

732A73 Bayesian Learning

Computer lab 1

Johannes Hedström & Mikael Montén

STIMA
Department of Computer and Information Science
Linköpings universitet

2024-05-24

Contents

1	Daniel Bernoulli	1
1.1	a	1
1.2	b	3
1.3	c	4
2	Log-normal distribution and the Gini coefficient.	6
2.1	a	6
2.2	b	7
2.3	c	8
2.4	d	8
3	Bayesian inference for the concentration parameter in the von Mises distribution	11
3.1	a)	11
3.2	b	13
4	Appendix	15

```
library(coda)
library(invgamma)
```

1 Daniel Bernoulli

Let $y_1, \dots, y_n | \theta \sim \text{Bern}(\theta)$, $s = 22, n = 70$. Assume $\text{Beta}(\alpha_0, \beta_0)$ prior for θ and $\alpha_0 = \beta_0 = 8$.

```
s <- 22 # successes
f <- 70 - 22 # failures
n <- 70 # trials

# Beta priors
A0 <- 8
B0 <- 8
```

1.1 a

Draw 10000 random values from the posterior and verify graphically that the posterior mean and standard deviation converges to the true values as the number of draws grows larger.

Posterior : $\theta | y \sim \text{Beta}(\alpha_0 + s, \beta_0 + f)$

Beta distribution has expected value $E[X] = \frac{\alpha}{\alpha + \beta}$ and variance $\text{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2 + (\alpha + \beta + 1)}$. Since we know the posterior we can deduce true mean and standard deviation to be the following;

$$\text{True posterior mean} = \frac{\alpha_0 + s}{\alpha_0 + s + \beta_0 + f} = \frac{\alpha_0 + s}{\alpha_0 + \beta_0 + n}$$

$$\text{True posterior sd} = \sqrt{\frac{(\alpha_0 + s) \cdot (\beta_0 + f)}{(\alpha_0 + \beta_0 + n)^2 (\alpha_0 + \beta_0 + n + 1)}}$$

* Draw 10000 random values from the posterior and calculate sample means & standard deviations as functions of accumulating number of drawn values

```
nDraws <- 10000 # nr of draws

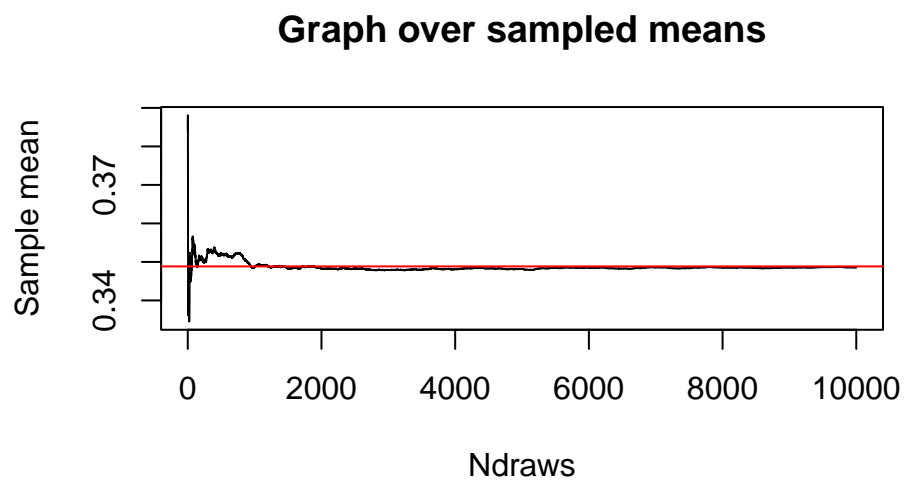
true_mean <- (A0 + s) / (A0 + B0 + n) # calculating true mean and sd for posterior dist
true_sd <- sqrt(((A0+s)*(B0+f))/((A0+B0+n)^2 * (A0+ B0+n+1)))

posterior_sample <- rbeta(nDraws, A0+s, B0+f) # draw values from posterior

# Calculate sample cumulative means and standard deviations to show convergence
sample_means <- cumsum(posterior_sample)/(1:nDraws)
sample_sds <- sqrt(cumsum((posterior_sample - sample_means)^2) / (1:nDraws))
```

