Task 5

5.1 Mixtures of beta priors

[25]

Estimate the probability θ of teen recidivism based on a study in which there were n = 43 individuals released from incarceration and y = 15 re-offenders within 36 months.

a) Using a beta$(2, 8)$ prior for θ, plot p(θ), p($y|\theta$) and p($\theta|y$) as functions of θ. Find the posterior mean, mode, and standard deviation of θ. Find a 95% quantile-based confidence interval.

b) Repeat a), but using a beta$(8, 2)$ prior for θ.

c) Consider the following prior distribution for θ:

$$p(\theta) = \frac{1}{4}\frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}[3\theta(1-\theta)^7 + \theta^7(1-\theta)]$$

which is a $75 - 25\%$ mixture of a beta$(2, 8)$ and a beta$(8, 2)$ prior distribution. Plot this prior distribution and compare it to the priors in a) and b). Describe what sort of prior opinion this may represent.

d) For the prior in **c)**:

i. Write out mathematically p(θ) $\times$ p($y|\theta$) and simplify as much as possible.

ii. The posterior distribution is a mixture of two distributions you know. Identify these distributions.

iii. On a computer, calculate and plot p(θ) $\times$ p($y|\theta$) for a variety of θ values. Also find (approximately) the posterior mode, and discuss its relation to the modes in **a)** and **b)**.

e) Find a general formula for the weights of the mixture distribution in **d)**ii, and provide an interpretation for their values.

[25] [DHo09]pg. 228 Exercise 3.4

Solution):

a) Based on the description of our problem, I deem the binomial model as a good fit for our teen recidivism study as a teen can either go back to jail or not and the binomial model provides the probability of getting a certain number of "successes" in a certain number of trials given a certain probability of "success"[26]. I note that this study will be focused on the 36-month time frame for our updated beliefs cause the data indicate they have been sampled from that 36-month period. The use of our theoretical knowledge is of key importance as the need for conjugate priors is unparalleled. Using the equation 2.2 we know that $\theta|Y \sim \text{beta}(\alpha = 2 + y = 15, \beta = 8 + n = 43 - y = 15) = \text{beta}(17, 36)$.

To answer **a)** the plot of p($y|\theta$) is (5.36) and I'll use the Monte Carlo sampling and present its usefulness in plotting the densities of p(θ) (5.35) in R:

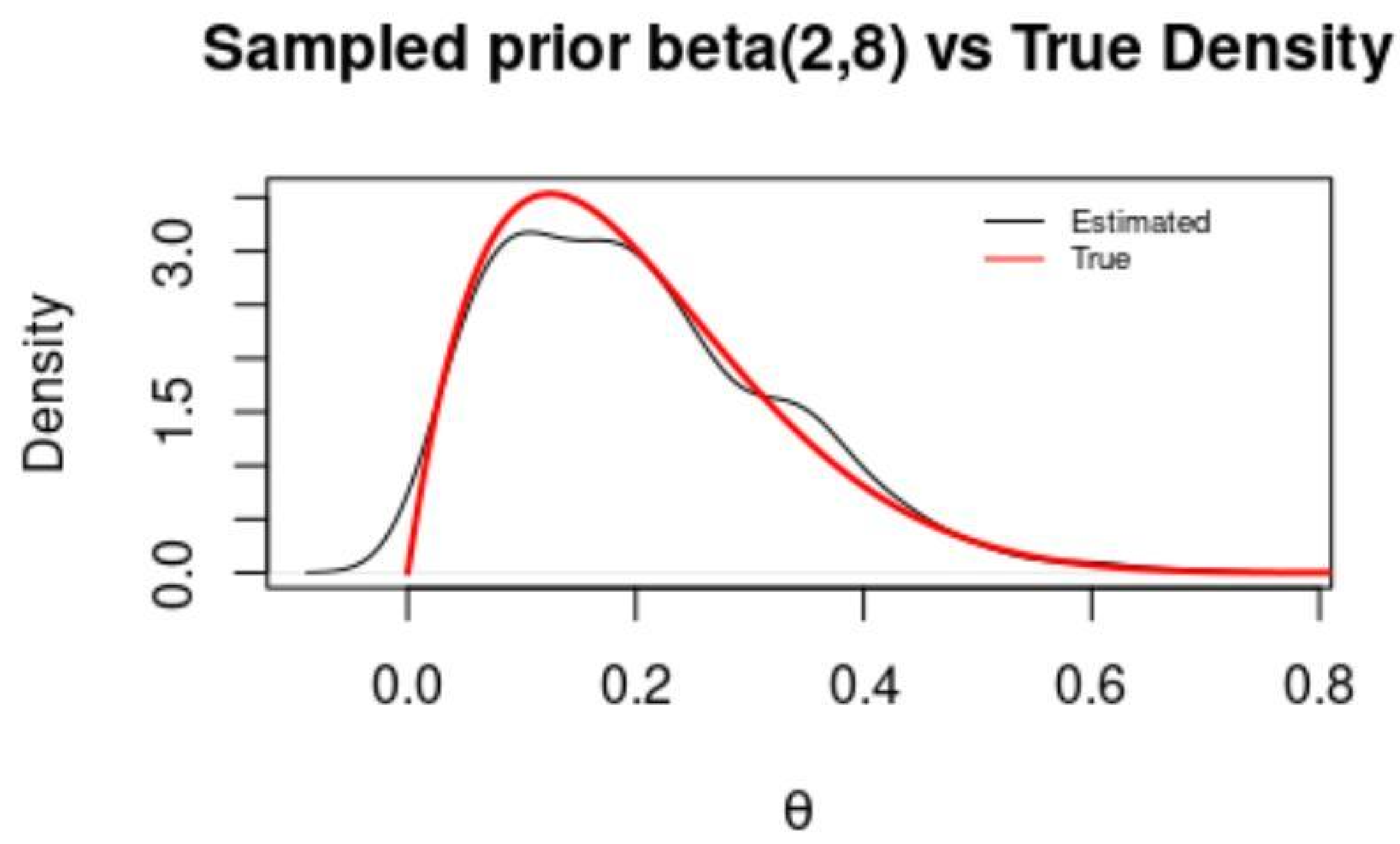


Figure 5.35: Prior Density plot.

```
# Sampling
prior_sample <- rbeta(500, 2, 8)

# Estimating the density
prior_density_estimation <- density(prior_sample)

# Plotting the estimated density
plot(prior_density_estimation,
     main="Sampled prior beta(2,8) vs True Density",
```

[26] In this case we should have avoided labeling this scenario as a success

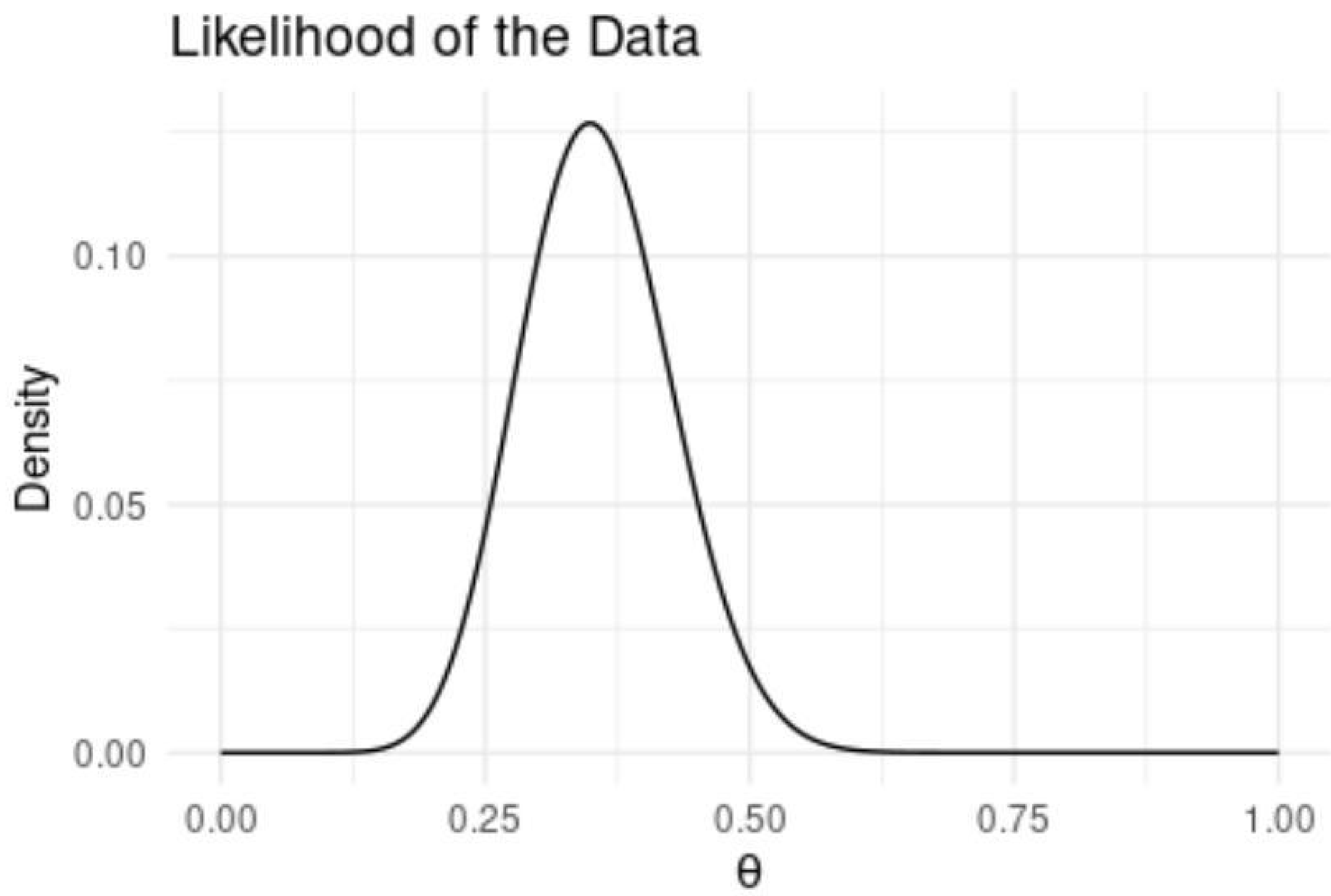


Figure 5.36: Likelihood Density plot.

```
     xlab=expression(theta), ylab="Density")

# Plotting the true density
theta_seq <- seq(0, 1, length.out = 1000)
true_density <- dbeta(theta_seq, 2, 8)
lines(theta_seq, true_density, col="red", lwd=2)

# Find the maximum value
max_density <-
  max(c(true_density, prior_density_estimation$y))

# Plot
plot(prior_density_estimation,
     main="Sampled prior beta(2,8) vs True Density",
     xlab=expression(theta), ylab="Density",
     ylim=c(0, max_density))
lines(theta_seq, true_density, col="red", lwd=2)

legend("topright", legend=c("Estimated", "True"),
       col=c("black", "red"), lty=1, cex=0.6,
       box.lty=0, inset = c(0.1, 0.01))
```

