta^2 - 2n\bar{Y} + \sum_{i=1}^{n} Y_i^2}{2}\right\}$$

$$\propto (2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{(\theta-\bar{Y})^2}{2/n} + \frac{n\bar{Y}^2 - \sum_{i=1}^{n} Y_i^2}{2}\right\}$$

$$\propto \exp\left\{-\frac{(\theta-\bar{Y})^2}{2/n}\right\} \quad \text{where } \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

So, $\theta|y \sim N(\bar{Y}, \frac{1}{n})$,

① the support of posterior distribution is R

② $\hat{\theta}_{MAP} = \arg\max_{\theta} p(\theta|y)$

Analytically, $\hat{\theta}_{MAP} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$

Numerically,
- first, set a range (a, b) containing $\hat{\theta}_{MAP}$ based on $\bar{Y}$
- second, equally divide (a, b) into a series of $N+1$ values. such as $a, a+\frac{b-a}{N}, a+\frac{2(b-a)}{N}, \ldots, b$
- third, calculate the $p(\theta|y)$ when $\theta = a, a+\frac{b-a}{N}, \ldots, b$
- fourth, select the maximum of $p(\theta|y)$ where the value of θ is $\hat{\theta}_{MAP}$

1.2 $f(y|\theta) = N(\theta, 1)$, $p(\theta) \propto \mathbb{I}(\theta \in [-1, 1])$

The posterior distribution is

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

$$\propto \exp\left\{-\frac{(\theta-\bar{Y})}{2/n}\right\} \mathbb{I}(\theta \in [-1,1]) \quad \text{where } \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

① the support of posterior distribution is $[-1, 1]$

② $\hat{\theta}_{MAP} = \underset{\theta}{\arg\max}\, p(\theta|y) \quad \theta \in [-1, 1]$

Analytically: $\begin{cases} \text{when } \bar{Y} \le -1, & \hat{\theta}_{MAP} = -1 \\ \text{when } -1 < \bar{Y} < 1 & \hat{\theta}_{MAP} = \bar{Y} \\ \text{when } \bar{Y} \ge 1 & \hat{\theta}_{MAP} = 1 \end{cases}$

Numerically: first, set a range $[-1, 1]$ for θ
second, equally divide $[-1, 1]$ into $N+1$ values, $-1, -1+\frac{2}{N}, -1+\frac{4}{N}, \cdots, 1$.
third, calculate the $p(\theta|y)$ when $\theta = -1, -1+\frac{2}{N}, -1+\frac{4}{N}, \cdots, 1$
fourth, select the maximum of $p(\theta|y)$ where the value of θ is $\hat{\theta}_{MAP}$.

1.3 $f(y|\theta) = g(y-\theta)$, where g is Exp(1) pdf, $p(\theta) \propto \mathbb{I}(\theta \in [0, 10])$

The posterior distribution is

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$
$$\propto \prod_{i=1}^{n} e^{-(Y_i-\theta)}\, \mathbb{I}(\theta \in [0, 10])$$
$$\propto e^{-\sum_{i=1}^{n}(Y_i-\theta)}\, \mathbb{I}(\theta \in [0, 10])$$
$$\propto e^{-n(\bar{Y}-\theta)}\, \mathbb{I}(\theta \in [0, 10]) \quad \text{where } \bar{Y} = \sum_{i=1}^{n} Y_i$$

① the support of posterior distribution is $[0, 10]$

② $\hat{\theta}_{MAP} = \underset{\theta}{\arg\max}\, p(\theta|y)$

Analytically. ~~[illegible]~~ Since $p(\theta|y) = e^{n(\theta-\bar{Y})}\, \mathbb{I}(\theta \in [0, 10])$ is monotonously increasing by θ, when θ is 10, $p(\theta|y)$ reaches the maximum.
So, $\hat{\theta}_{MAP} = 10$

Numerically: first, set a range $[0, 10]$ for θ
second, equally divide $[0, 10]$ into $N+1$ values, $0, \frac{10}{N}, \frac{20}{N}, \cdots, 10$.
third, calculate the $p(\theta|y)$ when $\theta = 0, \frac{10}{N}, \frac{20}{N}, \cdots, 10$.
fourth, select the maximum of $p(\theta|y)$ where the value of θ is $\hat{\theta}_{MAP}$.

1.4 $f(y|\theta) = \theta^{-1}\, \mathbb{I}(y \in [0, \theta])$, $p(\theta) \propto 1$

The posterior distribution is
$$p(\theta|y) \propto p(y|\theta)p(\theta)$$
$$\propto \prod_{i=1}^{n} \theta^{-1}\, \mathbb{I}(y_i \in [0, \theta]) \cdot 1$$
$$\propto \theta^{-n}\, \mathbb{I}(\theta \ge \max(Y_i)) \quad \text{where } 0 \le y_i \le \theta, \ i = 1, 2, \cdots, n$$

① the support of posterior distribution is ~~$(0,+\infty)$~~ $[\max\{Y_i, i=1,2,\cdots,n\}, +\infty)$

② $\hat{\theta}_{MAP} = \arg\max_{\theta} p(\theta|y)$

Analytically, since $p(\theta|y) \propto \theta^{-n}$ is monotonously decreasing by θ, when θ is $\max\{Y_i, i=1,2,\cdots,n\}$ ~~infinite small positive value~~, $p(\theta|y)$ reaches the maximum.

Numerically, first, set a range $(0, \max\{Y_i, i=1,2,\cdots,n\}]$ for θ ~~since $p(\theta|y)$ is monotonously decreasing by θ~~,
second, equally divide the range into N values, $\frac{\max\{Y_i\}}{N}, \frac{2\max\{Y_i\}}{N}, \cdots, \max\{Y_i\}$
third, calculate the $p(\theta|y)$ when $\theta = \frac{\max\{Y_i\}}{N}, \frac{2\max\{Y_i\}}{N}, \cdots, \max\{Y_i\}$.
forth, select the maximum of $p(\theta|y)$ where the value of θ is $\hat{\theta}_{MAP}$.

1.5 $f(y|\theta) \propto \phi(y-\theta)\cdot \mathbb{I}(|y-\theta|\le 1)$, where $\phi(\cdot)$ is the pdf of $N(0,1)$; $p(\theta)\propto 1$.