- Verify graphically that the posterior mean converges

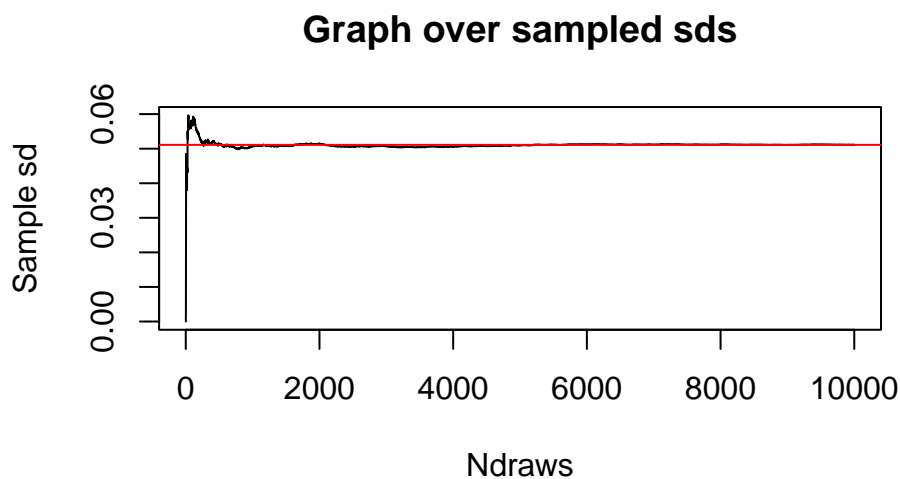
```
plot(sample_means,type='line',xlab='Ndraws',ylab='Sample mean', main='Graph over sampled means')  
abline(h=true_mean,col='red')
```



There is a short burn-in of around 1000 draws before the sample mean of θ stabilizes and it converges to the true posterior mean after around 2000 draws.

- Verify graphically that the posterior sd converges

```
plot(sample_sds,type='line',xlab='Ndraws',ylab='Sample sd', main='Graph over sampled sds')
abline(h=true_sd,col='red')
```



The sample standard deviation of θ stabilizes after around 1000 draws but doesn't fully converge to the true posterior standard deviation until around 5000 draws.

1.2 b

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

```
set.seed(12345)

# draws from posterior
post <- rbeta(nDraws, A0+s, B0+f)

# mean of samples over 0.3
prob <- mean(post>0.3)

prob_exact <- 1 - pbeta(0.3,A0+s, B0+f)

df <- data.frame('Posterior prob' =prob, 'Exact value from beta post' = prob_exact)
colnames(df) <- c('Posterior prob','Exact value from beta post')

knitr::kable(df)
```

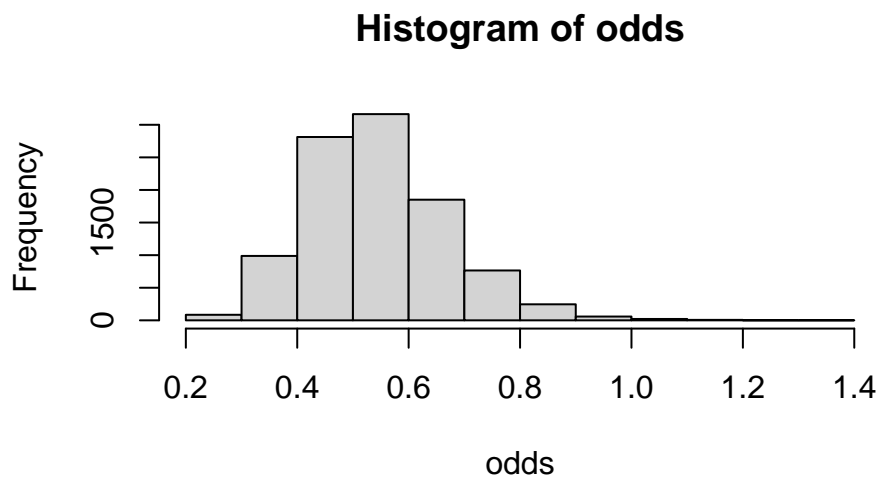
Posterior prob	Exact value from beta post
0.8294	0.8285936

