

Bayesian Learning Lab 1

Qinyuan Qi(qinqi464)

Satya Sai Naga Jaya Koushik Pilla (satpi345)

2024-04-11

1. Daniel Bernoul

We have $y_1, \dots, y_n | \theta \sim \text{Bern}(\theta)$, and have obtained a sample with $s = 22$ successes in $n = 70$ trials. A $\text{Beta}(\alpha_0, \beta_0)$ prior for θ and let $\alpha_0 = \beta_0 = 8$ were also given.

(a) Draw random numbers from the posterior $\theta | y = \text{Beta}(\alpha_0 + s, \beta_0 + f)$, and verify that the posterior mean and sd converge to the true mean and sd as the number of draws increases.

First of all, we need to calculate the true mean and sd of Beta distribution, the following are the formulas.

$$E[\theta] = \frac{\alpha}{\alpha + \beta} = \frac{\alpha_0}{\alpha_0 + s + \beta_0 + f}$$

$$sd[\theta] = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}} = \sqrt{\frac{(\alpha_0 + s)(\beta_0 + f)}{(\alpha_0 + s + \beta_0 + f)^2(\alpha_0 + s + \beta_0 + f + 1)}}$$

Then we draw 10000 random values and calculate the mean and sd of the posterior distribution.

We found that the mean and sd of the posterior distribution did not converge to the true mean and sd as the number of draws increased but within a close range near the true values.

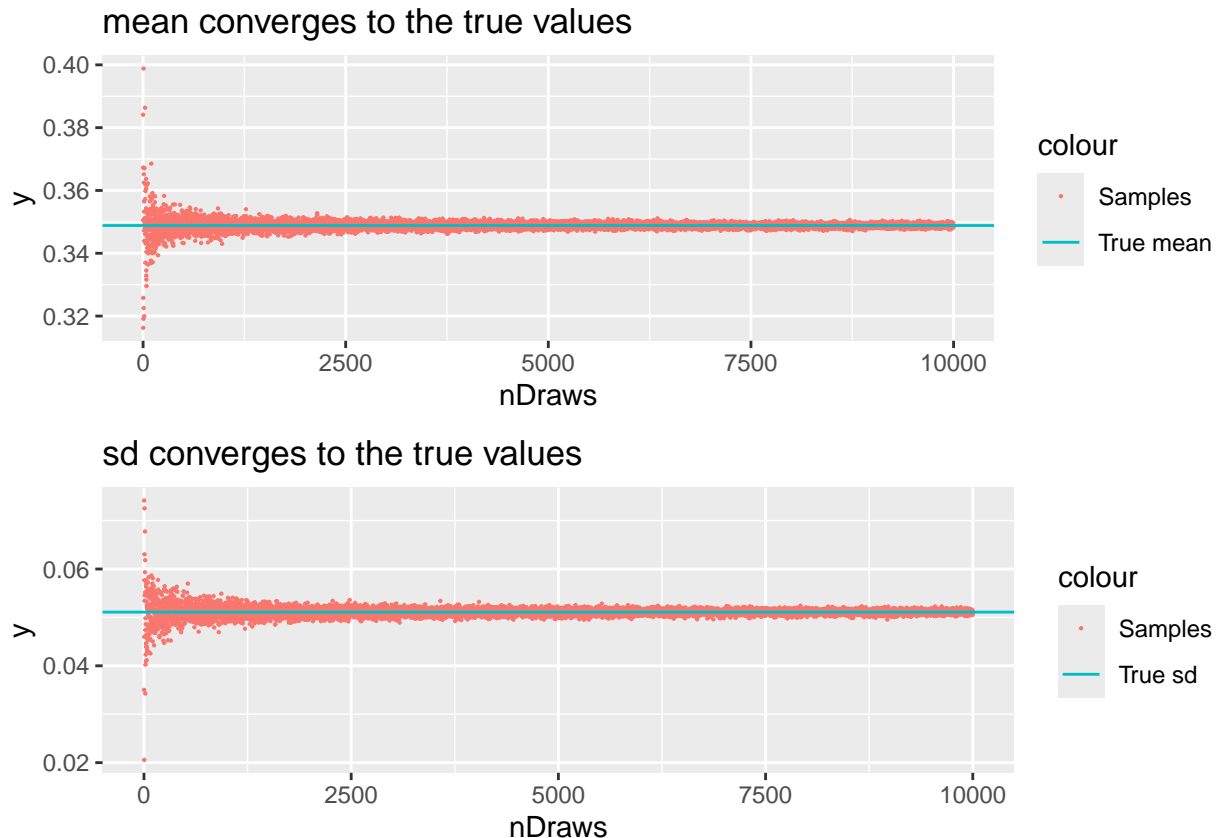
```
#-----  
# Code for question 1a  
#-----  
# given values  
nDraws_1a <- 10000  
alpha0 <- 8  
beta0 <- 8  
s <- 22  
n <- 70  
f <- n - s  
alpha_new <- alpha0 + s  
beta_new <- beta0 + f  
  
# calculate the true mean and standard deviation  
true_mean <- alpha_new/(alpha_new + beta_new)  
true_var <- (alpha_new*beta_new) / ((alpha_new + beta_new)^2 *  
  (alpha_new + beta_new + 1))  
true_sd <- sqrt(true_var)  
  
set.seed(12345)
```

```

# init the vectors to store the means and sds
means <- c()
sds <- c()

# calculate means and sds
for (i in 1:nDraws_1a) {
  beta_draws <- rbeta(n = i, shape1 = alpha_new, shape2 = beta_new)
  means[i] <- mean(beta_draws)
  sds[i] <- sd(beta_draws)
}