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$
$$\propto \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(Y_i-\theta)^2}{2}\right\}\mathbb{I}(|Y_i-\theta|\le 1)\cdot 1$$
$$\propto (2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{\sum_{i=1}^{n}(Y_i-\theta)^2}{2}\right\}\prod_{i=1}^{n}\mathbb{I}(|Y_i-\theta|\le 1)$$
$$\propto (2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{n\theta^2-2n\bar{Y}+\sum_{i=1}^{n}Y_i^2}{2}\right\}\prod_{i=1}^{n}\mathbb{I}(|Y_i-\theta|\le 1)$$
$$\propto \exp\left\{-\frac{(\theta-\bar{Y})^2}{2/n}+\frac{n\bar{Y}^2-\sum_{i=1}^{n}Y_i^2}{2}\right\}\prod_{i=1}^{n}\mathbb{I}(|Y_i-\theta|\le 1)$$
~~$\propto \exp\left\{-\frac{(\theta-\bar{Y})^2}{2/n}\right\}$~~ where $\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$, $|Y_i-\theta|\le 1$, $i=1,2,\cdots,n$

$$|Y_i-\theta|\le 1 \Rightarrow Y_i-1\le\theta\le Y_i+1$$
$$\propto \exp\left\{-\frac{(\theta-\bar{Y})^2}{2/n}\right\}\mathbb{I}(\max\{Y_i\}-1\le\theta\le\min\{Y_i\}+1) \qquad \max\{Y_i\}-1\le\theta\le\min\{Y_i\}+1$$

① the support of posterior distribution is $[\max\{Y_i\}-1, \min\{Y_i\}+1]$, $i=1,2,\cdots,n$

② $\hat{\theta}_{MAP} = \arg\max_{\theta} p(\theta|y)$

Analytically,
- when $\bar{Y}\le\max\{Y_i\}-1$, $\hat{\theta}_{MAP}=\max\{Y_i\}-1$
- when $\max\{Y_i\}-1<\bar{Y}<\min\{Y_i\}+1$, $\hat{\theta}_{MAP}=\bar{Y}$
- when $\bar{Y}\ge\min\{Y_i\}+1$, $\hat{\theta}_{MAP}=\min\{Y_i\}+1$

Numerically, first, set a range $[\max\{Y_i\}-1, \min\{Y_i\}+1]$ for θ
second, equally divide the range into $N+1$ values, $\max\{Y_i\}-1, \max\{Y_i\}-1+\frac{2+\min\{Y_i\}-\max\{Y_i\}}{N}, \cdots, \min\{Y_i\}+1$
third, calculate the $p(\theta|y)$ when $\theta=$ the $N+1$ values accordingly
fourth, select the maximum of $p(\theta|y)$ where the value of θ is $\hat{\theta}_{MAP}$.

1.6. $f(y|\theta) \propto \phi(y) \cdot \mathbb{I}(\theta_1 \le y \le \theta_2)$, $p(\theta) \propto 1$ and $\Theta = \{(\theta_1, \theta_2) \in \mathbb{R}^2 : \theta_1 < \theta_2\}$. $\phi(\cdot) \sim N(0,1)$

the posterior distribution is:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

$$\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{Y_i^2}{2}\right] \mathbb{I}(\theta_1 \le Y_i \le \theta_2) \cdot 1$$

$$\propto (2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{\sum_{i=1}^{n} Y_i^2}{2}\right\} \prod_{i=1}^{n} \mathbb{I}(\theta_1 \le Y_i \le \theta_2)$$

$$\propto \mathbb{I}(\theta_1 \le \min\{Y_i, i=1,2,\cdots,n\} \le \max\{Y_i, i=1,2,\cdots,n\} \le \theta_2)$$

① the support of posterior distribution is $\{(\theta_1, \theta_2) | \theta_1 \le \min\{Y_i\} \le \max\{Y_i\} \le \theta_2\}$

② $\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\, p(\theta|y)$

Analytically, $\hat{\theta}_{1MAP} \le \min\{Y_i, i=1,2,\cdots,n\} < \theta_2$.

$\hat{\theta}_{2MAP} \ge \max\{Y_i, i=1,2,\cdots,n\} > \theta_1$

Numerically:
- first, set a range $(-\infty, \min\{Y_i\}]$ for θ_1, $[\max\{Y_i\}, +\infty)$ for θ_2
- second, equally divide the two ranges into $N*N$ groups.
- third, calculate the values of $p(\theta|y)$ when θ_1, θ_2 = the $N*N$ groups.
- fourth, select the maximum of $p(\theta|y)$ where the values of θ_1, θ_2 are $\hat{\theta}_{1MAP}, \hat{\theta}_{2MAP}$

In summary, the support of posterior density not only depends on the prior, but also relies on the likelihood $p(y|\theta)$ when the values of $Y_i, i=1,2,\cdots,n$ are restricted/limited by θ.

Problem 2: Exploring frequentist and Bayesian asymptotics: the good, the bad and the ugly.

In the frequentist case, the MLE of θ for an experiment is

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

In the Bayesian case, the expression of the posterior distribution $p(\theta|y)$ can be written as

$$p(\theta|y) \propto p(y|\theta)p(\theta) \propto \theta^{\sum_{i=1}^{n} y_i}(1-\theta)^{n-\sum_{i=1}^{n} y_i}\mathrm{I}(\theta \epsilon \Theta)$$

2.1 Scenario 1: the parameter space $\boldsymbol{\Theta} = (\mathbf{0}, \mathbf{1})$

When the sample size $n = 4 * 4^i, i = 1,2,3,4$, then

(1) In the frequentist case:

The sampling distributions (histogram & density plot) under different sample sizes are shown in **Figure 1**, which approximately follow the Gaussian Distribution, particularly for large sample size. Furthermore, the larger the sampe size is, the more the MLE approaches to the truth (0.5). MLE is not a continuous rv.