The approximate posterior probability of $\theta > 0.3$ given y is 0.8294 which is very close to the exact value from the beta posterior 0.8286. The exact probability is found with the conjugate as pbeta calculates $P(\theta \leq 0.3)$ whereas we are interested in $P(\theta > 0.3)$.

1.3 c

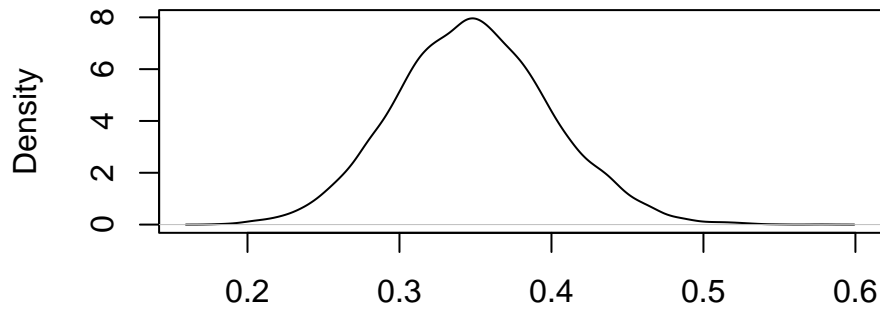
Draw 10000 random values from the posterior of the odds $\phi = \frac{\theta}{1-\theta}$ by using the previous random draws from the Beta posterior for θ and plot the posterior distribution of ϕ .

```
odds <- post / (1- post)
hist(odds)
```



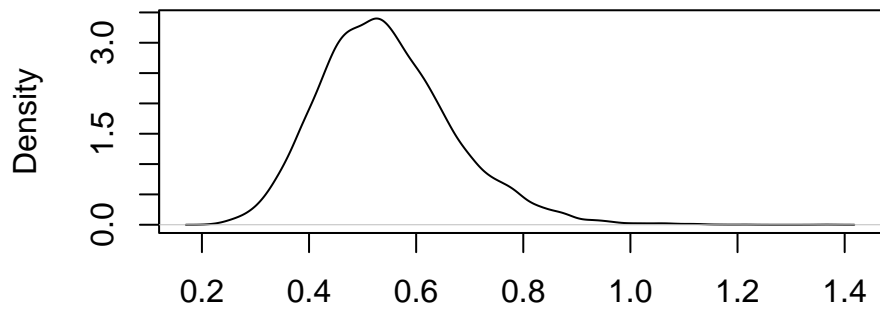
```
plot(density(post), main = "Density of posterior"); plot(density(odds), main = "Density of odds posterior")
```

Density of posterior



N = 10000 Bandwidth = 0.007262

Density of odds posterior



N = 10000 Bandwidth = 0.01711

The odds ratio transformation slightly skews the distribution to the right and the distribution is also wider with less density concentration around the mean.

2 Log-normal distribution and the Gini coefficient.

Assume that you have asked 8 randomly selected persons about their monthly income (in thousands SEK) and obtained the observations below. A common model for non-negative continuous variables is the log-normal distribution $\log N(\mu, \sigma^2)$ has density function

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

where $y > 0$, $-\infty < \mu < \infty$ and $\sigma^2 > 0$. Log-normal distribution is related to the normal distribution as:

$$\text{if } y \sim \log N(\mu, \sigma^2) \text{ then } \log y \sim N(\mu, \sigma^2)$$

Let y be log-normal distributed where $\mu = 3.6$ is assumed to be known but variance σ^2 is unknown with non-informative prior $p(\sigma^2) \propto 1/\sigma^2$. The posterior for σ^2 is the $Inv - \chi^2(n, \tau^2)$ where

