

732A91 - Bayesian learning

Lab 1 - “The bayesics”

Alexander Karlsson (aleka769)

Thijs Quast (thiqu264)

30-05-2019

Contents

Assignment 1	3
Bernoulli ... again. Let y_1, \dots, y_n distributed as $\text{Bern}(\theta)$, and assume that you have obtained a sample with $s = 14$ successes in $n = 20$ trials. Assume $\text{Beta}(\alpha_0, \beta_0)$ prior for θ and let $\alpha_0 = \beta_0 = 2$	3
a Draw random numbers from the posterior θy distributed as $\text{Beta}(\alpha_0 + s, \beta_0 + f)$, $y = (y_1, \dots, y_n)$, and verify graphically that the posterior mean and standard deviation converges to the true values as the number of random draws grows large.	3
b Use simulation ($n\text{Draws} = 10000$) to compute the posterior probability $\Pr(\theta < 0.4 \text{ given } y)$ and compare with the exact value [Hint: <code>pbeta()</code>].	4
c Compute the posterior distribution of the log-odds $\phi = \log(\theta/(1-\theta))$ by simulation ($n\text{Draws} = 10000$). [Hint: <code>hist()</code> and <code>density()</code> might come in handy]	5
Assignment 2	5
Log-normal distribution and the Gini coefficient. Assume that you have asked 10 randomly selected persons about their monthly income (in thousands Swedish Krona) and obtained the following ten observations: 14, 25, 45, 25, 30, 33, 19, 50, 34 and 67. A common model for non-negative continuous variables is the log-normal distribution. The log-normal distribution $\text{logNormal}(\mu, \sigma^2)$ has density function	5
for $y > 0$, $\mu > 0$ and $\sigma^2 > 0$. The log-normal distribution is related to the normal distribution as follows: if y distributed as $\text{log Normal}(\mu, \sigma^2)$ then $\log y$ distributed as $N(\mu, \sigma^2)$. Let the sample be iid; y_1, \dots, y_n given μ, σ^2 distributed as $\text{logN}(\mu, \sigma^2)$, where $\mu = 3.5$ is assumed to be known but σ^2 is unknown with noninformative prior $p(\sigma^2)$ proportional to $1/\sigma^2$. The posterior for σ^2 is the Inverse-scaled-chi squared(n, τ^2) distribution, where:	6
a Simulate 10000 draws from the posterior of σ^2 (assuming $\mu = 3.5$) and compare with the theoretical Inverse-scaled-chi-squared(n, τ^2) posterior distribution.	6
b Use the posterior draws in a) to compute the posterior distribution of the Gini coefficient G for the current data set.	7
c Use the posterior draws from b) to compute a 95% equal tail credible interval for G . An 95% equal tail interval (a, b) cuts off 2.5% percent of the posterior probability mass to the left of a , and 97.5% to the right of b . Also, do a kernel density estimate of the posterior of G using the density function in R with default settings, and use that kernel density estimate to compute a 95% Highest Posterior Density interval for G . Compare the two intervals.	8
Assignment 3	10

Bayesian inference for the concentration parameter in the von Mises distribution. This exercise is concerned with directional data. The point is to show you 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 directions at a given location on ten different days. The data are recorded in degrees: (40, 303, 326, 285, 296, 314, 20, 308, 299, 296), where North is located at zero degrees (see Figure 1 on the next page, where the angles are measured clockwise). To fit with Wikipedias description of probability distributions for circular data we convert the data into radians $-\pi \leq y \leq \pi$. The 10 observations in radians are (-2.44, 2.14, 2.54, 1.83, 2.02, 2.33, -2.79, 2.23, 2.07, 2.02). Assume that these data points are independent observations following the von Mises distribution	10
where $I_0(k)$ is the modified Bessel function of the first kind of order zero [see ?besselI in R]. The parameter μ ($-\pi \leq \mu \leq \pi$) is the mean direction and $k > 0$ is called the concentration parameter. Large k gives a small variance around μ , and vice versa. Assume that μ is known to be 2.39. Let k distributed $\text{Exponential}(\lambda = 1)$ a priori, where λ is the rate parameter of the exponential distribution (so that the mean is $1/\lambda$).	10
a Plot the posterior distribution of k for the wind direction data over a fine grid of k values. . . .	10
b Find the (approximate) posterior mode of k from the information in a).	11