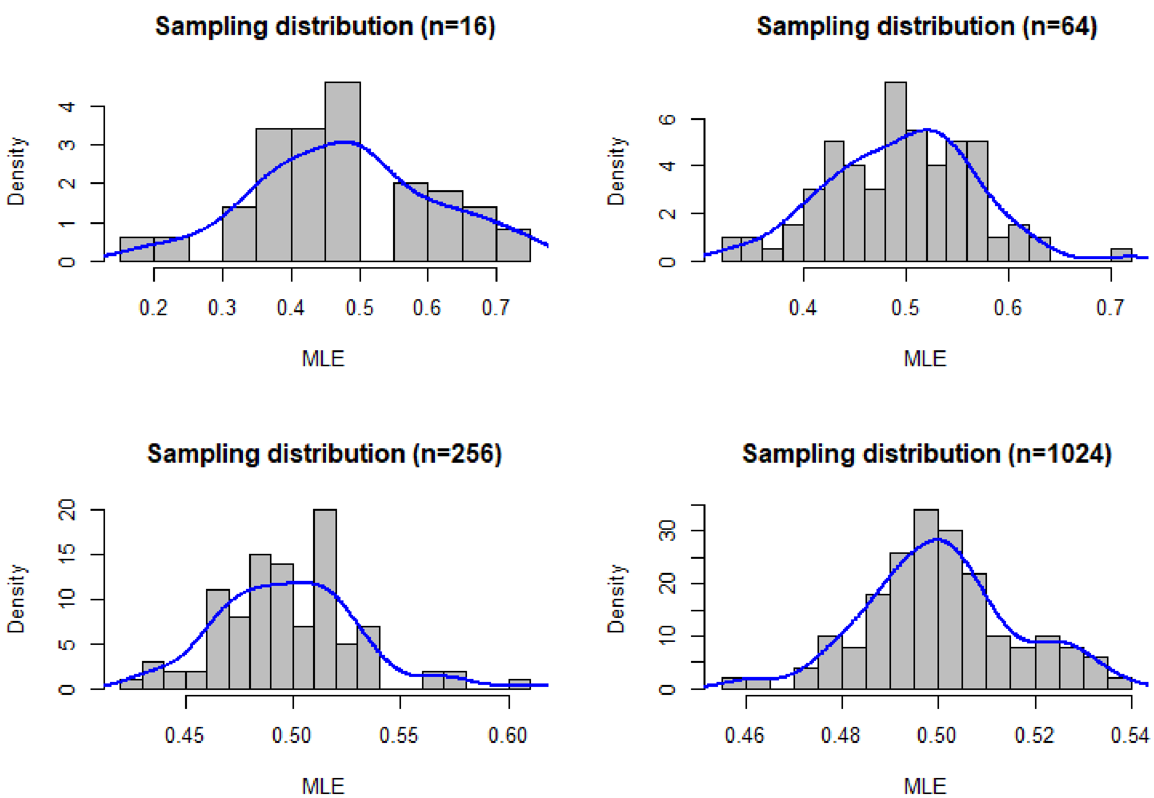

Figure 1: Plots of the sampling distribution of the MLE $\hat{\theta}$

(2) In the Bayesian case:

The posterior distributions under different sample sizes are shown in **Figure 2**, which follow the Beta Distribution (approximate Gaussian Distribution). Furthermore, the larger the sampe size is, the more the posterior estimate is concentrated to the truth (0.5). θ is a continuous rv.

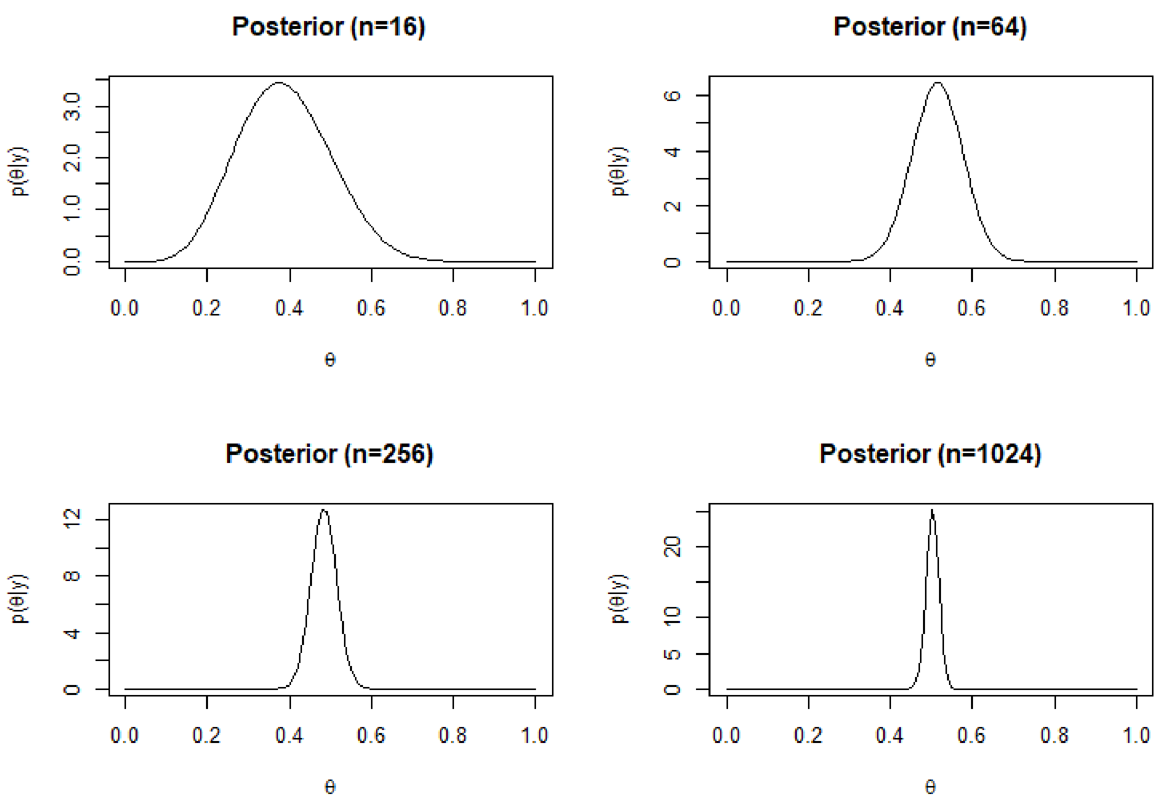

Figure 2: Plots of the posterior distribution $p(\theta|y)$

2.2 Scenario 2: the parameter space $\boldsymbol{\Theta = (0, 0.4]}$

When the sample size $n = 4 * 4^i, i = 1,2,3,4$, then

(1) In the frequentist case:

The sampling distributions (histogram & density plot) under different sample sizes are shown in **Figure 3**, which do not follow the Gaussian Distribution, particularly for large sample size. Furthermore, when the sampe size is less than 256, the MLE is equal to 0.4. When Θ=(0,0.4), MLE will gradually approach to 0.4. In contrast, when the sample size is 1024, we cannot find any efficient MLE based on the data samples. The MLE is not a continuous rv.