A more holistic approach that provides the posterior (5.37) in R:

```r
n <- 43; y <- 15
alpha <- 2; beta <- 8

theta_seq <- seq(0, 1, length.out = 100)

# The prior, likelihood, and posterior
prior <- dbeta(theta_seq, alpha, beta)
likelihood <- dbinom(y, n, theta_seq)
posterior <- dbeta(theta_seq, alpha + y, beta + n - y)

# The plot of the prior, likelihood and posterior
plot_data <- data.frame(
  Theta = rep(theta_seq, 3),
  Density = c(prior, likelihood, posterior),
  Distribution = rep(c("Prior", "Likelihood",
                       "Posterior"), each = 100))

ggplot(data = plot_data, aes(x = Theta, y = Density,
                             color = Distribution)) +
  geom_line() +
  labs(title = "Distribution Plots", x = expression(theta),
       y = "Density") +
  theme_minimal()

# The posterior mean, mode and standard deviation
posterior_mean <- (alpha + y) / (alpha + beta + n)
posterior_mode <- (alpha + y - 1) / (alpha + beta + n - 2)
posterior_sd <-
  sqrt((y + alpha) *
         (n - y + beta) /
         ((alpha + beta + n)^2 * (alpha + beta + n + 1)))

# The 95% credible interval
CI <- qbeta(c(0.025, 0.975), alpha + y, beta + n - y)
```

Providing posterior mean$= 0.321$, mode$= 0.314$, $sd = 0.064$ and 95% credible inter-

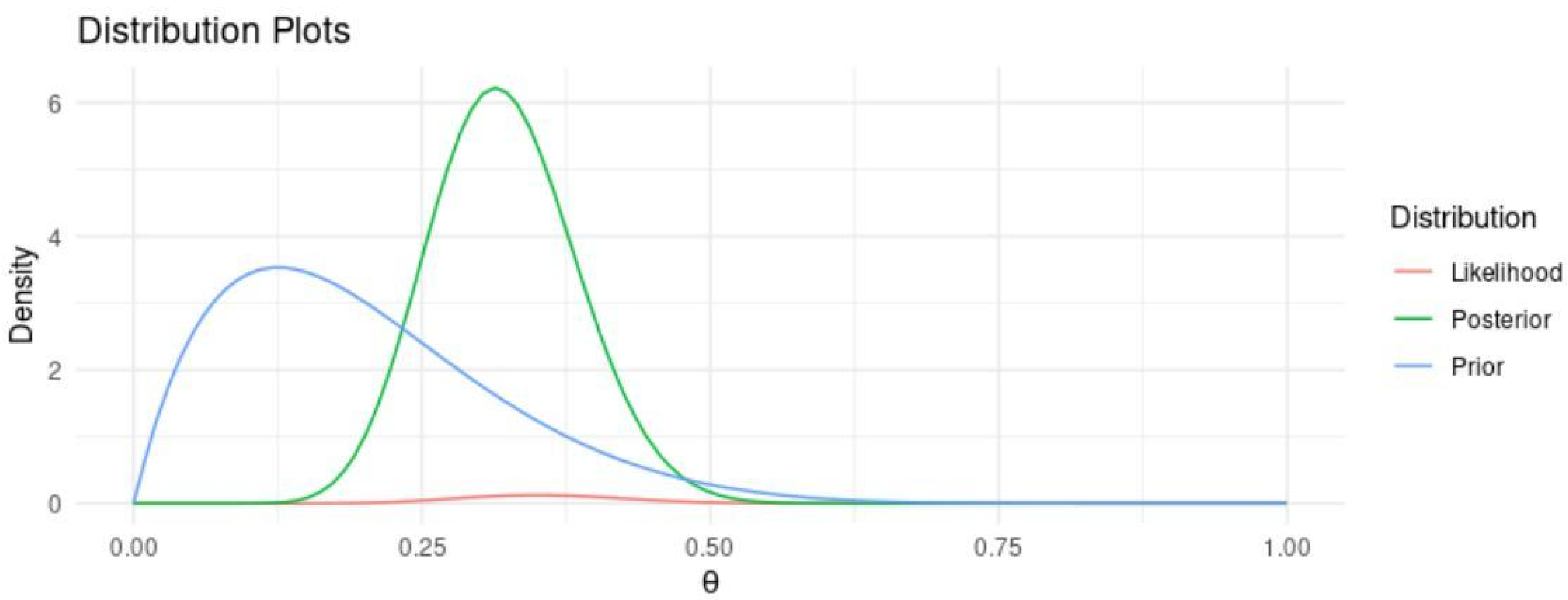


Figure 5.37: All the plots asked.

val $[0.203, 0.451]$

b) The idea is the same, but the prior is different, the graph5.38 represents the holistic answer approach. The same exact code where alpha= 8 and beta= 2

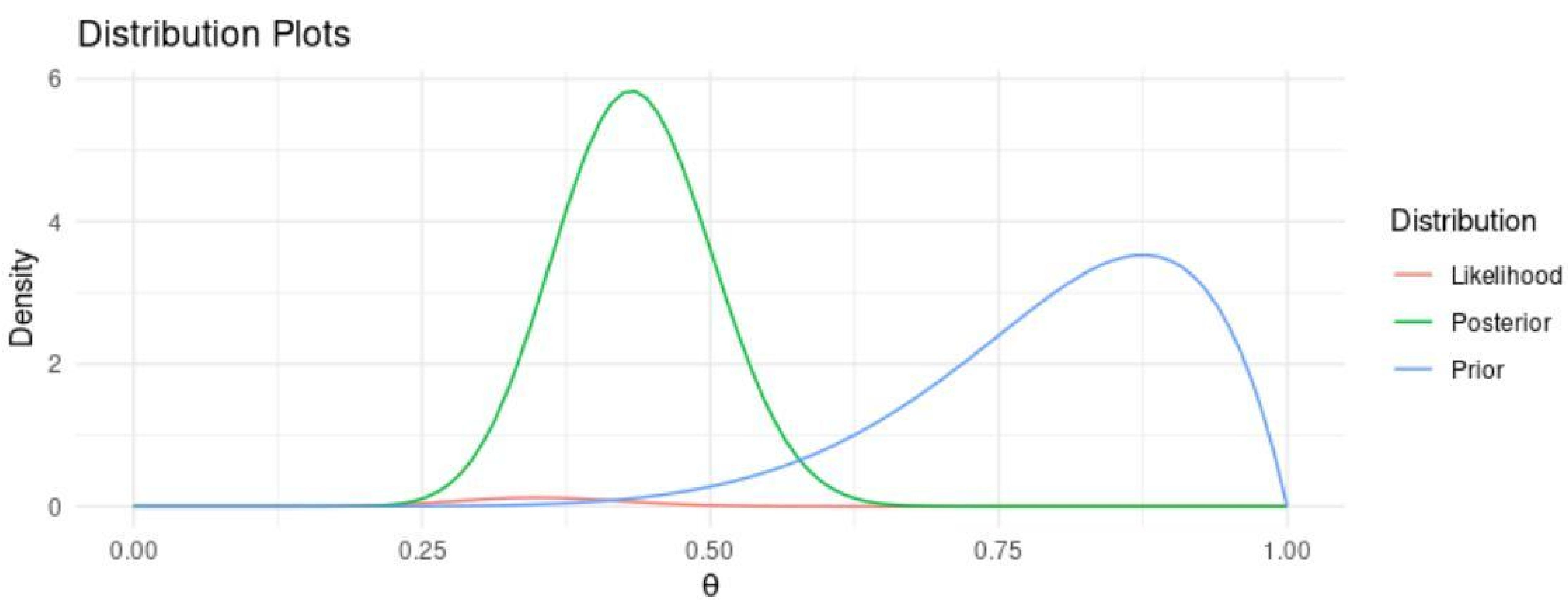


Figure 5.38: All the plots for the other prior.

c) The idea I presume is to make a weighted average from different distributions. In real-world problems, it could be explained as if you had some intel about the recidivism predictive model and you wanted to apply the intel in another recidivism problem, e.g. for another county, though you also want to keep the problem individuality, you contemplate using a weighted average prior. The reflection of the above can be visualized with the plot (5.39)

```
# Suppose the previous code...
# The complex prior function
complex_prior <- function(theta) {
  (1/4) * (gamma(10) / (gamma(2) * gamma(8))) *
```

```
    (3 * theta * (1 - theta)^7 + theta^7 *
       (1 - theta))}

complex_prior_values <- complex_prior(theta_seq)

# The plot complex prior
plot_data <- data.frame(
  Theta=rep(theta_seq, 3),
  Density=c(dbeta(theta_seq, 2, 8),
            dbeta(theta_seq, 8, 2),
            complex_prior_values),
  Distribution = rep(c("Beta(2, 8)", "Beta(8, 2)",
                       "Complex Prior"), each = 100))

ggplot(data = plot_data,
       aes(x = Theta, y = Density,
           color = Distribution)) + geom_line() +
  labs(title = "Prior Comparison", x = expression(theta),
       y = "Density") +
  theme_minimal()
```