```



(b) Draw 10000 random values from the posterior to compute the posterior probability $Pr(\theta > 0.3|y)$ and compare with the exact value from the Beta posterior

```

#-----
# Code for question 1b
#-----
set.seed(12345)

nDraws_1b <- 10000
sample_b <- rbeta(n = nDraws_1b, shape1 = alpha_new, shape2 = beta_new)
prob_sample_b <- length(sample_b[which(sample_b > 0.3)]) / nDraws_1b
exact_value_1b <- 1 - pbeta(q = 0.3, shape1 = alpha_new, shape2 = beta_new)

```

Code as above, and we have the following values listed in the table. The expected value is the exact value of the posterior probability $Pr(\theta > 0.3|y)$, and the simulated value is the value we calculated from the 10000

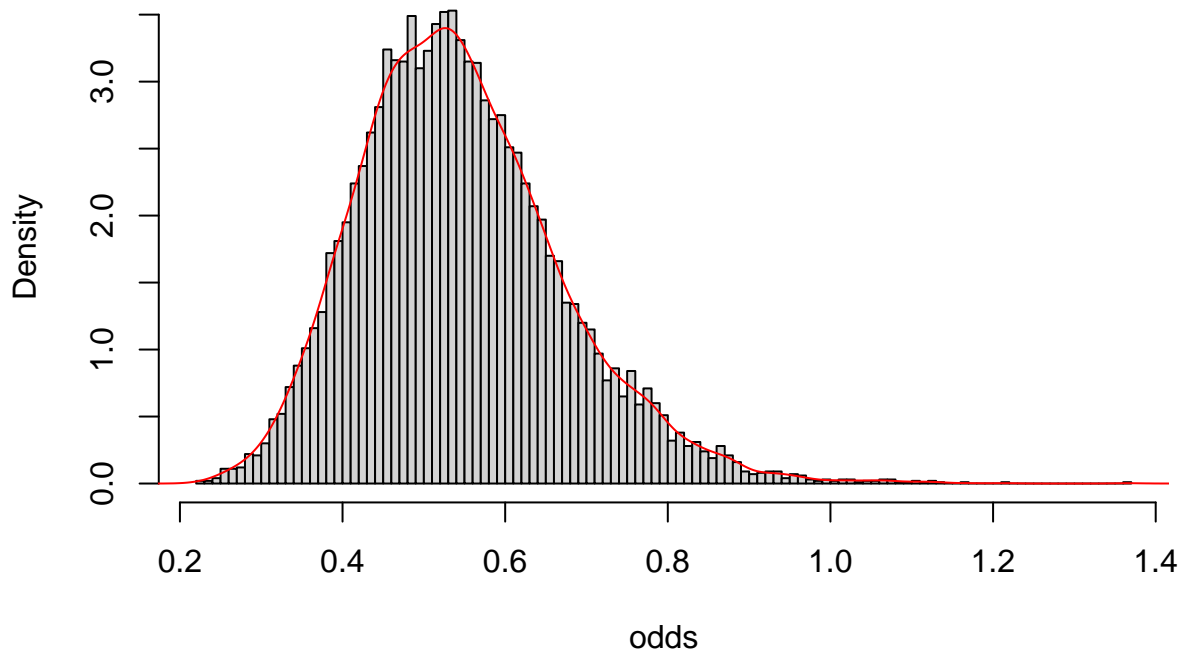
random values. And the expected value is a little bigger than the simulated value. And the exact value is will be calculated as $Pr(\theta > 0.3|y) = 1 - Pr(\theta \leq 0.3|y)$.