Assignment 1

Bernoulli ... again. Let y_1, \dots, y_n distributed as $\text{Bern}(\theta)$, and assume that you have obtained a sample with $s = 14$ successes in $n = 20$ trials. Assume $\text{Beta}(\alpha_0, \beta_0)$ prior for θ and let $\alpha_0 = \beta_0 = 2$.

a Draw random numbers from the posterior $\theta|y$ distributed as $\text{Beta}(\alpha_0 + s, \beta_0 + f)$, $y = (y_1, \dots, y_n)$, and verify graphically that the posterior mean and standard deviation converges to the true values as the number of random draws grows large.

Model:

$$y_1, \dots, y_n | \theta \sim \text{Bernoulli}(\theta)$$

Number of successes : 14, total number of observations = 20

Likelihood:

$$p(y_1, \dots, y_n | \theta) = \prod_{i=1}^n p(y_i | \theta) = \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1 - y_i} = \theta^{\sum_{i=1}^n y_i} (1 - \theta)^{n - \sum_{i=1}^n y_i} = \theta^s (1 - \theta)^f,$$

where $\sum_{i=1}^n y_i$ is the number of successive trials (denoted as s) and $1 - \sum_{i=1}^n y_i$ is the number of non-successive trials (denoted as f)

Prior is a gamma distribution:

$$\text{Prior}(\theta) \sim \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

Where $\alpha = 2$, $\beta = 2$

Posterior distributed as $\text{Beta}(16, 8)$

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

$$\text{Posterior}(\theta) \propto \theta^{\alpha+s-1} e^{-(\beta+f)\theta}$$

Theoretical mean:

$$E[\theta] = \frac{\alpha}{\alpha + \beta}$$

Theoretical standard deviation:

$$SD[\theta] = \sqrt{\frac{\alpha\beta}{(\alpha\beta)^2(\alpha + \beta + 1)}}$$

```
alpha_prior <- 2
beta_prior <- 2

alpha_posterior <- 2+14
beta_posterior <- 2+6
```

```

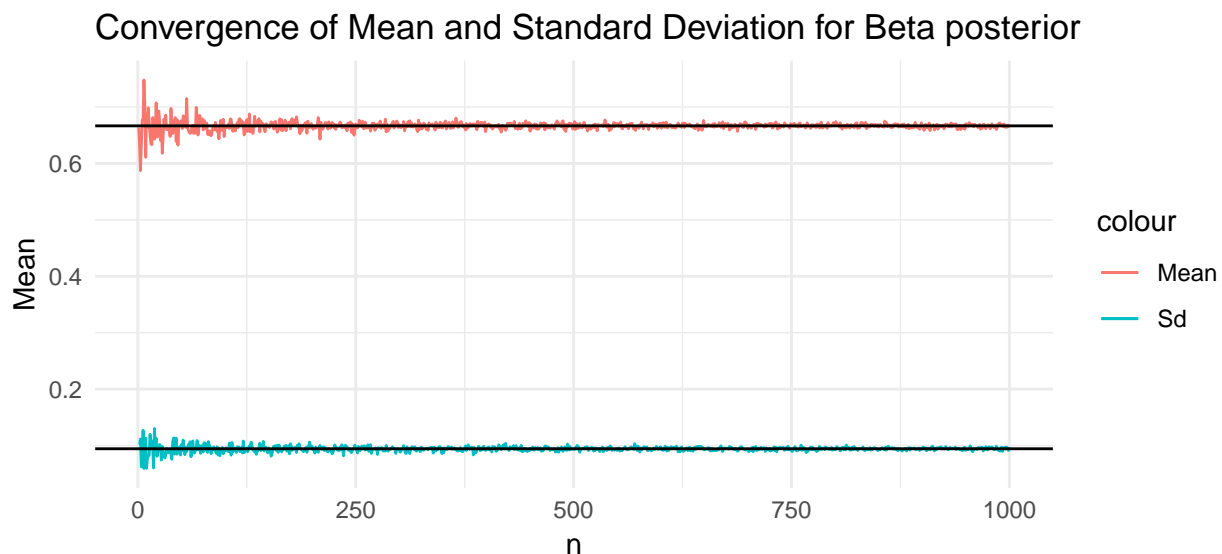
theoretical_mean <- alpha_posterior/(alpha_posterior + beta_posterior)
theoretical_sd <- sqrt(alpha_posterior*beta_posterior/((alpha_posterior+beta_posterior)^2 *
                                                         (alpha_posterior + beta_posterior + 1)))

df <- data.frame("Mean" = 0, "Sd" = 0, "n" = 0)

for (i in 1:1000){
  draw <- rbeta(i, alpha_posterior, beta_posterior)
  df[i,] <- c(mean(draw), sqrt(var(draw)), i)
}

library(ggplot2)
plot <- ggplot(data = df, aes(x = n, y = Mean, col= "Mean")) + geom_line()
plot <- plot + geom_abline(intercept = theoretical_mean, slope = 0)
plot <- plot + geom_line(aes(x = n, y = Sd, col="Sd"))
plot <- plot + geom_abline(intercept = theoretical_sd, slope = 0)
plot <- plot + ggtitle("Convergence of Mean and Standard Deviation for Beta posterior")
plot