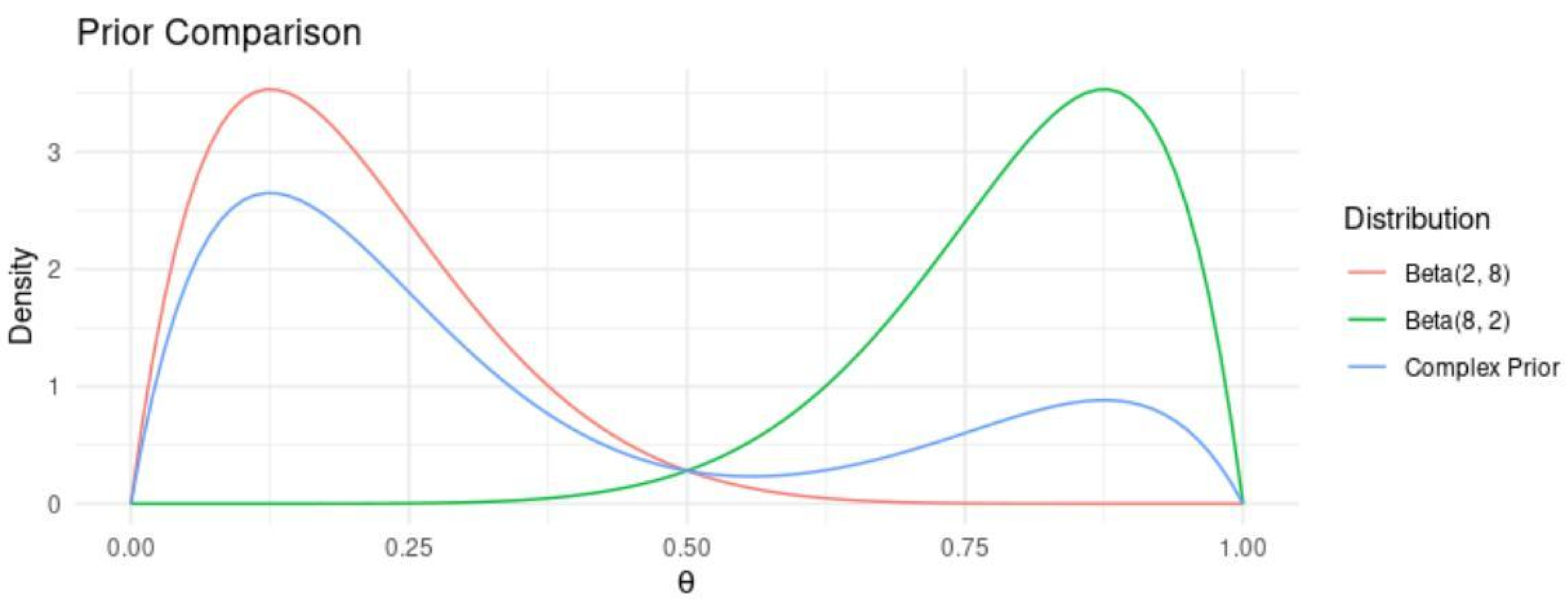


Figure 5.39: Visual aid in comparing all the priors in question.

d)i) To answer this question I will need the below representation of $\Gamma()$ function.

$$\Gamma(n) = (n-1)! \Rightarrow$$

$$\mathrm{p}(\theta) \times \mathrm{p}(y|\theta) = \frac{\binom{43}{15}}{4} \cdot \frac{\Gamma(10)}{\Gamma(2)\Gamma(8)}[3 \cdot \theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}] =$$
$$\frac{29 \cdot 31}{14 \cdot 13 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3}[3 \cdot \theta^{16}(1-\theta)^{35} + \theta^{22}(1-\theta)^{29}]$$

ii) In the part **a)** and **b)** we worked with beta$(2, 8)$ and beta$(8, 2)$, the posterior we work in **d)**, comes from a mixture of those priors and I fell confident in conjuring that the posterior will be a mixture of the corresponding posteriors beta() derived from these conjugate priors.
To prove it I need to see how the conjugates are affected and the reconfiguration of the weights. In our example the numbers have a funny way of working out, cause the constant of these to beta distributions and their corresponding posteriors is the same. Using the Bayes' theorem:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \Rightarrow \tag{5.48}$$
$$p(\theta|y) \propto \theta^{15}(1-\theta)^{28}(0.75 \cdot \mathrm{Beta}(2, 8) + 0.25 \cdot \mathrm{Beta}(8, 2)) \Rightarrow \tag{5.49}$$
$$p(\theta|y) \propto \theta^{15}(1-\theta)^{28}(0.75 \cdot \frac{\Gamma(10)}{\Gamma(2)\cdot\Gamma(8)} \cdot \theta^{2-1}(1-\theta)^{8-1} + \tag{5.50}$$
$$0.25 \cdot \frac{\Gamma(10)}{\Gamma(8)\cdot\Gamma(2)} \cdot \theta^{8-1}(1-\theta)^{2-1}) \Rightarrow \tag{5.51}$$
$$p(\theta|y) \propto 0.75 \cdot \theta^{16}(1-\theta)^{35} + 0.25 \cdot \theta^{22}(1-\theta)^{29} \Rightarrow \tag{5.52}$$
$$p(\theta|y) \propto 0.75 \cdot \theta^{17-1}(1-\theta)^{36-1} + 0.25 \cdot \theta^{23-1}(1-\theta)^{30-1} \Rightarrow \tag{5.53}$$
$$p(\theta|y) \propto \mathrm{w}_1 \cdot \mathrm{Beta}(17, 36) + \mathrm{w}_2 \cdot \mathrm{Beta}(23, 30) \Rightarrow \tag{5.54}$$
$$p(\theta|y) \propto \mathrm{w}_1 \cdot \mathrm{Beta}(2+15, 43+8-15) + \mathrm{w}_2 \cdot \mathrm{Beta}(8+15, 43+2-15) \tag{5.55}$$

Keep in mind

$$w_1 = \frac{\frac{0.75}{\Gamma(23)\cdot\Gamma(30)}}{\frac{0.75}{\Gamma(23)\cdot\Gamma(30)} + \frac{0.25}{\Gamma(17)\cdot\Gamma(36)}} = \frac{0.75\Gamma(17)\Gamma(36)}{0.75\Gamma(17)\Gamma(36) + 0.25\Gamma(22)\Gamma(30)}$$

and

$$w_2 = \frac{\frac{0.25}{\Gamma(17)\cdot\Gamma(36)}}{\frac{0.75}{\Gamma(23)\cdot\Gamma(30)} + \frac{0.25}{\Gamma(17)\cdot\Gamma(36)}} = \frac{0.25\Gamma(22)\Gamma(30)}{0.75\Gamma(17)\Gamma(36) + 0.25\Gamma(22)\Gamma(30)}$$

and we'll see more about it in **e)**.
iii) As the below code presents, we generate a sequence of θ values and use them to create posterior values which we then plot.

```r
library(ggplot2)
# The function for p(theta|y)
p_theta_given_y <- function(theta) {
  0.75 * theta^16 * (1-theta)^35 +
    0.25 * theta^22 * (1-theta)^29}

theta_vals <- seq(0, 1, length.out = 1000)
p_vals <- p_theta_given_y(theta_vals)

data <- data.frame(theta = theta_vals, p = p_vals)

# The plot p(theta|y) vs theta
ggplot(data, aes(x = theta, y = p)) +
  geom_line() +
  labs(title = "Posterior Distribution",
       x = expression(theta),
       y = expression(p(theta~'|'~y)))
```

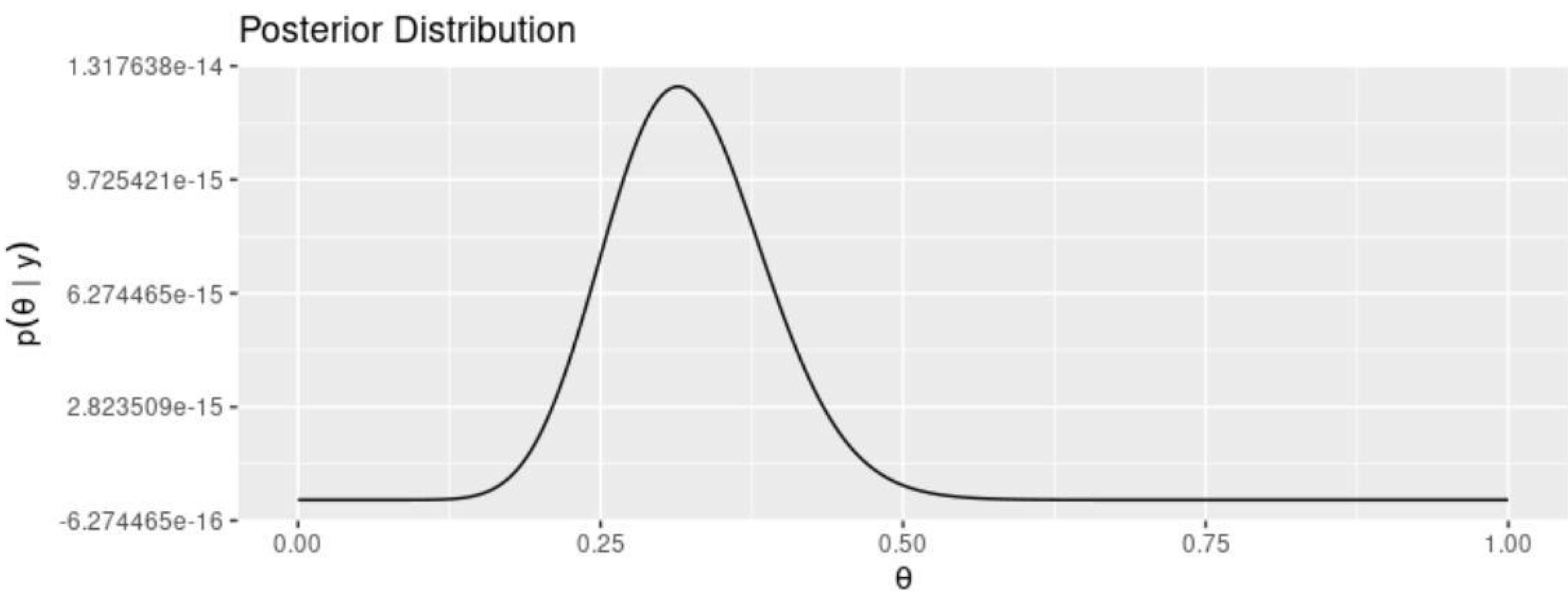


Figure 5.40: Visual aid of **d)iii)**.