Expected value	Simulated vaule
0.8286	0.8294

(c) Draw 10000 random values from the posterior of the odds $\frac{\theta}{1-\theta}$ and plot the posterior distribution θ .

```
#-----
# Code for question 1c
#-----
set.seed(12345)
nDraws_1c <- 10000
sample_1c <- rbeta(n = nDraws_1c, shape1 = alpha_new, shape2 = beta_new)
odds <- sample_1c/(1-sample_1c)
```

Histogram of odds



2. Log-normal distribution and the Gini coefficient

We asked 8 randomly selected persons about their income(in thousand SEK) and got the following data.

(33, 24, 48, 32, 55, 74, 23, 17)

We know the model for non-negative continuous variables is the log-normal distribution $\log N(\mu, \sigma^2)$ which has the following density function.

$$p(y|\mu, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(y) - \mu)^2}{2\sigma^2}\right)$$

where $y > 0, -\infty < \mu < \infty, \sigma^2 > 0$

And we also know that if $y \sim \text{logN}(\mu, \sigma^2)$, $\mu = 3.6$, but σ unknown but with a prior $p(\sigma^2) \sim 1/\sigma^2$.

The posterior for σ^2 is the $\text{Inv} - \chi^2(n, \tau^2)$ distribution where .

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

(a) Draw 10000 random values from the posterior of σ^2 by assuming $\mu = 3.6$ and plot the posterior distribution