```



b Use simulation ($n\text{Draws} = 10000$) to compute the posterior probability $\Pr(\theta < 0.4 \text{ given } y)$ and compare with the exact value [Hint: `pbeta()`].

```

sample_10000 <- rbeta(10000, alpha_posterior, beta_posterior)
sample_0.4 <- sample_10000[sample_10000 < 0.4]

simulated_probability <- length(sample_0.4)/length(sample_10000)

theoretical_probability <- pbeta(0.4, alpha_posterior, beta_posterior)

library(knitr)
kable(data.frame("Simulated Probability" = simulated_probability,
                  "Theoretical Probability" = theoretical_probability))

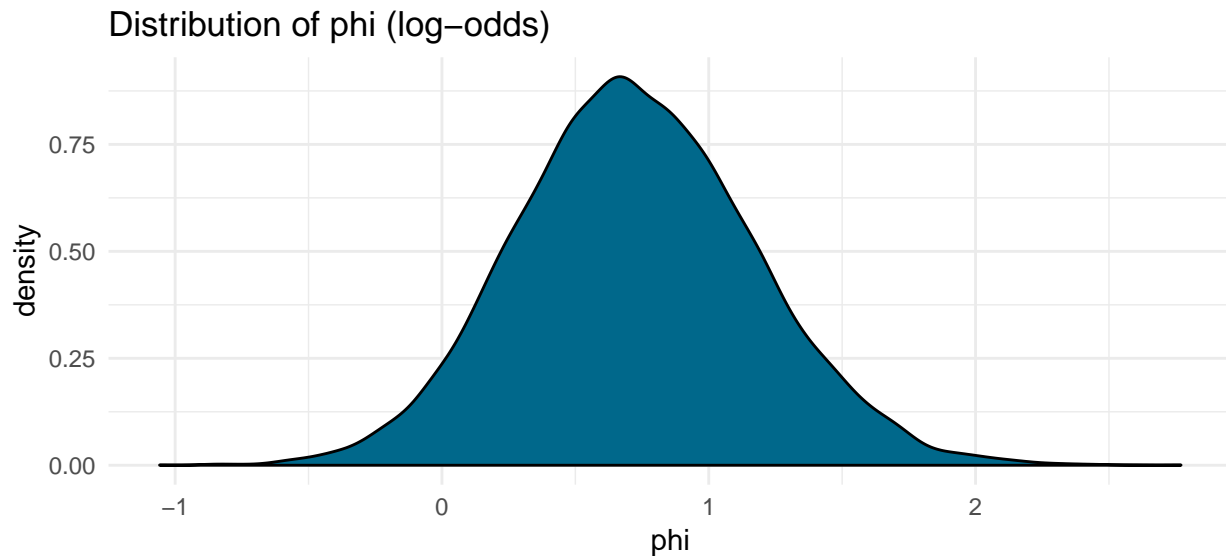
```