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

```
income <- c(33,24,48,32,55,74,23,17) # observed incomes

# tau distribution
tau2 <- function(y,n,mu){
  sum((log(y)-mu)^2)/n
}
```

2.1 a

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

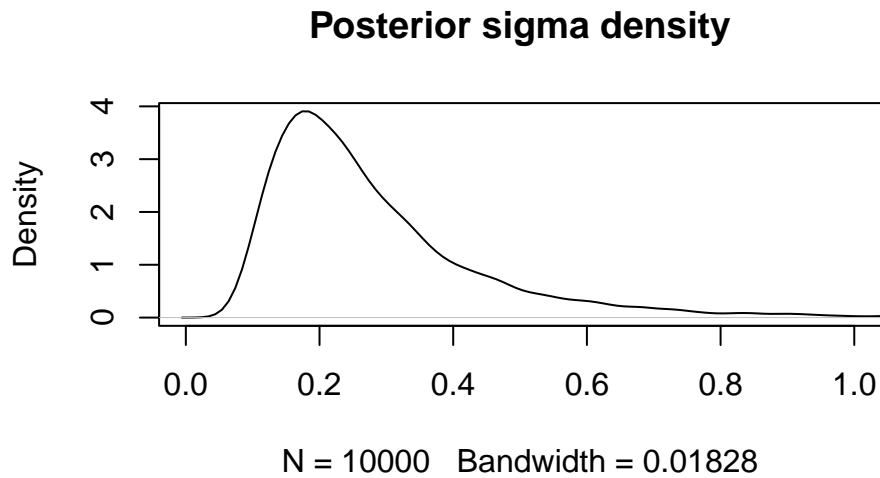
Posterior $\sigma^2 = \frac{n \cdot \tau^2}{X}$ where $X \sim \chi^2(n)$ due to τ^2 having n degrees of freedom.

```
n <- 8 # amount of observed incomes
mu <- 3.6

tau <- tau2(income,n,mu) # tau calculated using mu

post_sigma <- n*tau/rchisq(10000,n)

plot(density(post_sigma),main="Posterior sigma density",xlim=c(0,1))
```

Posterior σ^2 has a right skewed density with mode at slightly below 0.2.

2.2 b

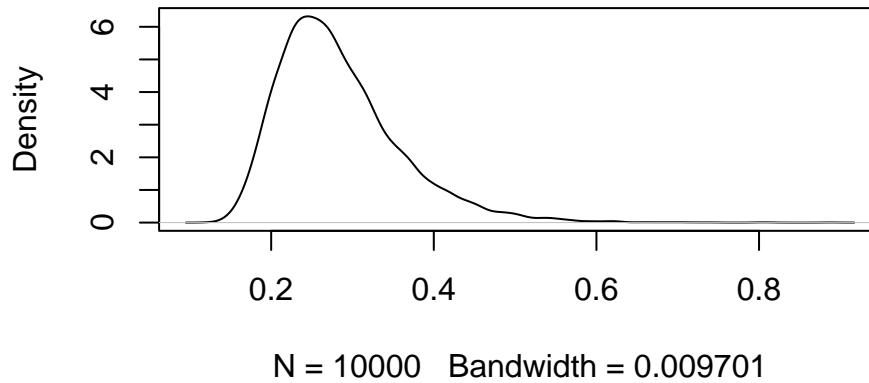
Most common measure of income inequality is Gini coefficient, $0 < G < 1$, where $G = 0$ is a completely equal income distribution and $G = 1$ completely unequal. It can be shown that,

$$G = 2\Phi(\sigma/\sqrt{2}) - 1 \text{ when income is distributed } \log N(\mu, \sigma^2)$$

$\Phi(z)$ is the cdf for the standard normal distribution with mean zero and unit variance. Use the posterior draws in a) to compute the posterior distribution of the Gini coefficient G for the current data.

```
gini <- 2*pnorm(q=sqrt(post_sigma)/sqrt(2), mean=0, sd=1) -1
plot(density(gini), main = "Gini coefficient posterior distribution")
```

Gini coefficient posterior distribution



The Gini distribution has a slight right skewed income with a mode at around 0.25. This indicates that income is distributed rather equally for the subjects but a low amount making substantially more than the others.