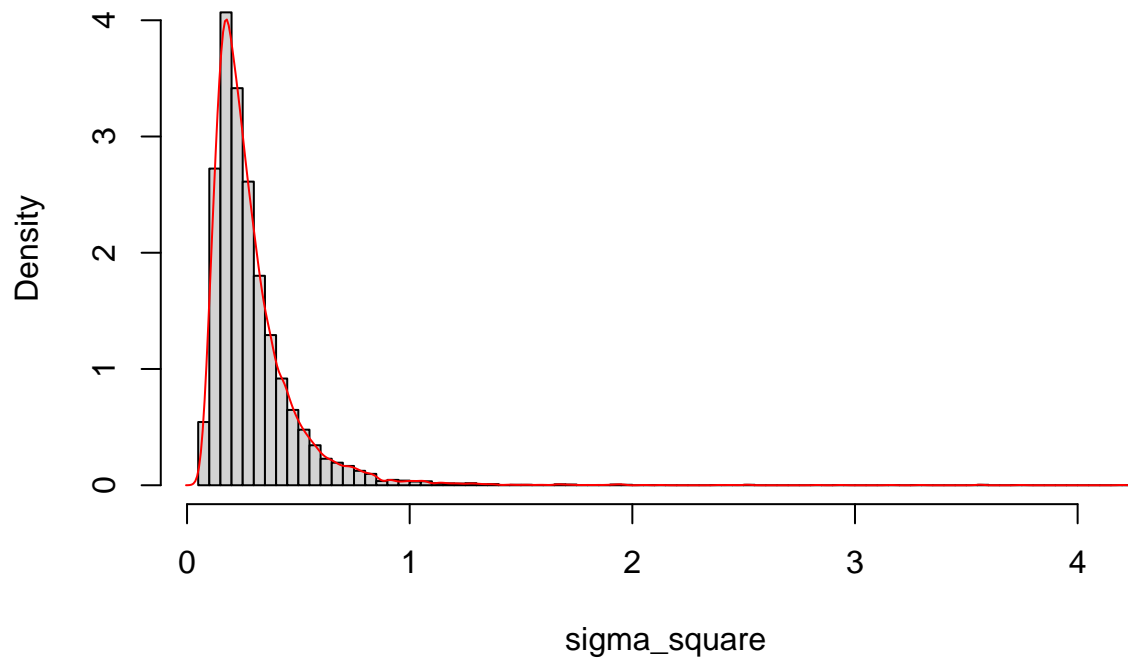
According to the formula of posterior for σ^2 , we have the following code.

```
#-----
# Code for question 2a: calc tau value
#-----
set.seed(12345)

# given values
observations <- c(33, 24, 48, 32, 55, 74, 23, 17)
n_observations <- length(observations)
nDraws_2a <- 10000
mu <- 3.6
nDraws_2b <- 10000

# calculate tau square value and sigma_square
tau_square <- sum((log(observations)-mu)^2) / n_observations
X <- rchisq(nDraws_2b, n_observations)
sigma_square <- n_observations * tau_square / X
```

Histogram of sigma_square

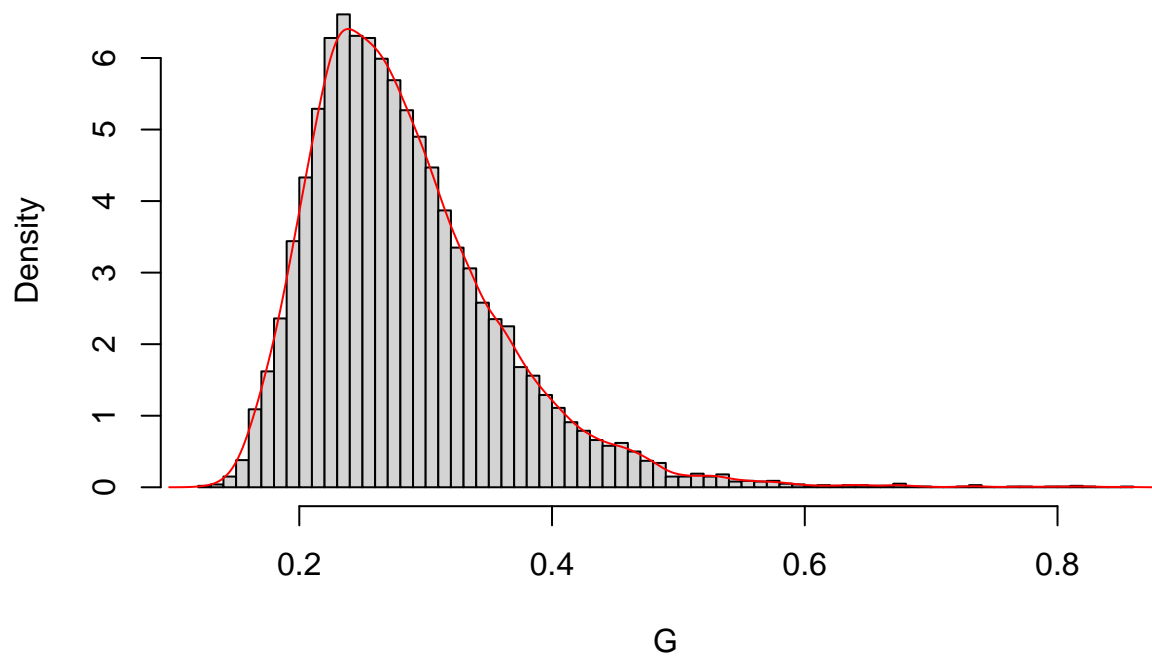


(b) Use the posterior draws in a) to compute the posterior distribution of the Gini coefficient G for the current data set