For the mode:

```r
> # The theta that maximizes p(theta|y)
> max_theta <- theta_vals[which.max(p_vals)]
> print(max_theta)
[1] 0.3143143
```

A visual aid plot (5.41), which concludes **d)**:

```
# The function for p(theta/y) for the mixed prior
p_theta_given_y_mixed <- function(theta) {
  0.75 * theta^16 * (1-theta)^35 + 0.25 *
    theta^22 * (1-theta)^29}

# The p(theta/y) for the Beta(2,8)
p_theta_given_y_beta_2_8 <- function(theta) {
  theta^16 * (1-theta)^35}

# The p(theta/y) for the Beta(8,2)
p_theta_given_y_beta_8_2 <- function(theta) {
  theta^22 * (1-theta)^29}

theta_vals <- seq(0, 1, length.out = 1000)

# The p(theta/y) values for each prior
p_vals_mixed <- p_theta_given_y_mixed(theta_vals)
p_vals_beta_2_8 <- p_theta_given_y_beta_2_8(theta_vals)
p_vals_beta_8_2 <- p_theta_given_y_beta_8_2(theta_vals)

data <- data.frame(
  theta = theta_vals,
  p_mixed = p_vals_mixed,
  p_beta_2_8 = p_vals_beta_2_8,
  p_beta_8_2 = p_vals_beta_8_2
)

# The plot p(theta/y) vs theta for each prior
ggplot(data) +
  geom_line(aes(x = theta, y = p_mixed,
                colour = "Mixed Prior")) +
  geom_line(aes(x = theta, y = p_beta_2_8,
                colour = "Beta(2,8) Prior")) +
  geom_line(aes(x = theta, y = p_beta_8_2,
                colour = "Beta(8,2) Prior")) +
  labs(title = "Posterior Distributions",
```

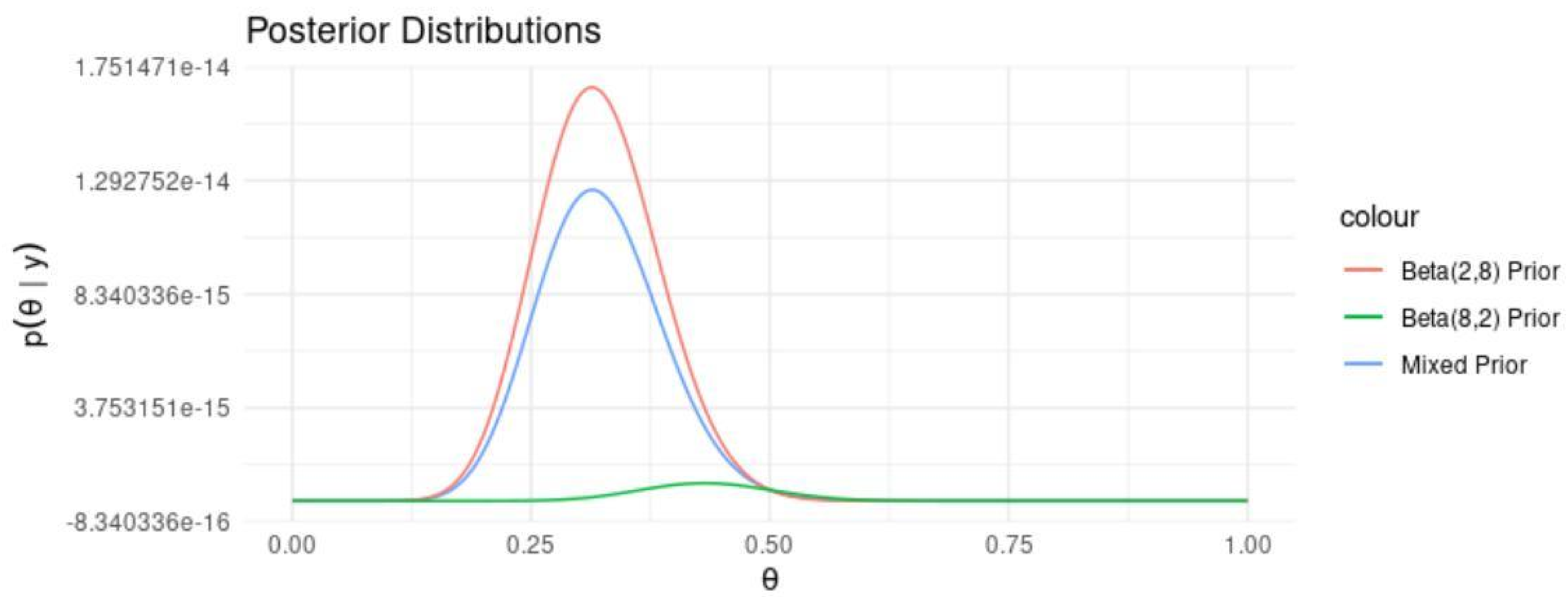


Figure 5.41: Visual aid for **d)iii)** comparisons.

```
      x = expression(theta),
      y = expression(p(theta~'|'~y))) +
  theme_minimal()
```

e) To answer that we need to go back into how we contracted our posterior in this problem and then go back to (4.6) in which the answer lies.

5.2 Mixtures of conjugate priors

[27]

Let $\mathrm{p}(y|\phi) = \mathrm{c}(\phi)\mathrm{h}(y)\exp\{\phi\cdot \mathrm{t}(y)\}$ be an exponential family model and let $p_1(\phi), \ldots \mathrm{p_K}(\phi)$ be K different members of the conjugate class of prior densities given in 2.2. A mixture of conjugate priors is given by $\tilde{p}(\theta) = \sum_{k=1}^{K} w_k \cdot \mathrm{p}_k(\theta)$, where the w_k's are all greater than zero and $\sum w_k = 1$ (see also Diaconis and Ylvisaker (1985)).

a) Identify the general form of the posterior distribution of θ, based on n i.i.d. samples from $\mathrm{p}(y|\theta)$ and the prior distribution given by $\tilde{p}$.

b) Repeat **a)** but in the special case that $\mathrm{p}(y|\theta) = \mathrm{dpois}(y, \theta)$ and $p_1, \ldots, p_K$ are gamma densities.

Solution:

a) To calculate the analytical posterior we need to assume the conjugates prior are of the form Diaconis and Ylvisaker worked and also assume $\theta = \phi$, so:

The Prior:

$$\begin{aligned}
\tilde{p}(\phi) &= \sum_{k=1}^{K} w_k p_k(\phi \mid n_{0,k}, t_{0,k}) \\
&= \sum_{k=1}^{K} (w_k \kappa(n_{0,k}, t_{0,k}) c(\phi)^{n_{0,k}} \exp(n_{0,k} t_{0,k} \phi))
\end{aligned}$$

then **the likelihood(model):**,

$$p(y_1, \ldots, y_n|\phi) = \prod_{k=1}^{n} h(y_i) c(\phi) e^{\phi t(y)} = c(\phi)^n e^{\phi \sum_{i=1}^{n} t(y_i)} \prod_{i=1}^{n} h(y_i)$$

thus, **the Posterior:**

$$\begin{aligned}
&p(\phi \mid y_1, \ldots, y_n) \propto p(\phi) p(y_1, \ldots, y_n \mid \phi) \\
&\propto \sum_{k=1}^{K} w_k \kappa(n_{0,k}, t_{0,k}) c(\phi)^{n_{0,k}} \exp\{n_{0,k} t_{0,k} \phi\} \times \exp\{\phi \sum_{i=1}^{n} t(y_i)\} c(\phi)^n \prod_{i=1}^{n} h(y_i) \\
&\propto \sum_{k=1}^{K} w_k \kappa(n_{0,k}, t_{0,k}) \prod_{i=1}^{n} h(y_i) c(\phi)^{n_{0,k}+n} \exp\left(\phi \times \left[n_{0,k} t_{0,k} + \sum_{i=1}^{n} t(y_i)\right]\right) \\
&\qquad \propto \sum_{k=1}^{K} w_k' p\left(\phi \,\middle|\, n_0 + n,\ n_0 t_0 + \sum_{i=1}^{n} t(y_i)\right)
\end{aligned}$$

[27] [DHo09]pg.229 Exercise 3.5

where, $w_k^{'} = \frac{w_k \kappa(n_{0,k}, t_{0,k}) \cdot \prod_{i=1}^{n} h(y_i)}{\kappa(k+n_{0,k}, n_{0,k}t_{0,k} + \sum_{k=1}^{K} t(y_i))}$ [28].