Simulated.Probability	Theoretical.Probability
0.0048	0.0039727

c Compute the posterior distribution of the log-odds $\phi = \log(\theta/(1-\theta))$ by simulation (nDraws = 10000). [Hint: hist() and density() might come in handy]

$$\phi = \log \frac{\theta}{1-\theta}$$

```
phi <- log(sample_10000/(1-sample_10000))
phi_df <- as.data.frame(phi)
plot_phi <- ggplot(phi_df, aes(x = phi, y = ..density..)) + geom_density(fill="deepskyblue4") +
  ggtitle("Distribution of phi (log-odds)")
plot_phi
```



Assignment 2

Log-normal distribution and the Gini coefficient. Assume that you have asked 10 randomly selected persons about their monthly income (in thousands Swedish Krona) and obtained the following ten observations: 14, 25, 45, 25, 30, 33, 19, 50, 34 and 67. A common model for non-negative continuous variables is the log-normal distribution. The log-normal distribution $\text{logNormal}(\mu, \sigma^2)$ has density function

$$y \sim \text{logN}(\mu, \sigma^2)$$

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

for $y > 0$, $\mu > 0$ and $\sigma^2 > 0$. The log-normal distribution is related to the normal distribution as follows: if y distributed as $\log \text{Normal}(\mu, \sigma^2)$ then $\log y$ distributed as $N(\mu, \sigma^2)$. Let the sample be iid; y_1, \dots, y_n given μ, σ^2 distributed as $\log N(\mu, \sigma^2)$, where $\mu = 3.5$ is assumed to be known but σ^2 is unknown with noninformative prior $p(\sigma^2)$ proportional to $1/\sigma^2$. The posterior for σ^2 is the Inverse-scaled-chi squared(n, τ^2) distribution, where:

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

a Simulate 10000 draws from the posterior of σ^2 (assuming $\mu = 3.5$) and compare with the theoretical Inverse-scaled-chi-squared(n, τ^2) posterior distribution.

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

Because in R we can sample X from a chi-squared distribution, we scale these values in such a way to obtain the desired inverse scaled chi-squared distribution.