From the statement of the question, we know that the Gini coefficient $G = 2\phi(\sigma/\sqrt{2}) - 1$, and ϕ is the CDF of the Normal distribution with mean zero and unit variance, so $\mu = 0, sd = 1$. we calculate $\phi(\sigma/\sqrt{2})$ first then calculate the $G = 2\phi - 1$, then we plot Gini coefficient value G .

```
#-----  
# Code for question 2b: compute the posterior distribution of the Gini  
# coefficient  
#-----  
phi <- pnorm(sqrt(sigma_square/2), mean = 0, sd = 1)  
G <- 2 * phi - 1
```

Histogram of G

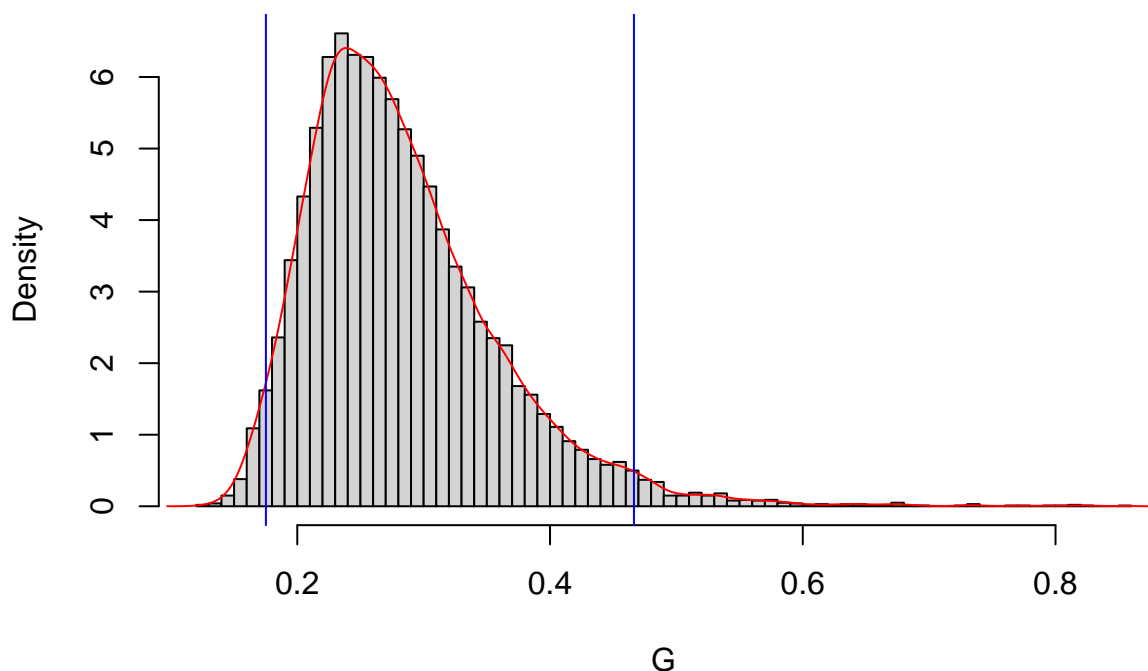


(c) Use the posterior draws from b) to compute a 95% equal tail credible interval for G.