2.3 c

Use the posterior draws from b) to compute a 95% equal tail credible interval for G . A 95% equal tail credible interval (a, b) cuts off 2.5% of the posterior probability mass to the left of a , and to the right of b .

```
eti <- quantile(gini, c(0.025, 0.975))  
  
knitr::kable(et, col.names = 'Interval')
```

	Interval
2.5%	0.1762792
97.5%	0.4656198

This interval shows the probability of the Gini coefficient G being outside of this interval as 2.5% per side, or 5% in total. The equal tail show the same density outside the interval on both sides.

2.4 d

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

```
set.seed(12345)  
gini_dens <- density(gini) # kernel density estimate of G posterior
```

```

sort_dens <- sort(gini_dens$y,decreasing=TRUE) # order the estimated density values

cdf <- cumsum(sort_dens)/sum(sort_dens) # cdf of sorted G posterior

cdf <- cdf[cdf<0.95] # 95% of the cdf

# subset the coordinates of points with highest pdf for 95% of cdf
index <- gini_dens$x[order(gini_dens$y,decreasing=TRUE)][1:length(cdf)]

vals <- data.frame("min" = min(index), "max" = max(index))
knitr::kable(vals, col.names = c("min", "max"))

```

min	max
0.1610018	0.4360971

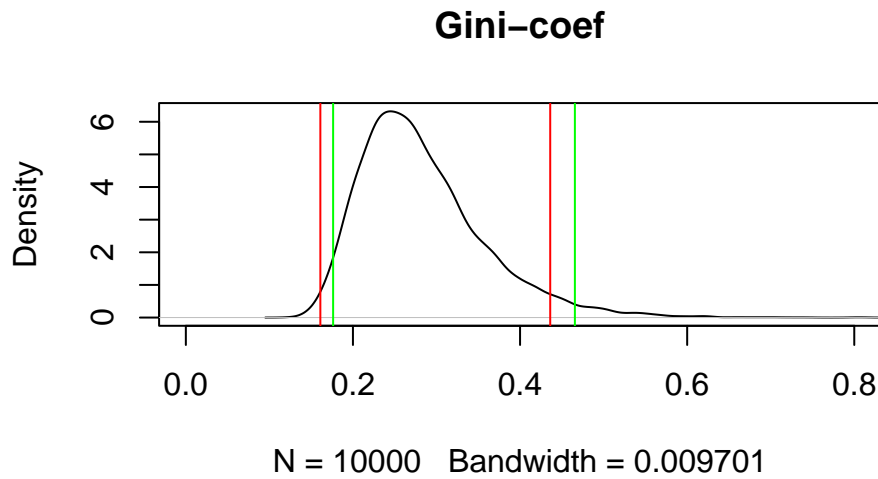
The above values show the min and max values for the HPDI, i.e. the interval for θ -values with the highest pdf.

```
set.seed(12345)
df2 <- data.frame(rbind(vals,eti))
rownames(df2) <- c('HPDI', 'ETCI')
knitr::kable(df2,caption = 'Comparison between the intervals')
```

Table 4: Comparison between the intervals

	min	max
HPDI	0.1610018	0.4360971
ETCI	0.1762792	0.4656198

```
plot(gini_dens, main='Gini-coef', xlim = c(0,0.8))
abline(v=eti[1],col='green')
abline(v=eti[2], col='green')
abline(v=vals[1],col='red')
abline(v=vals[2],col='red')
```



The table and figure above indicate the differences between the intervals. In the plot, the green lines show the 95% EQTI and the red lines the HPDI. The HPDI is slightly narrower than the EQTI, and is also located more to the left. In the figure this is visualized clearly, and the resulting gain of density is also shown. Whereas the EQTI disregards a lot of density on the lower side of the distribution, the HPDI picks this up for a smaller loss of density at the higher side.