$$X \sim \chi^2(n) \Rightarrow \frac{\tau^2 v}{X} \sim \text{scaled inv } \chi^2$$

```
invScaledChisqPDF = function(tau2, n, x){
  exp1 = (((tau2 * n) / 2)^(n/2)) / gamma(n/2)
  exp2 = exp((-n*tau2) / (2*x)) / x^(1 + n/2)
  return(exp1 * exp2)
}

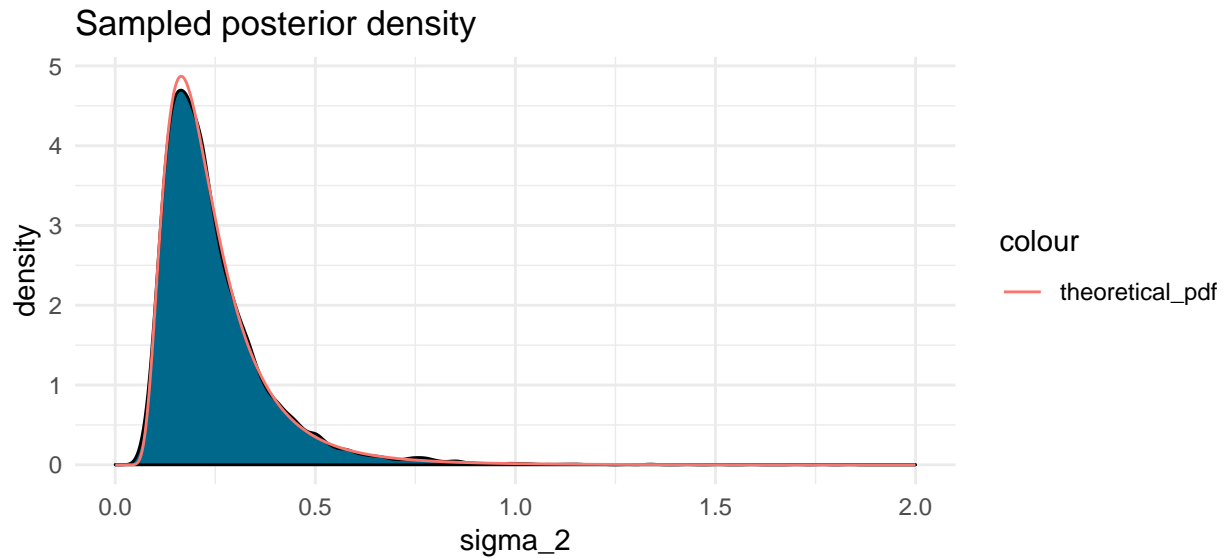
y = c(14, 25, 45, 25, 30, 33, 19, 50, 34, 67)
n <- 10
v <- n
mu <- 3.5

tau_squared <- sum((log(y) - mu)^2) / n

chi_squared <- rchisq(10000, df = v)
scaled_chi_squared <- (tau_squared * v) / chi_squared
theoretical_data <- seq(0, 1.9998, by = 2/10000)
theoretical_pdf <- invScaledChisqPDF(tau2 = tau_squared, n = n, x = theoretical_data)

df_chi <- data.frame("sigma_2" = scaled_chi_squared,
                    "x" = theoretical_data, "theoretical_pdf" = theoretical_pdf)

plot_chi <- ggplot(df_chi, aes(x = sigma_2, y = ..density..)) + geom_density(fill = "deepskyblue4")
plot_chi <- plot_chi + geom_line(aes(x = theoretical_data, y = theoretical_pdf, col="theoretical_pdf"))
plot_chi <- plot_chi + ggtitle("Sampled posterior density")
plot_chi
```



```
theoretical_mean_2 <- v*tau_squared/(v-2)
actual_mean <- mean(scaled_chi_squared)
theoretical_variance <- ((2*v^2)*(tau_squared^2))/((v-2)^2*(v-4))
sampled_variance <- var(scaled_chi_squared)

kable(data.frame("Theoretical Mean" = theoretical_mean_2, "Sampled Mean" = actual_mean,
                  "Theoretical variance" = theoretical_variance,
                  "Sampled variance" = sampled_variance))
```

Theoretical.Mean	Sampled.Mean	Theoretical.variance	Sampled.variance
0.2473497	0.2469987	0.020394	0.0194036

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

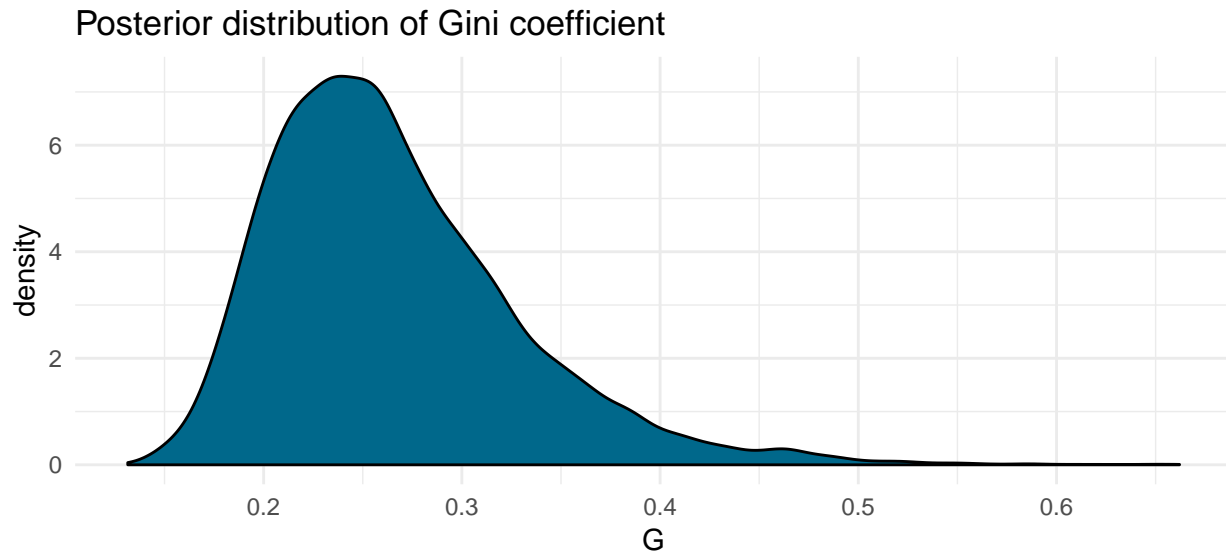
To calculate the Gini-coefficient posterior distribution, we use the cumulative distribution function ($\Phi(\sigma)$) for the obtained sampled standard deviation values in 2a) according to:

$$G = 2 \cdot \Phi(\sigma/\sqrt{2}) - 1,$$

where $\Phi(\dots)$ is the standard normal cumulative distribution function (i.e. $\mu = 0$ and $\sigma = 1$).

```
G <- 2*pnorm(q = sqrt(scaled_chi_squared)/sqrt(2), mean = 0, sd = 1)-1
df_G <- data.frame("G" = G)
```

```
plot_G <- ggplot(df_G, aes(G)) + geom_density(fill="deepskyblue4")
plot_G <- plot_G + ggtitle("Posterior distribution of Gini coefficient")
plot_G
```



c Use the posterior draws from b) to compute a 95% equal tail credible interval for G . An 95% equal tail interval (a, b) cuts off 2.5% percent of the posterior probability mass to the left of a , and 97.5% to the right of b . Also, do a kernel density estimate of the posterior of G using the density function in R with default settings, and use that kernel density estimate to compute a 95% Highest Posterior Density interval for G . Compare the two intervals.

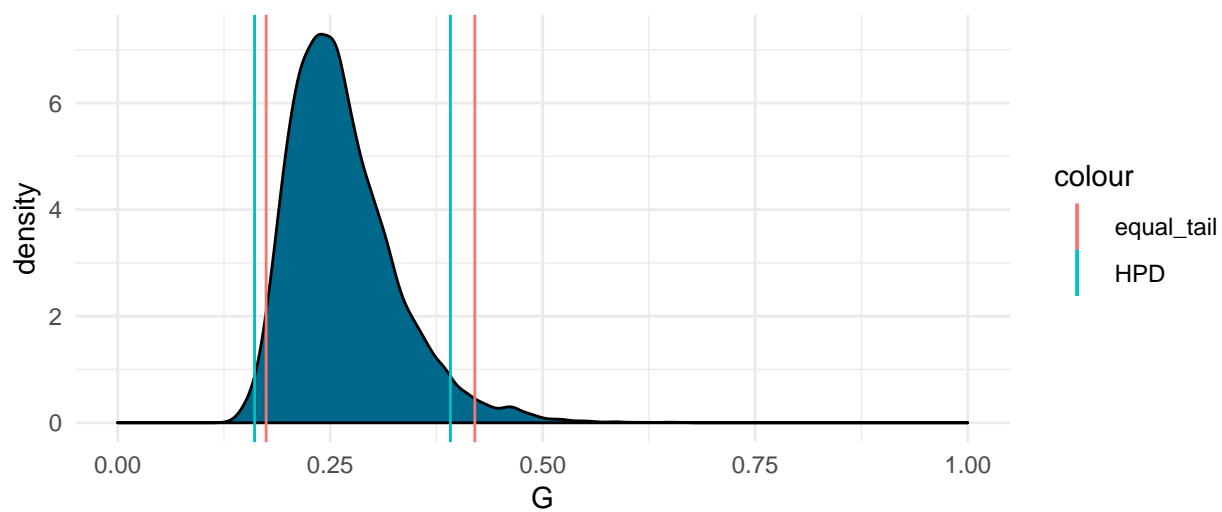
```
quantiles <- quantile(G, probs = c(0.025, 0.975))
min_quantile <- quantiles[1]
max_quantile <- quantiles[2]
equal_tail <- c(quantiles)

density_approximation <- approxfun(density(G))
densities <- density_approximation(G)
densities_df <- data.frame("x" = G, "density" = densities)
densities_df <- densities_df[order(densities_df$density, decreasing = TRUE),]
percentage <- 0.95
n_rows <- round(0.95 * nrow(densities_df))
densities_df <- densities_df[1:n_rows,]
x_min <- min(densities_df$x)
x_max <- max(densities_df$x)
HPD <- c(x_min, x_max)

plot_G <- ggplot(df_G, aes(G)) + geom_density(fill="deepskyblue4") + xlim(0, 1)
plot_G <- plot_G + geom_vline(aes(xintercept = equal_tail[1], col = "equal_tail"))
plot_G <- plot_G + geom_vline(aes(xintercept = equal_tail[2], col = "equal_tail"))

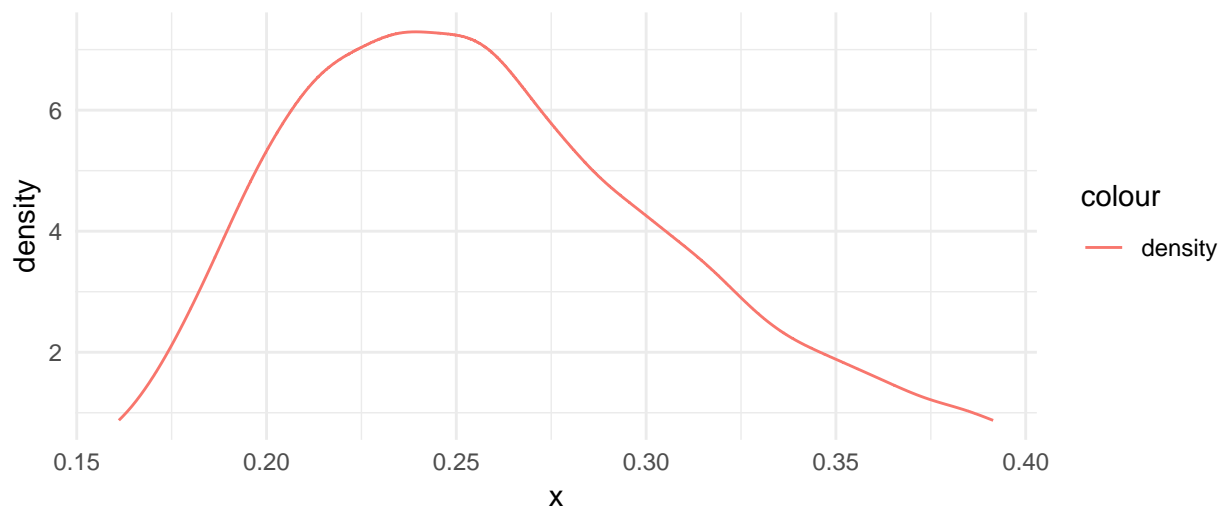
plot_G <- plot_G + geom_vline(aes(xintercept = HPD[1], color = "HPD"))