95% equal tail credible interval for G is the interval between 0.025 and 0.975 of the distribution of G. We use the quantile method to calculate the interval and plot the histogram of G.

```
#-----  
# Code for question 2c: compute a 95% equal tail credible interval for G  
#-----  
credible_interval <- quantile(G, c(0.025, 0.975))
```

Histogram of G



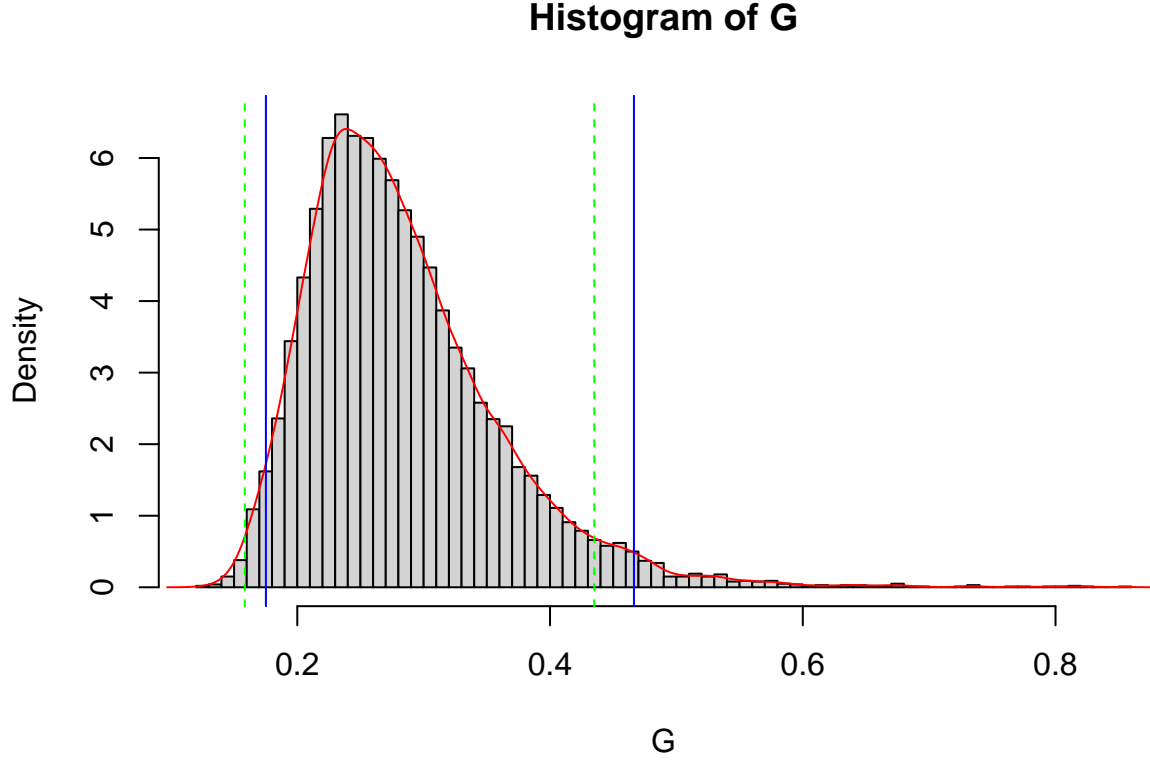
The 95% equal tail credible interval for G is: (0.1751706, 0.4664171)

(d) Use the posterior draws from b) to compute a 95% Highest Posterior DensityInterval (HPDI) for G. Compare the two intervals in (c) and (d)

First, we calculate the kernel density estimate of the posterior of G using the density function, then make a data frame using density values and x values. Then we order it by y using descending order, calculate the cumulative sum, and filter the x value using the 95% HPDI. Based on the HPDI, we plot the histogram of G and the 95% HPDI interval.

```
#-----
# Code for question 2d: compute a 95% Highest Posterior
# DensityInterval (HPDI) for G
#-----
density_G <- density(G)

# find the HPDI
# create the density data frame
density_G_df <- data.frame("x" = density_G$x, "y" = density_G$y)
# order it by y using descending order
density_G_df <- density_G_df[order(density_G_df$y, decreasing = TRUE),]
# calculate the cumulative sum
density_G_df$cumsum <- cumsum(density_G_df$y)
# find the 95% HPDI
HPDI <- density_G_df$x[density_G_df$cumsum < 0.95 * sum(density_G_df$y)]
hdpi_low <- min(HPDI)
hdpi_high <- max(HPDI)
```



95% HPDI interval for G is (0.1584821, 0.435012) which is green lines in the plot. The 95% equal tail credible interval for G is: (0.1751706, 0.4664171) which is blue lines in the plot.