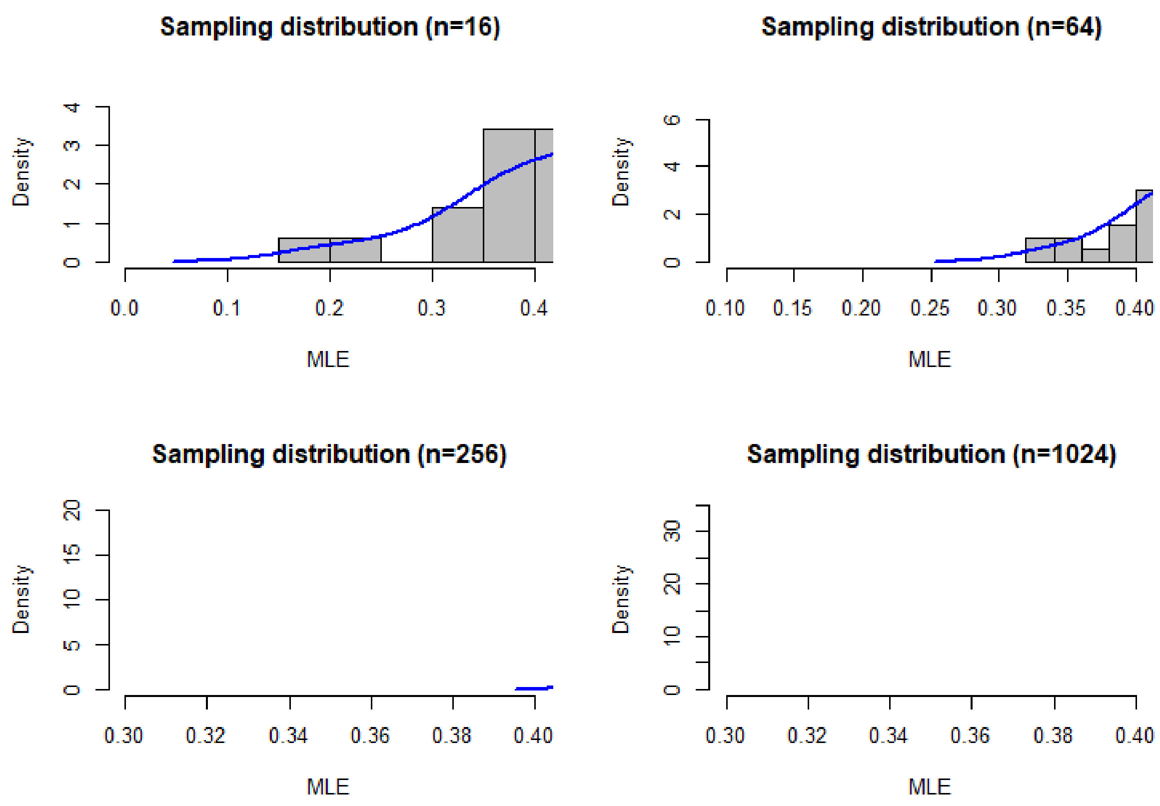

Figure 3: Plots of the sampling distribution of the MLE $\hat{\theta}$

(2) In the Bayesian case:

The posterior distributions under different sample sizes are shown in **Figure 4**, which do not follow the Gaussian Distribution, particularly for large sample size. When the sampe size is 16, the posterior estimate is equal to about 0.38; When the sampe size is higher than 16, the posterior estimate is equal to 0.4. Under Θ=(0,0.4), when the sampe size is 16, the posterior estimate is equal to about 0.38. When the sampe size is higher than 16, the posterior estimate gradually approaches to 0.4. θ is a continuous rv.

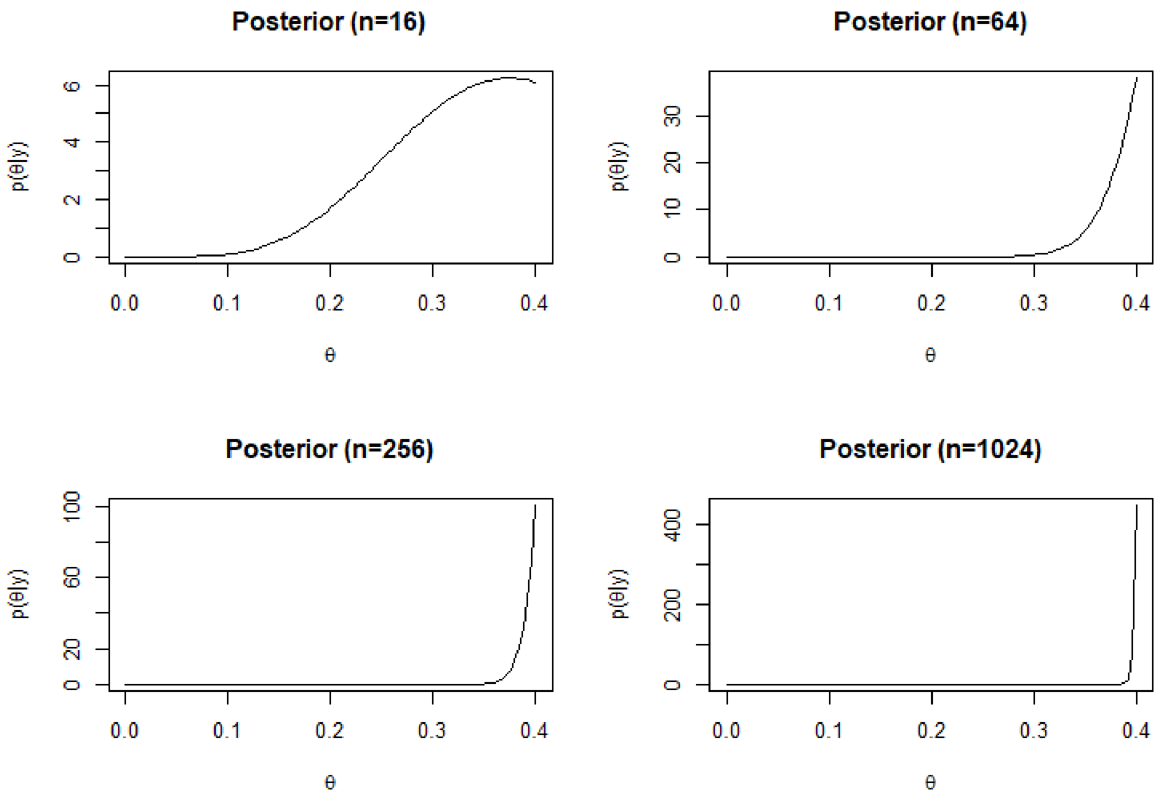

Figure 4: Plots of the posterior distribution $p(\theta|y)$

2.3 Scenario 3: the parameter space $\boldsymbol{\Theta = (0, 0.5]}$

When the sample size $n = 4 * 4^i, i = 1,2,3,4$, then

(1) In the frequentist case:

The sampling distributions (histogram & density plot) under different sample sizes are shown in **Figure 5**, which approximately follow single-sided Gaussian Distribution, particularly for large sample size. Furthermore, the larger the sampe size is, the more the MLE approaches to the truth (0.5). MLE is not a continuous rv.

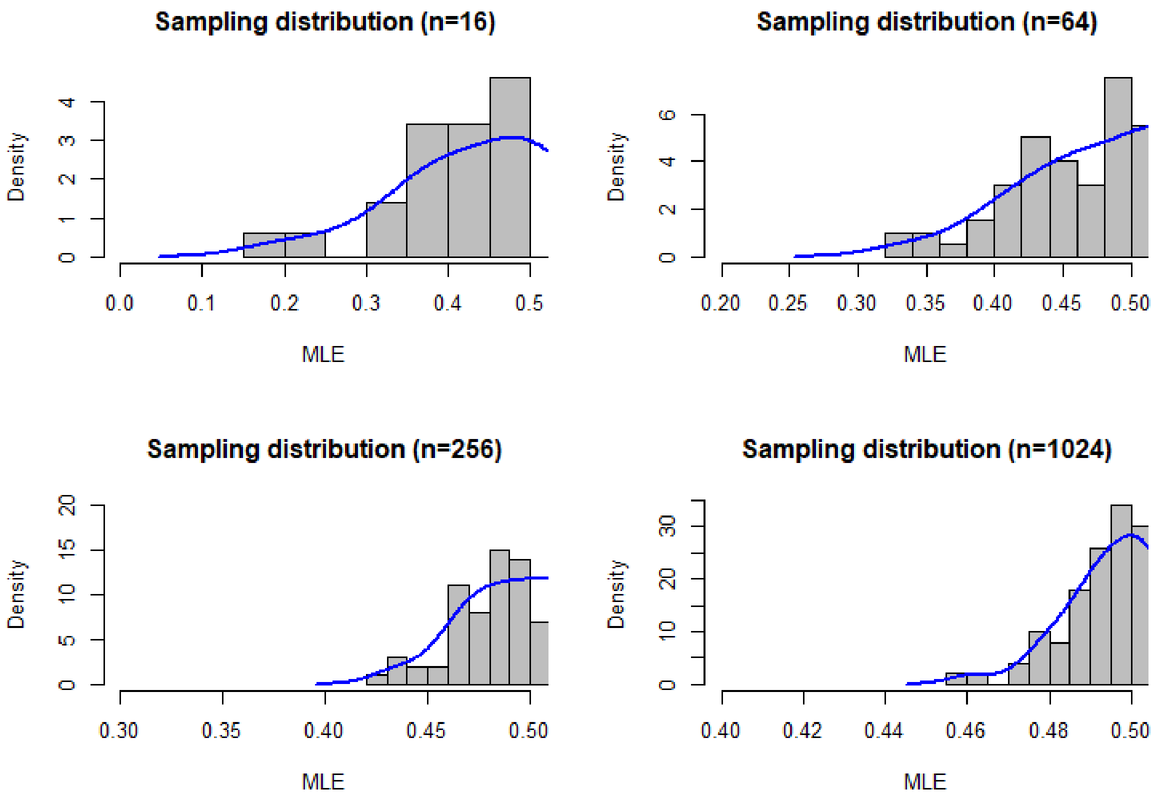

Figure 5: Plots of the sampling distribution of the MLE $\hat{\theta}$

(2) In the Bayesian case:

The posterior distributions under different sample sizes are shown in **Figure 6**, which follow the single-sided Beta Distribution (approximately single-sided Gaussian Distribution). Furthermore, the larger the sampe size is, the more the posterior estimate is concentrated to the truth (0.5). θ is a continuous rv.