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

This exercise is concerned with directional data and the point is to show that the posterior distribution for somewhat weird models can be obtained by plotting it over a grid of values. The data points are observed wind direction at a given location on ten different days and are recorded in degrees as 20, 314, 285, 40, 308, 314, 299, 296, 303, 326.

North is located at zero degrees. According to Wikipedias description of probability distributions for circular data we convert the data into radian $-\pi \leq y \leq \pi$. The observations in radians are $-2.79, 2.33, 1.83, -2.44, 2.23, 2.33, 2.07, 2.02, 2.14, 2.54$

Assume that these data points conditional on (μ, κ) are independent observations from the following

$$\text{von Mises distribution : } p(y | \mu, \kappa) = \frac{\exp[\kappa \cdot \cos(y - \mu)]}{2\pi I_0(\kappa)}, -\pi \leq y \leq \pi$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind of order zero (`?bessell` in R). The parameter μ ($-\pi < \mu < \pi$) is the mean direction and $\kappa > 0$ is called the concentration parameter. Large κ gives a small variance around μ , and vice versa. Assume that μ is known to be 2.4. Let $\kappa \sim \text{Exponential}(\gamma = 0.5)$ a priori, where γ is the rate parameter of the exponential distribution (so that the mean is $1/\gamma$).

3.1 a)

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

$$\begin{aligned} p(y | \mu, \kappa) &= \prod_{i=1}^n \frac{\exp[\kappa \cdot \cos(y - \mu)]}{2\pi I_0(\kappa)} \\ &= \frac{1}{(2\pi I_0(\kappa))^n} \cdot \exp(\kappa \cdot \sum_{i=1}^n \cos(y - \mu)) \\ &= \frac{1}{(2\pi)^n} \cdot \frac{1}{(I_0(\kappa))^n} \cdot \exp(\kappa \cdot \sum_{i=1}^n \cos(y - \mu)) \end{aligned}$$

$$\text{This is a constant and can therefore be ignored} = \frac{1}{(2\pi)^n}$$

$$\text{Likelihood} = \frac{\exp(\kappa \cdot \sum_{i=1}^n \cos(y - \mu))}{(I_0(\kappa))^n}.$$

$$\text{A priori, } \kappa \sim \text{Exp}(\gamma = 0.5) \implies \text{Prior} = p(\kappa) \propto \lambda e^{-\lambda\kappa} = 0.5e^{-0.5\kappa}$$

$$\text{Posterior} = p(\kappa | y, \mu) \propto p(y | \mu, \kappa) \cdot p(\kappa) = \frac{\exp(\kappa \cdot \sum_{i=1}^n \cos(y - \mu))}{(I_0(\kappa))^n} \cdot \lambda e^{-\lambda\kappa}$$

$\lambda = 0.5$ is a constant and can be ignored.

$$\text{Posterior} = p(\kappa \mid y, \mu) \propto p(y \mid \mu, \kappa) \cdot p(\kappa) = \frac{\exp(\kappa \cdot \sum^n \cos(y - \mu)) \cdot \exp(-\lambda\kappa)}{(I_0(\kappa))^n}$$

Following rules for multiplying exponentials

$$p(\kappa \mid y, \mu) \propto p(y \mid \mu, \kappa) \cdot p(\kappa) = \frac{\exp(\kappa \cdot \sum^n \cos(y - \mu) - \lambda\kappa)}{(I_0(\kappa))^n}$$

Below we calculate the Posterior $\propto f(\kappa)$.