Comparing the two intervals and plots above, we find that equal tail credible intervals cut off 2.5% of samples from each tail, while HDPI cuts off the data points that have the lowest density and keeps 95% of the data points. Since this distribution is not symmetrical, the HPDI interval is a better way to calculate the 95% interval in this case.

3. Bayesian inference for the concentration parameter in the von Mises distribution

We have observed wind directions at a given location on ten different days as follows.

$$(20, 314, 285, 40, 308, 314, 299, 296, 303, 326)$$

Then those 10 angle-based(north is 0 degree) data are converted to radians-based data as follows.

$$(-2.79, 2.33, 1.83, -2.44, 2.23, 2.33, 2.07, 2.02, 2.14, 2.54)$$

According to the statement of the question, those data points conditional on (μ, k) are independent observations from the von Mises distribution as follows.

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

Where $I_0(k)$ is the modified Bessel function of the first kind of order 0. The parameter $\mu(-\pi \leq \mu \leq \pi)$ is the mean direction and $k > 0$ is the concentration parameter, and have $\mu = 2.4$.

We also have $k \sim \text{Exponential}(\lambda = 0.5)$ a priori, where λ is the rate parameter of the exponential distribution.

(a) Derive the expression for what the posterior $p(k|y, \mu)$ is proportional to. Hence, derive the function $f(k)$ such that $p(k|y, \mu) \propto f(k)$. Then, plot the posterior distribution of k for the wind direction data over a fine grid of k values.

We know that

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

So we derive the Likelihood formula of von Mises distribution as follows.

$$\begin{aligned} & \prod_{i=1}^n \frac{\exp[k \cdot \cos(y_i - \mu)]}{2\pi I_0(k)} \\ &= \left(\frac{1}{2\pi I_0(k)}\right)^n \exp\left[k \cdot \sum_{i=1}^n \cos(y_i - \mu)\right] \\ &= \frac{1}{(2\pi)^n} \frac{1}{(I_0(k))^n} \exp\left[k \cdot \sum_{i=1}^n \cos(y_i - \mu)\right] \end{aligned}$$

Since $\frac{1}{(2\pi)^n}$ is not related to k , we can ignore it.

We have $k \sim \text{Exponential}$ a priori, the PDF of Exponential distribution (when $\lambda = 0.5 > 0$) is.

$$f(x; \lambda) = \lambda e^{-\lambda x}$$

With $\lambda = 1/2$, we have the following prior.

$$\lambda \exp(-\lambda k) = \frac{1}{2} \exp\left(-\frac{1}{2}k\right)$$

Since $1/2$ is not related to k , we can ignore it.

So we have posterior as follows.

$$\begin{aligned} & \frac{1}{(I_0(k))^n} \exp\left[k \cdot \sum_{i=1}^n \cos(y - \mu)\right] \cdot \exp\left(-\frac{1}{2}k\right) \\ &= \frac{1}{(I_0(k))^n} \exp\left[k \cdot \sum_{i=1}^n \cos(y - \mu) - \frac{1}{2}k\right] \end{aligned}$$