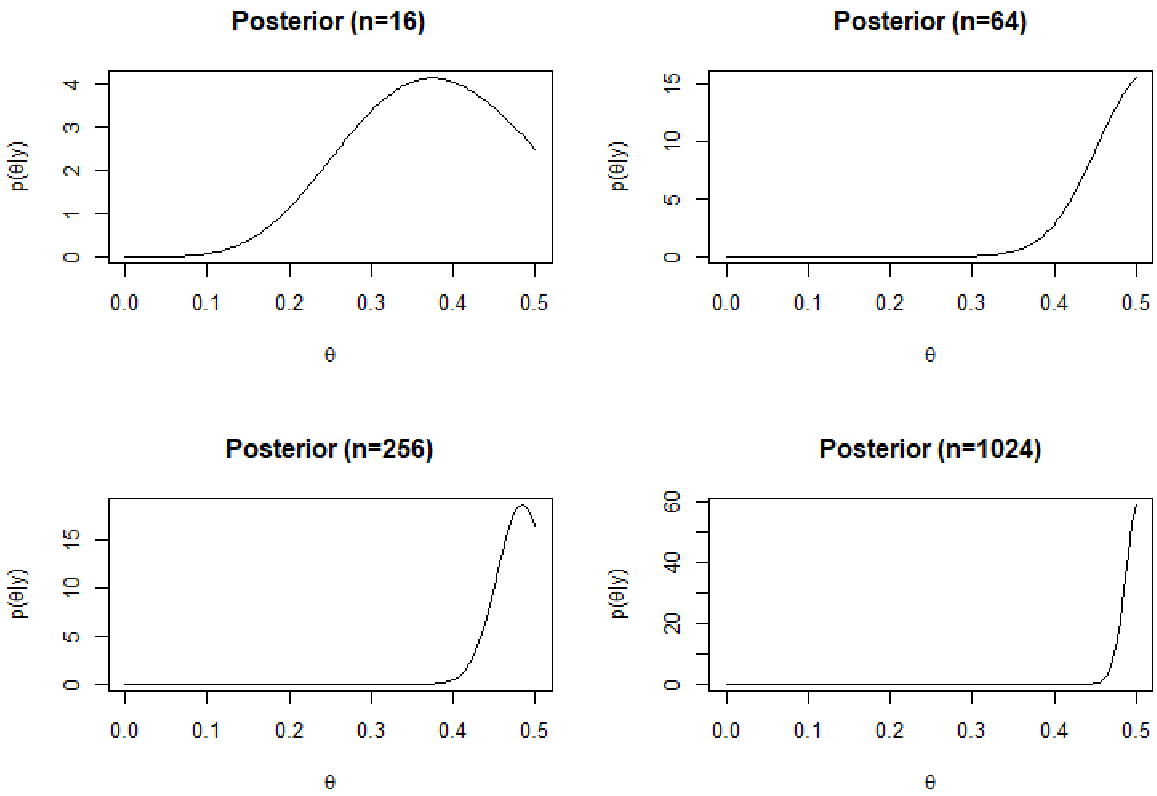

Figure 6: Plots of the posterior distribution $p(\theta|y)$

Problem 3: Black-box Gaussian approximation to a posterior.

Generate the data samples as follows:

set.seed(0); n=10; mut = 17; sigt = 1; y = rnorm(n, mut, sigt);

Then we use the generated 10 data samples to conduct the black-box Gaussian approximation to a posterior.

3.1 To implement the log-posterior as a function of (μ, τ), we set different values of (μ, τ) ranging from 16 to 18 and from -1 to 1, respectively, and then calculate the log-posterior under different (μ, τ) values, and the results is shown in **Figure 7**. We find that when $\mu = 17.4, \tau = 0.1$, the log-posterior value reaches the maximum. Furthermore, when the number of data samples gradually increases, the approximation to (μ, τ) approaches (17, 0).

```
## ----Problem 3.1--------------------------------------
mu_seq <- seq(16, 18, by=0.1)
log_sigma_seq <- seq(-1, 1, by=0.1)
log_posterior <- matrix(, nrow = length(mu_seq), ncol = length(log_sigma_seq))
for (i in 1:length(mu_seq)) {
  for (j in 1:length(log_sigma_seq)) {
    log_posterior[i,j] = log(prod(dnorm(y, mean=mu_seq[i], sd=exp(log_sigma_seq[j]))))
  }
}

persp(mu_seq, log_sigma_seq, log_posterior, xlab=expression(mu), ylab=expression(tau), zlab="p(mu,tau|y)",
      main='log-posterior', col='pink', shade=.4, theta = 30, phi = 15, ticktype='detailed')

column_max=as.integer(which.max(log_posterior)/length(log_sigma_seq))
row_max=which.max(log_posterior) - column_max*length(mu_seq)
mut_hat=mu_seq[row_max]
taut_hat=log_sigma_seq[column_max+1]
```

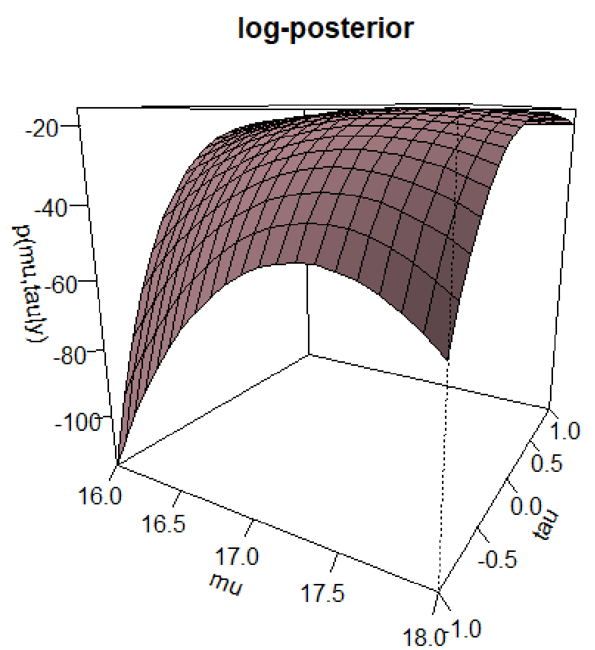

Figure 7: Plot of log-posterior as a function of (μ, τ)

3.2 Let $\tau = log\sigma$, the form of the log-posterior distribution is shown below:

$$logp(\mu,\tau|y) \propto -n*\tau - \frac{1}{2e^{2\tau}}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2 + n*(\bar{y}-\mu)^2\right]$$

Then we define the log-posterior as a function "obj" and use "optim" to find the posterior mode of μ,τ as 17.3589,0.1341.

```
# Approximation: "black box"
obj <- function(x) {-1*(-n*x[2] - 1/(2*exp(x[2])^2)*((n-1)*s_square + n*(mu_hat-x[1])^2))} # form of fun f
obj_origin <- function(x) {-n*x[2] - 1/(2*exp(x[2])^2)*((n-1)*s_square + n*(mu_hat-x[1])^2)}
init = c(16, 0.5)
out1 = optim(par=init, obj, lower=c(16,-1), upper=c(18,1), method="L-BFGS-B")
```

Then we use the second difference of the log-posterior to obtain the Hassien matrix with regard to μ,τ and make the inverse of the Hassien matrix to output the covariance matrix of μ,τ. Using the numerical method, the log-posterior follows:

$$p(\mu,\tau|y) \approx N\left(\begin{pmatrix}17.3589\\0.1341\end{pmatrix}\middle|\begin{pmatrix}0.5243 & -0.3621\\-0.3621 & 0.2005\end{pmatrix}\right)$$

```
eps = 10^(-3)
e_vector = matrix(c(1,0,0,1), nrow = 2, ncol =2)
Hess_fxx <- function(g, x, matrix, delta=eps) {
  Hess_mat <- c()
  for (i in 1:2) {
    for (j in 1:2) {
      Hess_mat = append(Hess_mat, (g(x+delta*matrix[i,])+g(x-delta*matrix[j,])-2*g(x))/delta^2)
    }
  }
  Hess_mat
}

myJ = Hess_fxx(obj_origin,out1$par,e_vector)
myJ = matrix(myJ, nrow=2, ncol=2)
black_mu_tau_var = solve(myJ)
```

3.3 Based on 3.2, we have obtained the "black-box" Gaussian approximation to a posterior. Then we use the expressions in the BDA3 example to obtain the "clear-box" Gaussian approximation as:

$$p(\mu,\tau|y) \approx N\left(\begin{pmatrix}17.3589\\0.1341\end{pmatrix}\middle|\begin{pmatrix}0.1453 & 0\\0 & 0.05\end{pmatrix}\right)$$

By comparing the "black-box" and "clear-box" Gaussian approximations, we find that the means of μ,τ are the same while their variances are much lower in "clear-box" Gaussian approximation than those in "black-box" Gaussian approximation. **Figure 8** further illustrates the finding. This also indicates that "black-box" Gaussian approximation to derivatives is sensitive to the sample mean. When the sample size gradually increases, the "black-box" Gaussian approximation will approach the "clear-box" Gaussina approximation, particularly for the covariance matrix of μ, τ.