plot_G <- plot_G + geom_vline(aes(xintercept = HPD[2], col = "HPD"))
plot_G <- plot_G + ggtitle("Posterior distribution of G with intervals")
plot_G
```


Posterior distribution of G with intervals



```
plot_HPF <- ggplot(densities_df, aes(x = x, y = density, color = "density")) + geom_line()
plot_HPF <- plot_HPF + ggtitle("Highest Posterior Density")
plot_HPF
```

Highest Posterior Density



Assignment 3

Bayesian inference for the concentration parameter in the von Mises distribution. This exercise is concerned with directional data. The point is to show you 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 directions at a given location on ten different days. The data are recorded in degrees: (40, 303, 326, 285, 296, 314, 20, 308, 299, 296), where North is located at zero degrees (see Figure 1 on the next page, where the angles are measured clockwise). To fit with Wikipedias description of probability distributions for circular data we convert the data into radians $-\pi$ smallerthanorequalto y smallerthanorequalto π . The 10 observations in radians are (-2.44, 2.14, 2.54, 1.83, 2.02, 2.33, -2.79, 2.23, 2.07, 2.02). Assume that these data points are independent observations following the von Mises distribution

Model:

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

where $I_0(k)$ is the modified Bessel function of the first kind of order zero [see ?besselI in R]. The parameter μ ($-\pi$ smallerthanorequalto μ smallerthanorequalto π) is the mean direction and $k > 0$ is called the concentration parameter. Large k gives a small variance around μ , and vice versa. Assume that μ is known to be 2.39. Let k distributed Exponential($\lambda = 1$) a priori, where λ is the rate parameter of the exponential distribution (so that the mean is $1/\lambda$).

a Plot the posterior distribution of k for the wind direction data over a fine grid of k values.