Even in the simpler problem of n=1 the problem exists as:

$$p(\phi|y) \propto w_k \kappa(n_{0,k}, t_{0,k}) h(y) c(\phi)^{n_{0,k}+1} \exp\{\phi \times (n_{0,k}t_{0,k} + t(y))\}$$

So the posterior is another weighted mixture. However the weights of the relative components are not preserved. [29]

b) [30] **The General Approach:** Suppose $y_1, \ldots, y_n$ i.i.d. $\sim$ Poisson(θ) and the priors to be of Gamma form, then:

Prior: $\sum_{k=1}^{K} w_k \kappa(n_{0,k}, t_{0,k}) c(\phi) e^{t_{0,k}\phi} \Rightarrow$

$$\tilde{p} = \sum_{k=1}^{K} \frac{b_k^{a_k}}{\Gamma(a_k)} \theta^{a_k - 1} e^{-b_k \theta}$$

The Likelihood:

$$p(y_1, \ldots, y_n) = \prod_{i=1}^{n} \frac{1}{y_i!} \theta^{y_i} e^{-\theta} = \theta^{\sum_{i=1}^{n} y_i} e^{-n\theta} \prod_{i=1}^{n} \frac{1}{y_i!}$$

The Posterior:

$$p(\theta|y_1, \ldots, y_n) \propto \sum_{k=1}^{K} \prod_{i=1}^{n} \frac{w_k}{y_i!} \frac{b_k^{a_k}}{\Gamma(a_k)} \theta^{a_k + \sum_{i=1}^{n} y_i - 1} e^{\theta(-b_k - n)} =$$
$$\sum_{k=1}^{K} \frac{\prod_{i=1}^{n} \frac{w_k}{y_i!} b_k^{a_k}}{\Gamma(a_k)} \frac{\Gamma(a_k + \sum_{i=1}^{n} y_i)}{(b_k + n)^{a_k + \sum_{i=1}^{n} y_i}} \text{dgamma}(\theta, a_k + \sum_{i=1}^{n} y_i, b_k + n)$$

so,

$$w_k^{'} = \frac{\prod_{i=1}^{n} \frac{w_k}{y_i!} b_k^{a_k}}{\Gamma(a_k)} \frac{\Gamma(a_k + \sum_{i=1}^{n} y_i)}{(b_k + n)^{a_k + \sum_{i=1}^{n} y_i}}$$

[28] This is probably wrong, is based on the material from Peter D. Hoff's book, where in pg. 51 he produced the posterior without taking into account that the number you raise in e is $n_0 \cdot t_0$ or there is a problem in the stated prior, due to limited time, I haven't seen the original Diaconis's and Ylvisaker's work, so I hope on some feedback

[29] I hope everyone noticed the product of y_i variables and contemplated if it belongs there, as I did, the answer to this I don't know and given the limited time for submission, I look forward to any feedback on this.

[30] There is a mistake in Peter D. Hoff, as in the prior we have $e^{t_0 \phi}$ and not $e^{t_0 \cdot k_0 \phi}$ as stated, otherwise the gamma prior doesn't makes much sense...

For one variable y and in the special case where p($y|\theta$) is a Poisson distribution and each p_k is a gamma distribution, we know that the gamma distribution is the conjugate prior for the Poisson distribution. This means that each posterior distribution p($\theta|y, p_k$) will also be a gamma distribution.
Given the likelihood function for the Poisson distribution:

$$p(y|\theta) = \frac{e^{-\theta}\theta^{y}}{y!} \tag{5.56}$$

Assuming each $p_k(\theta)$ follows a gamma distribution with α_k and β_k we have:

$$p_k(\theta) = \frac{\beta_k^{\alpha_k}\theta^{\alpha_k-1}e^{-\beta_k\theta}}{\Gamma(\alpha_k)} \tag{5.57}$$

So the posterior distribution will be given by a weighted sum of gamma distributions:

$$p(\theta|y) \propto \sum_{k=1}^{K} \frac{w_k b_k^{a_k}}{\Gamma(a_k)y} \cdot \frac{\Gamma(a_k+y)}{(b_k+1)^{a_k+y}} \text{Gamma}(\alpha_k+y, \beta_k+1) \tag{5.58}$$

So the general form of the posterior distribution will be a mixture of gamma distributions with updated weights.

$$w_k^{'} = \frac{w_k b_k^{a_k}}{\Gamma(a_k)y} \cdot \frac{\Gamma(a_k+y)}{(b_k+1)^{a_k+y}}$$

5.3 Mixtures of conjugate priors

[31]

For the posterior density from 5.1:

a) Make a plot of p($\theta|y$) or p($y|\theta$)p(θ) using the mixture prior distribution and a dense sequence of θ-values. Can you think of a way to obtain a 95% quantile-based posterior confidence interval for θ? You might want to try some sort of discrete approximation.

b) To sample a random variable z from the mixture distribution $wp_1(z) + (1 - w) \cdot p_0(z)$, first toss a w-coin and let x be the outcome (this can be done in R with

```
x <- rbinom(1, 1, w) ).
```

Then if $x = 1$ sample z from p_1, and if $x = 0$ sample z from p_0. Using this technique, obtain a Monte Carlo approximation of the posterior distribution p($\theta|y$) and a 95% quantile-based confidence interval, and compare them to the results in part **a)**.

Solution:

a) To plot p($\theta|y$) or p($y|\theta$)p(θ) using the mixture prior distribution (5.42), we can build upon the script used in the exercise (4.6). Then numerically find a 95% quantile-based posterior confidence interval for θ by using a method such as bootstrapping or, as hinted, some sort of discrete approximation where we numerically find the quantiles based on our grid of θ values. The code will be fairly similar to the corresponding code/answer of (5.1):

```
library(ggplot2)

# The new weights
w_1 = 0.75 * factorial(16) * factorial(35) /
  (0.75 * factorial(16) * factorial(35) + 0.25 *
     factorial(22) * factorial(29))
w_2 = 0.25 * factorial(22) * factorial(29) /
  (0.75 * factorial(16) * factorial(35) + 0.25 *
     factorial(22) * factorial(29))

# The function for p(theta|y) for the mixed prior
p_theta_given_y_mixed <- function(theta) {
  w_1 * dgamma(theta, 17, 36) + w_2 * dgamma(theta, 23, 30)
```

[31] [DHo09]pg.233 Exercise 4.4

```
}

theta_vals <- seq(0, 1, length.out = 10000)
p_vals_mixed <- sapply(theta_vals, p_theta_given_y_mixed)

# A data frame
data <- data.frame(
  theta = theta_vals,
  p_mixed = p_vals_mixed
)

# The plot p(theta|y) vs theta for the mixed prior
ggplot(data) +
  geom_line(aes(x = theta, y = p_mixed,
                colour = "Mixed Prior")) +
  labs(title = "Posterior Distribution",
       x = expression(theta),
       y = expression(p(theta~'|'~y))) +
  theme_minimal()


# Normalize the posterior density values to use as weights
weights <- p_vals_mixed / sum(p_vals_mixed)

# The weighted quantiles for the 95% confidence interval
weighted_quantile <- function(x, probs, weights) {
  sorted_data <- sort(x)
  cum_weights <- cumsum(weights[order(x)])
  approx(cum_weights, sorted_data, xout = probs)$y}

conf_interval <- weighted_quantile(
  theta_vals, c(0.025, 0.975), weights)

#Some useful info for b)
#The mean
posterior_mean <- sum(data$theta * data$p_mixed) /
```

```r
    sum(data$p_mixed)
#The mode
posterior_mode <- data$theta[which.max(data$p_mixed)]
#The median
cum_density <- cumsum(data$p_mixed) / sum(data$p_mixed)
posterior_median <- data$theta[which.min(abs(
cum_density - 0.5))]
```

```
> conf_interval
[1] 0.275507 0.740893
>posterior_mean
#[1] 0.4758319
> posterior_mode
#[1] 0.4447445
> posterior_median
#[1] 0.4648465
```

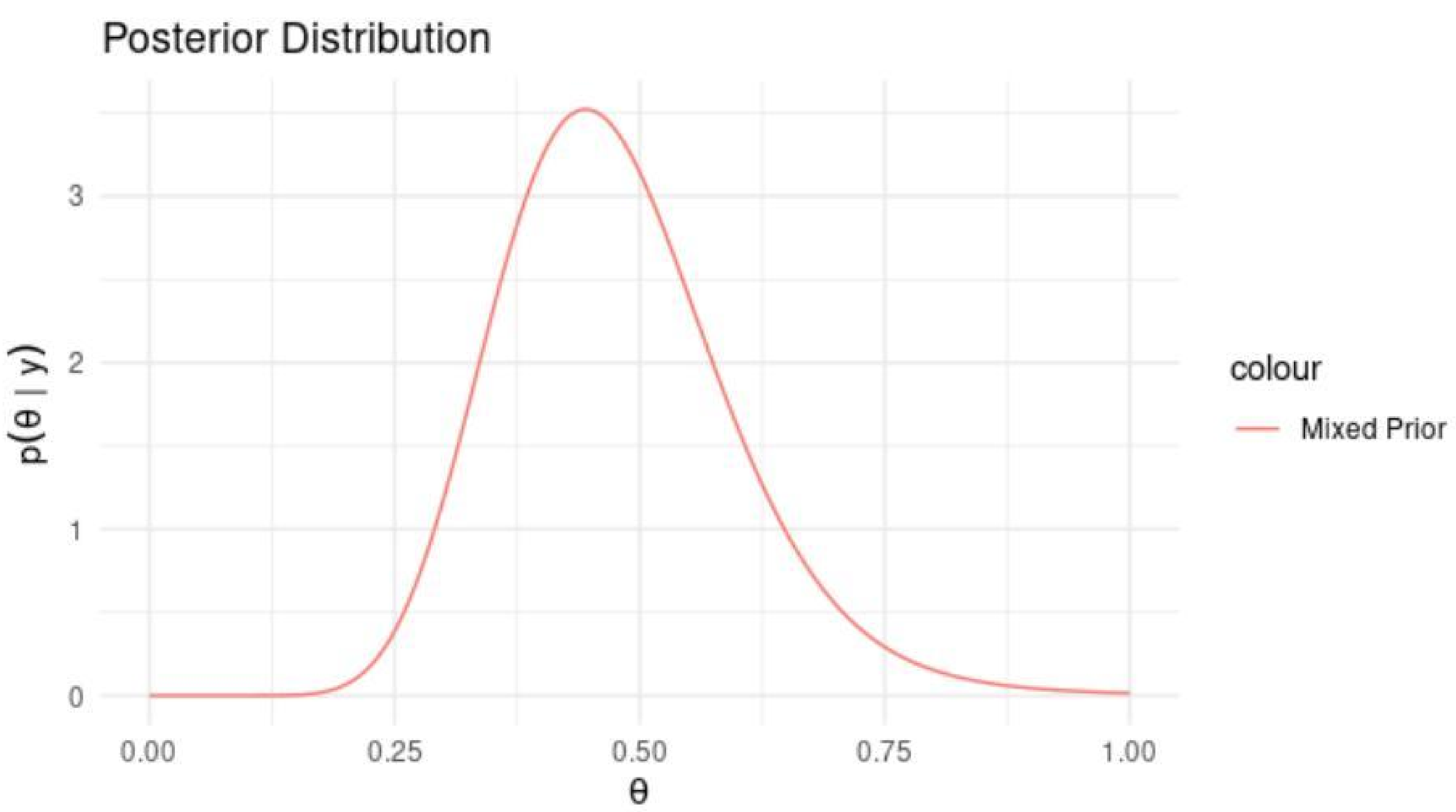


Figure 5.42: The plot for **a)**.