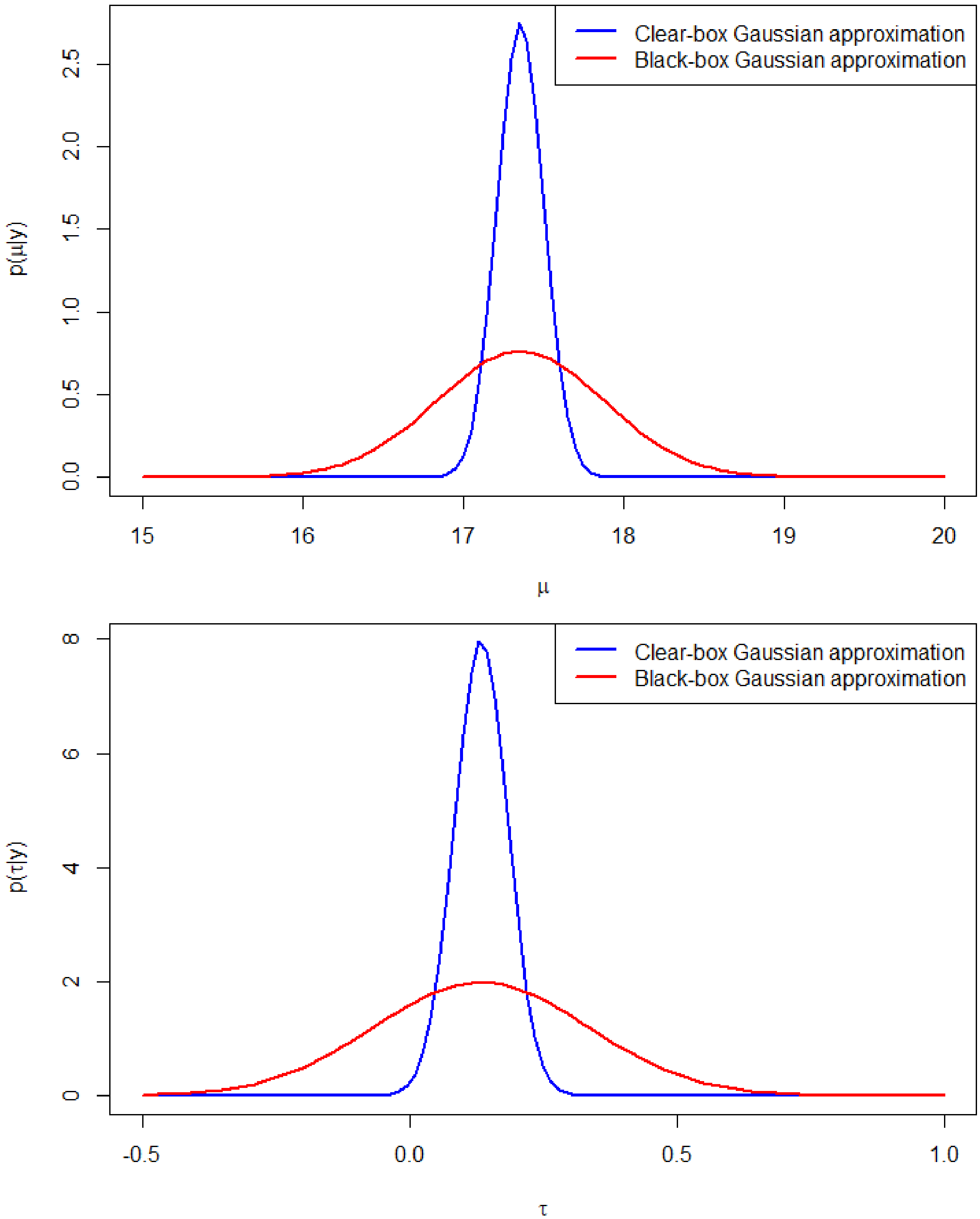

Figure 8: Normal approximation of μ, τ based on their log-posterior distribution

Problem 4: Shortest-length credible intervals, part II.

We need to write a function to compute the approximate shortest-length $100(1-\alpha)\%$ credible interval based on M iid samples from a posterior distribution pdf $p(\theta|y)$ where the expression of $p(\theta|y)$ is unknown. Here, we design the following algorithm about the M iid samples to return the shortest-length $100(1-\alpha)\%$ credible interval:

Step 1: Given M iid samples, we first estimate the density based on the M samples. Then we use the "cumsum" function and "equal-tailed" method to obtain the lower and upper bounds of $100(1-\alpha)\%$ credible interval.

```
## ----define function------------------------------------------------------------------
shortest_interval <- function(thetav, alpha) {
  # calculate density behind data samples
  prob = density(thetav)$y/(sum(density(thetav)$y))
  xx = density(thetav)$x
  lower = which(cumsum(prob) >= alpha/2)[1] # lower bound
  upper = which(cumsum(prob) >= 1-alpha/2)[1]-1 # upper bound
```