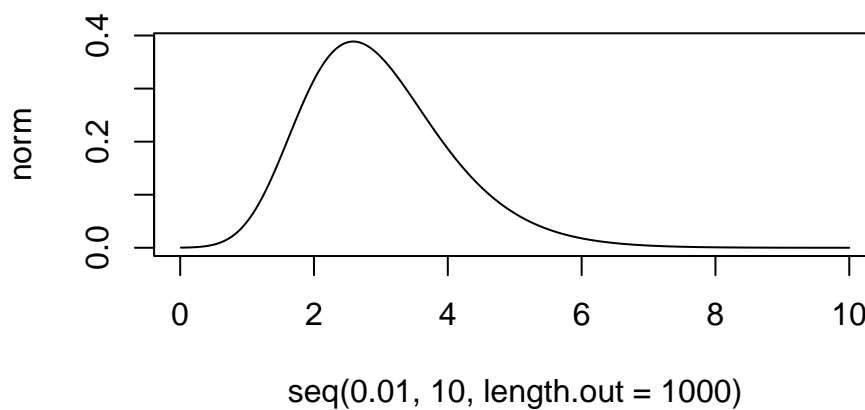
```
y <- c(-2.79, 2.33, 1.83, -2.44, 2.23, 2.33, 2.07, 2.02, 2.14, 2.54)
lambda <- 0.5
n <- 10
mu <- 2.4

posterior <- function(k){
  exp(k * (sum(cos(y-mu))) - k*lambda) / (besselI(k, 0)^n)
}
```

- Plotting the posterior distribution over a fine grid

```
# applying a grid of k values for the posterior function
posterior_3 <- sapply(seq(0.01,10,by=0.01),posterior)

norm <- posterior_3/(sum(posterior_3)*0.01)
plot(seq(0.01,10,length.out=1000),norm, type='l')
```



The posterior is right skewed and has a high density between around 2-4.

3.2 b

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

- The fastest way to do this is by finding index for where the posterior density vector is maximized

```
which.max(norm); norm[which.max(norm)]
```

```
## [1] 259
```

```
## [1] 0.3886893
```

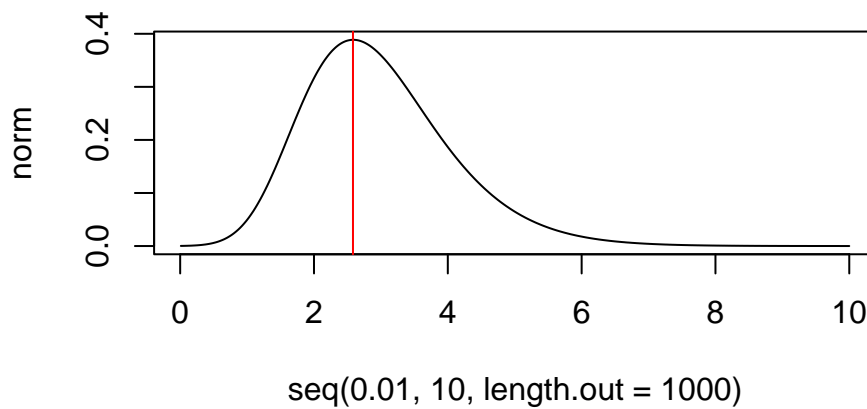
At element 259 the posterior density has value 0.003886893, which we use to find the maximum (i.e. mode) value of the $[0,10]$ sequence.

```
seq(0,10,length.out=1000)[which.max(norm)]
```

```
## [1] 2.582583
```

The approximate posterior mode of κ is 2.582583. If the sequenced numbers would increase the approximate value would be closer to true convergence.

```
plot(seq(0.01,10,length.out=1000),norm, type='l')  
abline(v=(seq(0,10,length.out=1000)[which.max(norm)]), col = "red")
```



This graph confirms that the value found of 2.162162 is the approximate posterior mode of κ .