Likelihood:

$$p(k|\mu, y) = \prod_{i=1}^n \left[\frac{\exp[k * \cos(y_i - \mu)]}{2\pi I_0(k)} \right] = \frac{\exp[k * \sum_{i=1}^n \cos(y_i - \mu)]}{[2\pi I_0(k)]^n}$$

$$Prior : p(k) \sim \lambda e^{-\lambda * k}$$

$$Posterior \propto Prior * Likelihood$$

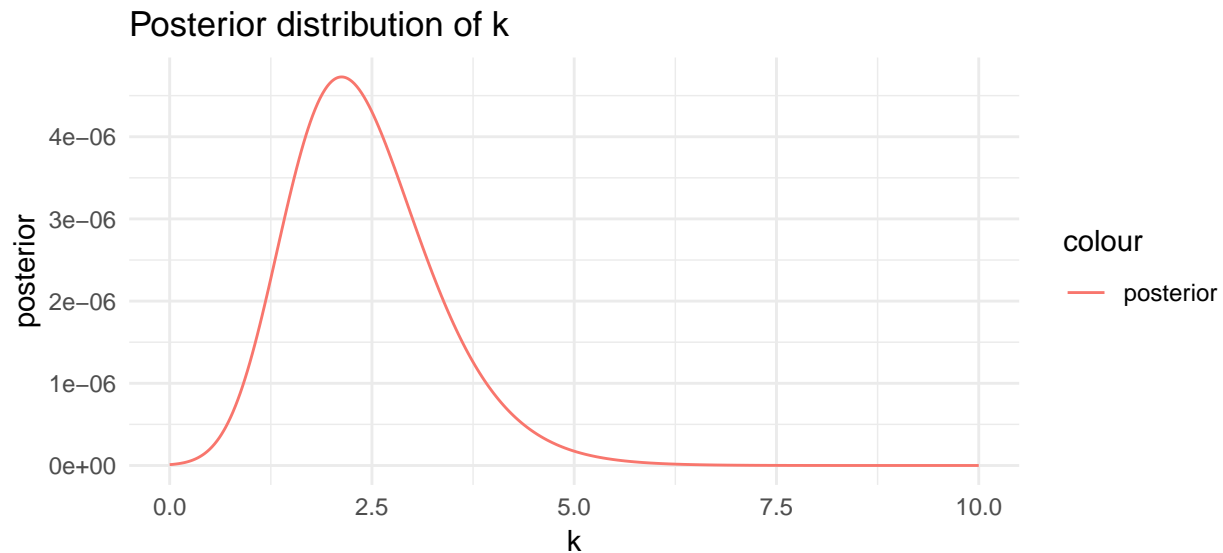
$$Posterior \propto \frac{\lambda * \exp[k \sum_{i=1}^n \cos(y_i - \mu) - \lambda * k]}{[2\pi I_0(k)]^n}$$

```
y_sampled <- c(-2.44, 2.14, 2.54, 1.83, 2.02, 2.33, -2.79, 2.23, 2.07, 2.02)
y_bar <- mean(y_sampled)
n <- length(y)
mu <- 2.39
k <- seq(0, 10, by = 0.01)
df_3 <- data.frame("k"=k, "likelihood" = 0)
```

```
for (i in 1:length(k)){
  df_3$likelihood[i] <- exp(k[i]*sum(cos(y_sampled - mu))) / ((2*pi*besselI(k[i], 0))^n)
}
```

```
df_3$prior <- dexp(k, rate = 1)
df_3$posterior <- df_3$prior * df_3$likelihood
```

```
plot_k <- ggplot(df_3, aes(x = k, y = posterior, col="posterior")) + geom_line()
plot_k <- plot_k + ggtitle("Posterior distribution of k")
plot_k
```



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

```
library(knitr)
m <- data.frame("Mode" = df_3$k[which.max(df_3$posterior)])
kable(m)
```

Mode
2.12