b) I'll start by defining the w_1, w_2 as the weights that correspond to the proportions of the two beta distributions in our mixture, then simulate the draws and plot 5.43 their histogram and calculate the confidence interval will be calculated as shown in the below R-code:

```r
N <- 50000
```

```
# The weights
w_1 <- 0.75 * factorial(16) * factorial(35) /
  (0.75 * factorial(16) * factorial(35) + 0.25 *
     factorial(22) * factorial(29))
w_2 <- 0.25 * factorial(22) * factorial(29) /
  (0.75 * factorial(16) * factorial(35) + 0.25 *
     factorial(22) * factorial(29))

# Sampling from the mixture distribution
samples <- numeric(N)
for (i in 1:N) {
  x <- rbinom(1, 1, w_1)
  if (x == 1) {
    # Sample from p_1 (posterior with beta(2,8) prior)
    samples[i] <- rbeta(1, 17, 36)
  } else {
    # Sample from p_0 (posterior with beta(8,2) prior)
    samples[i] <- rbeta(1, 23, 30)
  }
}

# The 95% quantile-based confidence interval for the samples
conf_interval_b <- quantile(samples, probs = c(0.025, 0.975))

# Plotting the histogram of the samples
hist(samples, breaks = 50, probability = TRUE,
     main = "Monte Carlo approximation of p(θ|y)",
     xlab = expression(theta), ylab = "Density")
abline(v = conf_interval_b, col = "red", lwd = 2)

#the mode using the histogram
hist_data <- hist(samples, breaks = 50, plot = FALSE)
mode_index <- which.max(hist_data$counts)
mode_b <- (hist_data$breaks[mode_index] +
             hist_data$breaks[mode_index + 1]) / 2
```

```
> conf_interval_b
#        2.5%       97.5%
#0.2049872  0.4566495
> mean(samples)
#[1]  0.3224323
> median(samples)
#[1]  0.3195992
> mode_b
#[1]  0.305
```

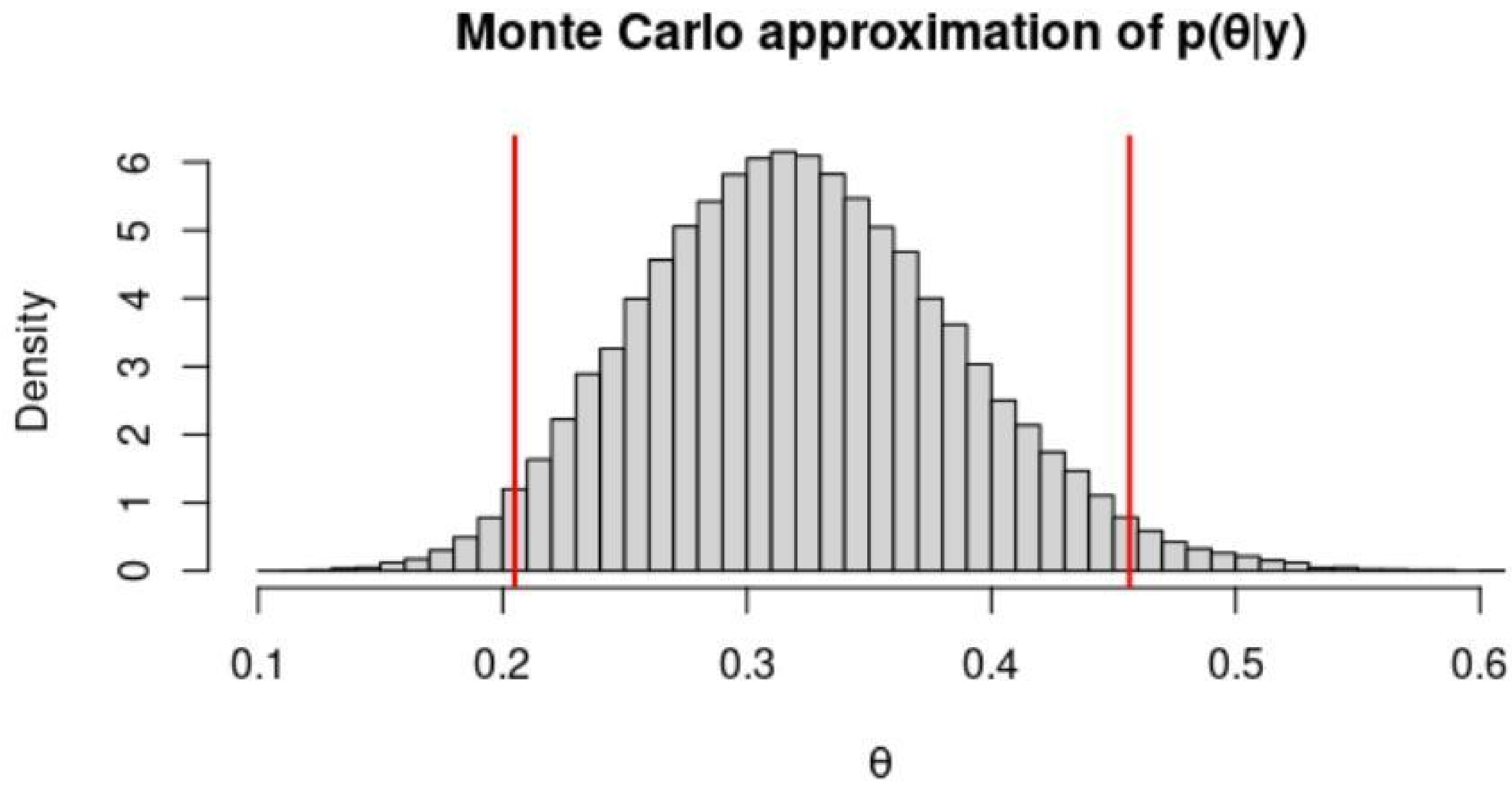


Figure 5.43: The plot for **b)**.

The 5.43 represents a Monte Carlo approximation of the posterior distribution p($\theta|y$). It is created by drawing a large number of samples (in this case, 50.000 samples) from the corresponding to the weight that won, posterior distribution. The difference between these two methods is great as in the **b)** approach the confidence interval, the mean, the median and the mode are different, the only logical explanation I can find given my approach is not wrong in both cases is that the outlier or the extreme values of those distributions affected it so, thought it doesn't not explain the great difference in the normalized densities values. [32].

[32] I am stunned as I notice the difference between these two methods and I feel like I made a mistake somewhere along the way, I genuinely hope and wish for your feedback on this one, because a mistake, or my uncertainty, here could take me back to Task4

Task 6

6.1 Studying

[33]

The files `school1.dat`, `school2.dat`, and `school3.dat` contain data on the amount of time students from three high schools spent on studying or homework during an exam period. Analyze data from each of these schools separately, using the normal model with a conjugate prior distribution, in which

$$\{\mu_0 = 5, \sigma_0^2 = 4, \kappa_0 = 1, \nu_0 = 2\}$$

and compute or approximate the following:

a) Posterior means and 95% confidence intervals for the mean θ and standard deviation σ from each school;

b) The posterior probability that $\theta_i < \theta_j < \theta_k$ for all six permutations {i, j, k} of {1, 2, 3};

c)The posterior probability that $\tilde{Y}_i < \tilde{Y}_j < \tilde{Y}_k$ for all six permutations {i, j, k} of {1, 2, 3}, where $\tilde{Y}_i$ is a sample from the posterior predictive distribution of school i.

d) Compute the posterior probability that θ_1 is bigger than both θ_2 and θ_3, and the posterior probability that $\tilde{Y}_1$ is bigger than both $\tilde{Y}_2$ and $\tilde{Y}_3$.

Solution:

a) Given the problem, we are provided with a set of parameters for the normal-inverse-gamma (NIG) prior distribution. Specifically, the parameters are:

$$\mu_0 = 5; \qquad \sigma_0^2 = 4; \qquad \kappa_0 = 1; \qquad \nu_0 = 2.$$

The interpretation of these parameters is as follows:

- μ_0: Prior mean of the normal distribution.
- σ_0^2: Prior variance of the normal distribution.

[33] [DHo09]pg. 235 Exercise 5.1

- κ_0: Represents the effective sample size of the prior information.
- ν_0: Represents the degrees of freedom for the inverse gamma distribution, which models the variance.