So we write the following code to calculate the posterior distribution of k .

```
#-----
# Code for question 3a: calc the posterior
#-----
# given values
degrees <- c(20, 314, 285, 40, 308, 314, 299, 296, 303, 326)
n_degrees <- length(degrees)
radians <- c(-2.79, 2.33, 1.83, -2.44, 2.23, 2.33, 2.07, 2.02, 2.14, 2.54)
mu_3a <- 2.4
mu_sample <- mean(radians)
```

```

sd_sample <- sd(radians)

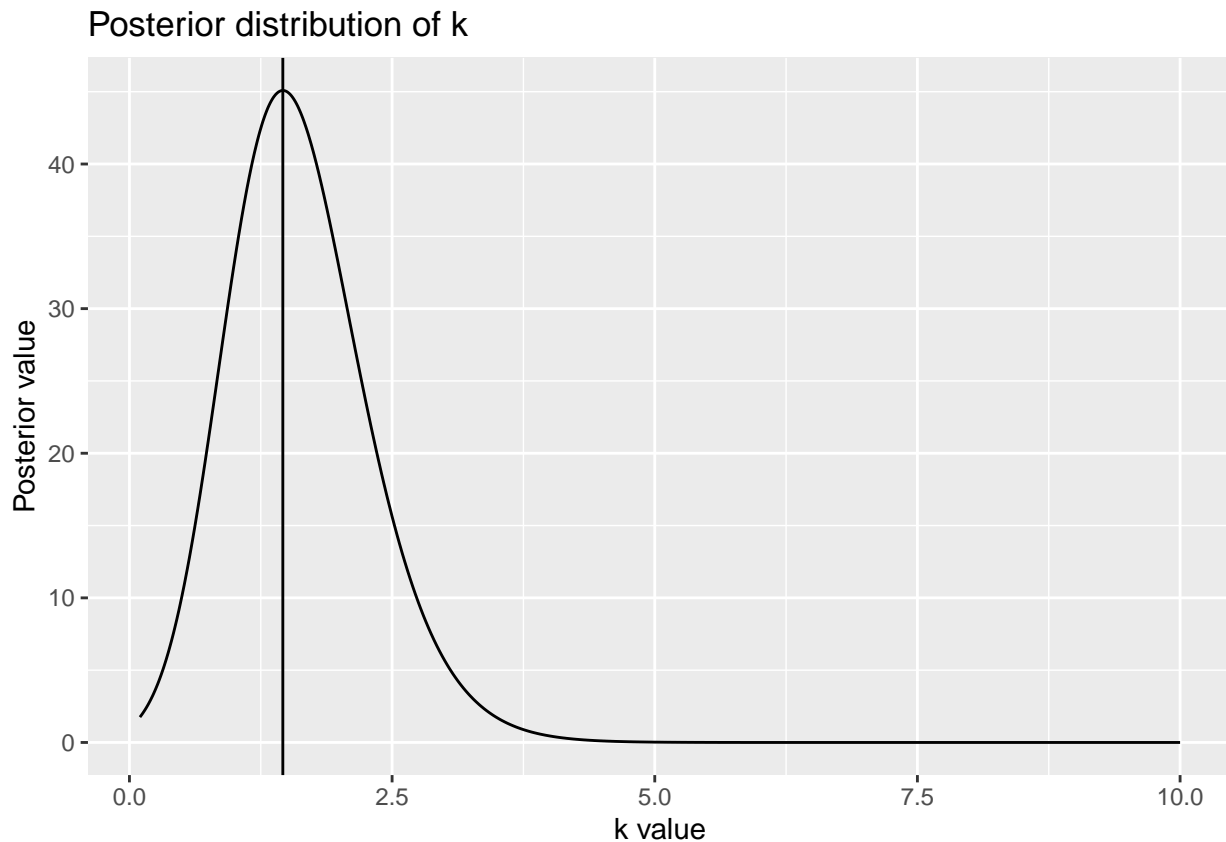
radians <- (radians-mu_sample)/sd_sample
mu_3a <- (mu_3a-mu_sample)/sd_sample
n_radians <- length(radians)

# grid of k values
k <- seq(from = 0.1,to = 10, by = 0.01) # k>0

# calculate the posterior
result_3a <- exp(k * sum(cos(radians - mu_3a)) - k / 2) /
  besseli(x = k, nu = 0) ^ n_radians
k_max_index <- which.max(result_3a)

```

Then we plot the posterior distribution of k for the wind direction data over a fine grid of k values 0 to 10 with step 0.1.



(b) Find the (approximate) posterior mode of k from the information in a.

```

#-----
# Code for question 3b: posterior mode of k
#-----
k_mode <- k[k_max_index]

```

According to the data in 3.b, we know that the k is 1.46.