4 Appendix

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning=FALSE, fig.width = 5, fig.height = 3, fig.al
set.seed(12345)
library(coda)
library(invgamma)

s <- 22 # successes
f <- 70 - 22 # failures
n <- 70 # trials

# Beta priors
A0 <- 8
B0 <- 8

nDraws <- 10000 # nr of draws

true_mean <- (A0 + s) / (A0 + B0 + n) # calculating true mean and sd for posterior dist
true_sd <- sqrt(((A0+s)*(B0+f))/((A0+B0+n)^2 * (A0+ B0+n+1)))

posterior_sample <- rbeta(nDraws,A0+s,B0+f) # draw values from posterior

# Calculate sample cumulative means and standard deviations to show convergence
sample_means <- cumsum(posterior_sample)/(1:nDraws)
sample_sds <- sqrt(cumsum((posterior_sample - sample_means)^2) / (1:nDraws))

plot(sample_means,type='line',xlab='Ndraws',ylab='Sample mean', main='Graph over sampled means')
abline(h=true_mean,col='red')
plot(sample_sds,type='line',xlab='Ndraws',ylab='Sample sd', main='Graph over sampled sds')
abline(h=true_sd,col='red')
set.seed(12345)

# draws from posterior
post <- rbeta(nDraws, A0+s, B0+f)

# mean of samples over 0.3
prob <- mean(post>0.3)

prob_exact <- 1 - pbeta(0.3,A0+s, B0+f)

df <- data.frame('Posterior prob' =prob, 'Exact value from beta post' = prob_exact)
colnames(df) <- c('Posterior prob','Exact value from beta post')

knitr::kable(df)

odds <- post / (1- post)
```

```

hist(odds)

plot(density(post), main = "Density of posterior"); plot(density(odds), main = "Density of odds posteri

income <- c(33,24,48,32,55,74,23,17) # observed incomes

# tau distribution
tau2 <- function(y,n,mu){

  sum((log(y)-mu)^2)/n
}

n <- 8 # amount of observed incomes
mu <- 3.6

tau <- tau2(income,n,mu) # tau calculated using mu

post_sigma <- n*tau/rchisq(10000,n)

plot(density(post_sigma),main="Posterior sigma density",xlim=c(0,1))

gini <- 2*pnorm(q=sqrt(post_sigma)/sqrt(2), mean=0, sd=1) -1

plot(density(gini), main = "Gini coefficient posterior distribution")
eti <- quantile(gini,c(0.025,0.975))

knitr::kable(eti,col.names = 'Interval')
set.seed(12345)
gini_dens <- density(gini) # kernel density estimate of G posterior

sort_dens <- sort(gini_dens$y,decreasing=TRUE) # order the estimated density values

cdf <- cumsum(sort_dens)/sum(sort_dens) # cdf of sorted G posterior

cdf <- cdf[cdf<0.95] # 95% of the cdf

# subset the coordinates of points with highest pdf for 95% of cdf
index <- gini_dens$x[order(gini_dens$y,decreasing=TRUE)][1:length(cdf)]

vals <- data.frame("min" = min(index), "max" = max(index))
knitr::kable(vals, col.names = c("min", "max"))
set.seed(12345)
df2 <- data.frame(rbind(vals,eti))
rownames(df2) <- c('HPDI', 'ETCI')
knitr::kable(df2,caption = 'Comparison between the intervals')

plot(gini_dens, main='Gini-coef', xlim = c(0,0.8))
abline(v=eti[1],col='green')

```

```

abline(v=eti[2], col='green')
abline(v=vals[1],col='red')
abline(v=vals[2],col='red')
y <- c(-2.79, 2.33, 1.83, -2.44, 2.23, 2.33, 2.07, 2.02, 2.14, 2.54)
lambda <- 0.5
n <- 10
mu <- 2.4

posterior <- function(k){

  exp(k * (sum(cos(y-mu)))-k*lambda)/ (besselI(k, 0)^n)

}
# applying a grid of k values for the posterior function
posterior_3 <- sapply(seq(0.01,10,by=0.01),posterior)

norm <- posterior_3/(sum(posterior_3)*0.01)
plot(seq(0.01,10,length.out=1000),norm, type='l')
which.max(norm);norm[which.max(norm)]
seq(0,10,length.out=1000)[which.max(norm)]
plot(seq(0.01,10,length.out=1000),norm, type='l')
abline(v=(seq(0,10,length.out=1000)[which.max(norm)]), col = "red")

```