Given new data, denoted as $y_1, y_2, \dots, y_n$, from each school, we can update these prior parameters to compute the posterior parameters. The sample statistics from the data are:

$$\begin{aligned} n &: \text{Sample size}, \\ \bar{y} &: \text{Sample mean}, \\ s^2 &: \text{Sample variance}. \end{aligned}$$

The formulas to update the parameters based on the sample data are:

$$\begin{aligned} \kappa_n &= \kappa_0 + n, \\ \nu_n &= \nu_0 + n, \\ \mu_n &= \frac{\kappa_0 \mu_0 + n\bar{y}}{\kappa_0 + n}, \\ \sigma_n^2 &= \frac{\nu_0 \sigma_0^2 + (n-1)s^2 + \kappa_0 n(\bar{y} - \mu_0)^2/(\kappa_0 + n)}{\nu_0 + n}. \end{aligned}$$

Note: [34] Given the normal likelihood and prior distributions:

$$\begin{aligned} \frac{1}{\sigma^2} &\sim \text{gamma}(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}) \\ \theta|\sigma^2 &\sim \text{normal}(\mu_0, \frac{\sigma^2}{\kappa_0}) \\ Y_1, \dots, Y_n &\sim \text{i.i.d. normal}(\theta, \sigma^2) \end{aligned}$$

The posterior distribution of θ and σ are going to be:

$$\begin{aligned} \{\theta|y_1, \dots, y_n, \sigma^2\} &\sim \text{normal}(\mu_n, \frac{\sigma^2}{\kappa_n}) \\ \{\frac{1}{\sigma^2}|y_1, \dots, y_n\} &\sim \text{gamma}(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}) \end{aligned}$$

Using this, we can compute the posterior mean and variance but also derive the 95% confidence intervals for θ and σ for each school.

[34] For more information see [DHo09]pg.74-75

```
library(stats); library(httr)

# The prior parameters
mu_0=5; sigma2_0=4; kappa_0 = 1; nu_0 = 2

# Fetching the data from the URL, using this method
# for security reasons on there end (the url cyber place)
fetch_data <- function(url) {
  response <- GET(url, config(ssl_verifypeer = FALSE))
  scan(textConnection(content(response, "text")),
       quiet = TRUE)}

# The posterior parameters
compute_posterior <- function(data) {
  n <- length(data)
  y_bar <- mean(data)
  s2 <- var(data)

  kappa_n <- kappa_0 + n
  nu_n <- nu_0 + n
  mu_n <- (kappa_0 * mu_0 + n * y_bar) / kappa_n
  sigma2_n <- 1/nu_n *
    (nu_0*sigma2_0 + (n-1)*s2 +
       kappa_0*n*(y_bar - mu_0)^2 / (kappa_0 + n))

  return(list(mu_n = mu_n, sigma2_n = sigma2_n,
              kappa_n = kappa_n, nu_n = nu_n))}

# URLs for data
urls <- list(
  school1 =
    'https://www2.stat.duke.edu/~pdh10/FCBS/Exercises/school1.dat',
  school2 =
    'https://www2.stat.duke.edu/~pdh10/FCBS/Exercises/school2.dat',
  school3 =
    'https://www2.stat.duke.edu/~pdh10/FCBS/Exercises/school3.dat')
```

```r
# Analysis per school to compute the posterior means
#and 95% confidence intervals
results <- list()
for (school in names(urls)) {
  data <- fetch_data(urls[[school]])
  posterior <- compute_posterior(data)

  # Monte Carlo sampling for theta and sigma
  s2_samples <- 1 /
    rgamma(5000, posterior$nu_n / 2,
           posterior$nu_n * posterior$sigma2_n / 2)
  theta_samples <- rnorm(5000,
                         posterior$mu_n,
                         sqrt(s2_samples / posterior$kappa_n))

  # The 95% confidence intervals
  theta_ci <- quantile(theta_samples,
                       probs = c(0.025, 0.5, 0.975))
  sigma_ci <- sqrt(quantile(s2_samples,
                            probs = c(0.025, 0.5, 0.975)))

  results[[school]] <- list(theta_ci = theta_ci,
                            sigma_ci = sigma_ci)}

# Print the results
for (school in names(results)) {
  cat("Results for", school, ":\n")
  cat("Theta - Posterior Mean:", posterior$mu_n, "\n")
  cat("Theta - 95% CI:", results[[school]]$theta_ci, "\n")
  cat("Sigma - Posterior Mean:",
      sqrt(posterior$sigma2_n), "\n")
  cat("Sigma - 95% CI:",
      results[[school]]$sigma_ci, "\n\n")}
```

With output [35]:

Results for school1 :

Theta - Posterior Mean: 7.812381

Theta - 95% CI: 7.731923 9.291251 10.81001

Sigma - Posterior Mean: 3.61849

Sigma - 95% CI: 3.003774 3.866621 5.166409

Results for school2 :

Theta - Posterior Mean: 7.812381

Theta - 95% CI: 5.248633 6.957673 8.766324

Sigma - Posterior Mean: 3.61849

Sigma - 95% CI: 3.330012 4.329596 5.867488

Results for school3 :

Theta - Posterior Mean: 7.812381

Theta - 95% CI: 6.158554 7.770961 9.42133

Sigma - Posterior Mean: 3.61849

Sigma - 95% CI: 2.793323 3.672032 5.106151

b)To compute those, we'll use the Monte Carlo samples of θ that we generated for each school in the previous step. We'll then compare these samples to determine the probability of each permutation using a two-step approach.
1. For each permutation $\{i, j, k\}$, we are going to count the number of times the condition $\theta_i < \theta_j < \theta_k$ is satisfied across the Monte Carlo samples.
2. Divide that count by the total number of samples to normalize and get the approximation of the posterior probability for each permutation.
The Code:

```
#...[Given the data and the posterior parameters
# from the previous code] :

# Function to for the probability of a permutation
compute_permutation_probability <- function(
    samples_i, samples_j, samples_k) {
  sum(samples_i < samples_j & samples_j < samples_k) /
    length(samples_i)
```

[35]The actual output was of the form of school$_{\text{number}}$ so I just replaced it with $\tilde{Y}_{\text{number}}$

```r
}

# Monte Carlo samples for theta for each school
theta_samples_list <- list()
for (school in names(urls)) {
  data <- fetch_data(urls[[school]])
  posterior <- compute_posterior(data)

  # Sampling for theta
  s2_samples <- 1 /
    rgamma(5000,posterior$nu_n / 2,
           posterior$nu_n *
             posterior$sigma2_n / 2)

  theta_samples <- rnorm(
    5000, posterior$mu_n,
    sqrt(s2_samples / posterior$kappa_n))

  theta_samples_list[[school]] <- theta_samples}

# The posterior probability for each permutation
permutations <- list(c("school1", "school2", "school3"),
                     c("school1", "school3", "school2"),
                     c("school2", "school1", "school3"),
                     c("school2", "school3", "school1"),
                     c("school3", "school1", "school2"),
                     c("school3", "school2", "school1"))

for (perm in permutations) {
  prob <- compute_permutation_probability(
    theta_samples_list[[perm[1]]],
    theta_samples_list[[perm[2]]],
    theta_samples_list[[perm[3]]])
  cat("Probability for",
      perm[1], "<", perm[2], "<", perm[3], ":", prob, "\n")}
```

With output:

$$\text{Probability for school1} < \text{school2} < \text{school3} : 0.0056$$
$$\text{Probability for school1} < \text{school3} < \text{school2} : 0.0028$$
$$\text{Probability for school2} < \text{school1} < \text{school3} : 0.0826$$
$$\text{Probability for school2} < \text{school3} < \text{school1} : 0.6780$$
$$\text{Probability for school3} < \text{school1} < \text{school2} : 0.0170$$
$$\text{Probability for school3} < \text{school2} < \text{school1} : 0.2140$$

c) The posterior probability that $\tilde{Y}_i < \tilde{Y}_j < \tilde{Y}_k$ for all six permutations $\{i, j, k\}$ of $\{1, 2, 3\}$, where $\tilde{Y}_i$ is a sample from the posterior predictive distribution of school i. The posterior predictive distribution for a new observation $\tilde{Y}$ in a normal model with known variance σ^2 and unknown mean θ is:

$$\tilde{Y}|y \sim N(\theta, \sigma^2 + \frac{\sigma^2}{n})$$

Where:

1. θ is the posterior mean.
2. σ^2 is the posterior variance.
3. n is the sample size

Given this, we can generate samples from the posterior predictive distribution for each school and then compute the probability for each permutation as we did in part **b)** using a three-step approach.
1. Generate samples from the posterior predictive distribution using the Monte Carlo samples of θ and σ^2 that we already have for each school.
2. For each permutation $\{i, j, k\}$, count the number of times the condition $\tilde{Y}_i < \tilde{Y}_j < \tilde{Y}_k$ is satisfied across the Monte Carlo samples.
3. Divide the count by the total number of samples to get the posterior probability for that permutation.
The Code:

```
# ... [Previous code for fetching data and computing
#posterior parameters]

# Function for the probability of a permutation
#for Y_tilde
```

```
compute_permutation_probability_Ytilde <- function(
    samples_i, samples_j, samples_k) {
  sum(samples_i < samples_j & samples_j < samples_k) /
    length(samples_i)}

# Monte Carlo samples for Y_tilde for each school
Ytilde_samples_list <- list()
for (school in names(urls)) {
  data <- fetch_data(urls[[school]])
  posterior <- compute_posterior(data)

  # Sampling for Y_tilde
  s2_samples <- 1 /
    rgamma(5000,
           posterior$nu_n / 2,
           posterior$nu_n * posterior$sigma2_n / 2)
  theta_samples <-
    rnorm(5000,
          posterior$mu_n,
          sqrt(s2_samples / posterior$kappa_n))
  Ytilde_samples <-
    rnorm(5000,
          theta_samples,
          sqrt(s2_samples + s2_samples/length(data)))

  Ytilde_samples_list[[school]] <- Ytilde_samples}

# The posterior probability for each permutation
for (perm in permutations) {
  prob <-
    compute_permutation_probability_Ytilde(
      Ytilde_samples_list[[perm[1]]],
      Ytilde_samples_list[[perm[2]]],
      Ytilde_samples_list[[perm[3]]])
  cat("Probability for",
      perm[1], "<",
```

```
    perm[2], "<",
    perm[3], ":", prob, "\n")}
```

With output:

$$
\begin{aligned}
&\text{Probability for } \tilde{Y}_1 < \tilde{Y}_2 < \tilde{Y}_3 : 0.1142\\
&\text{Probability for } \tilde{Y}_1 < \tilde{Y}_3 < \tilde{Y}_2 : 0.1026\\
&\text{Probability for } \tilde{Y}_2 < \tilde{Y}_1 < \tilde{Y}_3 : 0.1846\\
&\text{Probability for } \tilde{Y}_2 < \tilde{Y}_3 < \tilde{Y}_1 : 0.2676\\
&\text{Probability for } \tilde{Y}_3 < \tilde{Y}_1 < \tilde{Y}_2 : 0.1398\\
&\text{Probability for } \tilde{Y}_3 < \tilde{Y}_2 < \tilde{Y}_1 : 0.1912
\end{aligned}
$$

Note: In this part I would like to undress a dispute, even though our approach is correct, we would like to present an alternative solution and let you be the judge. To enhance our understanding and the presentability of the document I will just present the part of the code, that showcases the logical difference in these approaches.