Step 2: Since the "equal-tailed" method cannot guarantee the shortest-length $100(1-\alpha)\%$ credible interval, we design the following algorithm to improve the lower and upper bounds such that the credible interval is the shortest. The idea behind this algorihm is to find the lower and upper bounds whose density is almost the same. Therefore, when the lower and upper bounds under "equal-tailed" condition have the same density, the corresponding credible interval is the shortest. If the density of lower and upper bounds is not the same, we use the following codes to find the lower and upper bounds almost with the same density as the shortest credible interval.

```
if (prob[lower] > prob[upper]) {
  for (i in 1:lower) {
    if ((prob[lower-i] - prob[upper-i]) < 10^-4) {break}
  }
  lower = lower-i+1
  upper = upper-i+1
  dist = sum(prob[lower:upper]) - (1-alpha)
  for (j in 0:i) {
    if (prob[lower+j] > prob[upper-j]) {dist = dist - prob[upper-j]; lower = lower-1}
    else {dist = dist - prob[lower+j]; upper = upper+1}
    if (dist < 0) {break}
  }
  lower = lower+j+1
  upper = upper-j-1
}
if (prob[lower] < prob[upper]) {
  for (i in 1:(length(prob)-upper)) {
    if ((prob[lower+i] - prob[upper+i]) > 10^-4) {break}
  }
  lower = lower+i-1
  upper = upper+i-1
  dist = sum(prob[lower:upper]) - (1-alpha)
  for (j in 0:i) {
    if (prob[lower+j] > prob[upper-j]) {dist = dist - prob[upper-j]; lower = lower-1}
    else {dist = dist - prob[lower+j]; upper = upper+1}
    if (dist < 0) {break}
  }
  lower = lower+j+1
  upper = upper-j-1
}
c(xx[lower], xx[upper])
```

Finally, we test the following three functions ($\alpha = 0.05$) to return the $100(1-\alpha)\%$ approximate credible intervals.

(1) Test problem (i): samples from $Beta(4,8)$ pdf

Table 1 Results of test problem (i)

M	Approximate shortest interval	Length	Exact shortest interval	Length	Ratio	Coverage
50	[0.1074,0.5860]	0.4786	[0.0934,0.5880]	0.4946	0.9677	0.9400
100	[0.1210,0.5895]	0.4685	[0.0934,0.5880]	0.4946	0.9472	0.9304
200	[0.1093,0.5785]	0.4691	[0.0934,0.5880]	0.4946	0.9486	0.9342
400	[0.0882,0.5812]	0.4930	[0.0934,0.5880]	0.4946	0.9968	0.9488
800	[0.0846,0.5858]	0.5012	[0.0934,0.5880]	0.4946	1.0133	0.9531
1600	[0.0820,0.5928]	0.5108	[0.0934,0.5880]	0.4946	1.0323	0.9579
3200	[0.0897,0.5935]	0.5038	[0.0934,0.5880]	0.4946	1.0186	0.9548
6400	[0.0925,0.5924]	0.4999	[0.0934,0.5880]	0.4946	1.0107	0.9528

Notes of columns:

1. value of M
2. the end points (L and U) of your approximate shortest interval
3. length of your approximate shortest interval
4. the end points (L and U) of your exact shortest interval
5. length of your exact shortest interval

6. ratio of the length of your approximate interval (in the numerator) to that of the exact (in the denominator)

7. coverage means the exact probability that θ is in your approximate credible interval

(2) Test problem (ii): samples from standard normal truncated to the interval (-4, 1).

Table 2 Results of test problem (ii)

M	Approximate shortest interval	Length	Exact shortest interval	Length	Ratio	Coverage
50	[-1.7233,1.1551]	2.8784	[-1.7267,1]	2.7267	1.0556	0.9496
100	[-1.6228,1.0872]	2.7100	[-1.7267,1]	2.7267	0.9939	0.9379
200	[-1.5648,1.0636]	2.6285	[-1.7267,1]	2.7267	0.9640	0.9301
400	[-1.7736,1.0624]	2.8360	[-1.7267,1]	2.7267	1.0401	0.9548
800	[-1.7890,1.0531]	2.8422	[-1.7267,1]	2.7267	1.0423	0.9563
1600	[-1.8699,1.0493]	2.9192	[-1.7267,1]	2.7267	1.0706	0.9625
3200	[-1.8093,1.0506]	2.8599	[-1.7267,1]	2.7267	1.0488	0.9582
6400	[-1.7940,1.0475]	2.8415	[-1.7267,1]	2.7267	1.0420	0.9568

It is worth mentioning that the coverage is calculated by the ratio of probability that θ is in the approximate credible interval to probability that θ is in the support as shown in HW1. Furthermore, the approximate credible interval is first truncated by the support of θ in HW1.

```
prob_2 = integrate(function(x) exp(-x^2/2)/sqrt(2*pi), x.min, x.max)$value
```

```
if (SI_2_i[2] > 1) {coverage_SI_2 = append(coverage_SI_2, integrate(function(x) exp(-x^2/2)/sqrt(2*pi), SI_2_i[1], 1)$value/prob_2)}
else {coverage_SI_2 = append(coverage_SI_2, integrate(function(x) exp(-x^2/2)/sqrt(2*pi), SI_2_i[1], SI_2_i[2])$value/prob_2)}
```

(3) Test problem (ii): samples from standard normal truncated to the interval (-1, 1).

Table 3 Results of test problem (iii)

M	Approximate shortest interval	Length	Exact shortest interval	Length	Ratio	Coverage
50	[-1.0716,0.9784]	2.0500	[-0.9318,0.9318]	1.8636	1.1000	0.9923
100	[-1.1264,0.9881]	2.1145	[-0.9318,0.9318]	1.8636	1.1347	0.9958
200	[-1.0516,0.9904]	2.0420	[-0.9318,0.9318]	1.8636	1.0957	0.9966
400	[-1.0421,0.9759]	2.0180	[-0.9318,0.9318]	1.8636	1.0829	0.9914
800	[-1.0042,0.9625]	1.9666	[-0.9318,0.9318]	1.8636	1.0553	0.9864
1600	[-0.9646,0.9570]	1.9216	[-0.9318,0.9318]	1.8636	1.0311	0.9717
3200	[-0.9459,0.9441]	1.8900	[-0.9318,0.9318]	1.8636	1.0142	0.9599
6400	[-0.9440,0.9426]	1.8867	[-0.9318,0.9318]	1.8636	1.0124	0.9587

In summary, when the sample is fixed, the distribution of data samples and the coarseness of grid involved in density estimation certainly influence the accuracy of the implementation to find the shortest-length credible interval. When the sample size increases from 50 to 6400, the approximate credible interval gradually approaches the exact one, in terms of lower and upper bounds, length, and coverage, particularly for coverage that is closer to 0.95, which is a good example of data asymptotics, large-sample estimates.