The Difference:

```
# Generate Monte Carlo samples for Y_tilde for each school
Ytilde_samples_list <- list()
for (school in names(urls)) {
  data <- fetch_data(urls[[school]])
  posterior <- compute_posterior(data)

  # Monte Carlo sampling for Y_tilde
  s2_samples <- 1 /
    rgamma(5000,
           posterior$nu_n / 2,
           posterior$nu_n * posterior$sigma2_n / 2)
  theta_samples <-
    rnorm(5000,
          posterior$mu_n,
          sqrt(s2_samples / posterior$kappa_n))
  Ytilde_samples <-
    rnorm(5000,
          theta_samples,
          sqrt(s2_samples))  # <-- This line is changed
```

```
Ytilde_samples_list[[school]] <- Ytilde_samples}
```

With output:

Probability for $\tilde{Y_1} < \tilde{Y_2} < \tilde{Y_3}$: 0.1142
Probability for $\tilde{Y_1} < \tilde{Y_3} < \tilde{Y_2}$: 0.1056
Probability for $\tilde{Y_2} < \tilde{Y_1} < \tilde{Y_3}$: 0.1740
Probability for $\tilde{Y_2} < \tilde{Y_3} < \tilde{Y_1}$: 0.2748
Probability for $\tilde{Y_3} < \tilde{Y_1} < \tilde{Y_2}$: 0.1340
Probability for $\tilde{Y_3} < \tilde{Y_2} < \tilde{Y_1}$: 0.1974

As we can observe by the probabilities is difficult to know if the difference in approach made a mark because we drew Monte Carlo Samples, so the fact that the differences in value are of the 10^{-2} magnitude at best, I am unable to fully comprehend the actual effect it had in our predictions. A possible explanation could be that these ways of sampling answer different questions. An insight provided by the chat was that the approach I presented last, draws samples from the posterior predictive distribution of $\tilde{Y}$ for a new observation, from the same population (i.e., a new student from the same school), but not considering the variability within the observed data. Where the first approach (with the variance of the Post. Pred. Distr. set as $\sqrt{\sigma^2_{\text{sample}} + \frac{\sigma^2_{\text{sample}}}{n}}$) draws samples from the posterior predictive distribution of $\tilde{Y}$ for the average of a new sample of students of the same size as the original sample from the same school.

d) [36] Given our previous work, this is a straightforward extension of our previous algorithmic approach. We'll use the Monte Carlo samples of θ and $\tilde{Y}$ that we've already generated, count the number of samples where θ_1 and $\tilde{Y_1}$ are bigger than both of the other θ and $\tilde{Y}$ respectfully and then divide each count by the total. **The Code:**

```
# ... [Previous code for fetching data, computing posterior
# parameters, and generating Monte Carlo samples]

# Compute the posterior probability that theta_1 is bigger
# than both theta_2 and theta_3
```

[36] Based on the similarity of the probabilities and the different variances, we will proceed using the first approach.

```
prob_theta1_bigger <-
  sum(theta_samples_list[['school1']] >
        theta_samples_list[['school2']] &
        theta_samples_list[['school1']] >
        theta_samples_list[['school3']]) / 5000
cat("Posterior probability that theta_1 is bigger than both
    theta_2 and theta_3:",prob_theta1_bigger, "\n")

# Compute the posterior probability that Ytilde_1 is bigger
# than both Ytilde_2 and Ytilde_3
prob_Ytilde1_bigger <-
  sum(Ytilde_samples_list[['school1']] >
        Ytilde_samples_list[['school2']] &
        Ytilde_samples_list[['school1']] >
        Ytilde_samples_list[['school3']]) / 5000
cat("Posterior probability that Ytilde_1 is bigger than both
    Ytilde_2 and Ytilde_3:", prob_Ytilde1_bigger, "\n")
```

With output:

The posterior probability where θ_1 is bigger than both θ_2 and θ_3: 0.892
The posterior probability where $\tilde{Y_1}$ is bigger than both $\tilde{Y_2}$ and $\tilde{Y_3}$: 0.4588

6.2 Sensitivity analysis

[37]

Thirty-two students in a science classroom were randomly assigned to one of two study methods, A and B, so that $n_A = n_B = 16$ students were assigned to each method. After several weeks of study, students were examined on the course material with an exam designed to give an average score of 75 with a standard deviation of 10. The scores for the two groups are summarized by $\{\bar{y_A} = 75.2, s_A = 7.3\}$ and $\{\bar{y_B} = 77.5, s_B = 8.1\}$. Consider independent, conjugate normal prior distributions for each of θ_A and θ_B, with $\mu_0 = 75$ and $\sigma_0^2 = 100$ for both groups. For each $(\kappa_0, \nu_0) \in \{(1, 1), (2, 2), (4, 4), (8, 8), (16, 16), (32, 32)\}$ (or more values), obtain $\Pr(\theta_A < \theta_B | y_A, y_B)$ via Monte Carlo sampling. Plot this probability as a function of $(\kappa_0 = \nu_0)$. Describe how you might use this plot to convey the evidence that $\theta_A < \theta_B$ to people of a variety of prior opinions.

Solution

Step 1: Setting up the data and prior parameters

This code will provide the posterior parameters for both methods A and B for each pair of (κ_0, ν_0) values. We'll use these parameters and calculate the probabilities in question.

```
# Data and prior definitions
data_A <- c(y_bar_A = 75.2, s_A = 7.3, n_A = 16)
data_B <- c(y_bar_B = 77.5, s_B = 8.1, n_B = 16)
mu_0 <- 75; sigma2_0 <- 100
kappa_nu_values <- c(1, 2, 4, 8, 16, 32)

# Function to compute posterior parameters
compute_posterior <- function(
    data, kappa_0, nu_0, mu_0, sigma2_0) {
  n <- length(data)
  y_bar <- mean(data)
  s2 <- var(data)

  kappa_n <- kappa_0 + n
  nu_n <- nu_0 + n
```

[37] [DHo09] pg.237 Exercise 5.2

```r
  mu_n <- (kappa_0 * mu_0 + n * y_bar) / kappa_n
  sigma2_n <- 1/nu_n *
    (nu_0*sigma2_0 + (n-1)*s2 +
       kappa_0*n*(y_bar - mu_0)^2 / (kappa_0 + n))

  return(
    list(mu_n = mu_n,
         sigma2_n = sigma2_n,
         kappa_n = kappa_n, nu_n = nu_n))}

# Compute posterior parameters for each value
#of kappa_0 and nu_0
posterior_params_A <- list()
posterior_params_B <- list()

for (value in kappa_nu_values) {
  posterior_params_A[[as.character(value)]] <-
    compute_posterior(data_A, value, value, mu_0, sigma2_0)
  posterior_params_B[[as.character(value)]] <-
    compute_posterior(data_B, value, value, mu_0, sigma2_0)}

# Monte Carlo sampling to compute the
# Pr(theta_A < theta_B | y_A, y_B)
results <- numeric(length(kappa_nu_values))
names(results) <- as.character(kappa_nu_values)

for (value in kappa_nu_values) {
  posterior_A <- posterior_params_A[[as.character(value)]]
  posterior_B <- posterior_params_B[[as.character(value)]]

  s2_samples_A <- 1 / rgamma(
    5000, posterior_A$nu_n / 2,
    posterior_A$nu_n * posterior_A$sigma2_n / 2)
  theta_samples_A <- rnorm(
    5000, posterior_A$mu_n,
    sqrt(s2_samples_A / posterior_A$kappa_n))
```

```
  s2_samples_B <- 1 / rgamma(
    5000, posterior_B$nu_n / 2,
    posterior_B$nu_n * posterior_B$sigma2_n / 2)
  theta_samples_B <- rnorm(
    5000, posterior_B$mu_n,
    sqrt(s2_samples_B / posterior_B$kappa_n))

  results[as.character(value)] <-
    mean(theta_samples_A < theta_samples_B)}
```

```
> results
#      1       2       4       8      16      32
#0.5190  0.5168  0.5204  0.5032  0.5012  0.5072
```

The probabilities in respect to $\kappa_0, \nu_0 (= \kappa_0)$ are: for $\kappa_0 = \nu_0 = 1 \rightarrow 0.5298$, $\kappa_0 = \nu_0 = 2 \rightarrow 0.5064$, $\kappa_0 = \nu_0 = 4 \rightarrow 0.4994$, $\kappa_0 = \nu_0 = 8 \rightarrow 0.4930$, $\kappa_0 = \nu_0 = 16 \rightarrow 0.5006$ and $\kappa_0 = \nu_0 = 16 \rightarrow 0.5102$

A way to convey the evidence that $\theta_A > \theta_B$ through the plot 6.44 is to say to someone who believes otherwise " if you choose to go against that claim you will be wrong $\frac{5}{6}$ of the times based on the graph" as a joke and then say ", but seriously now a trend for the plot to move upwards as the $\kappa_0 = \nu_0$ gets bigger is evident".

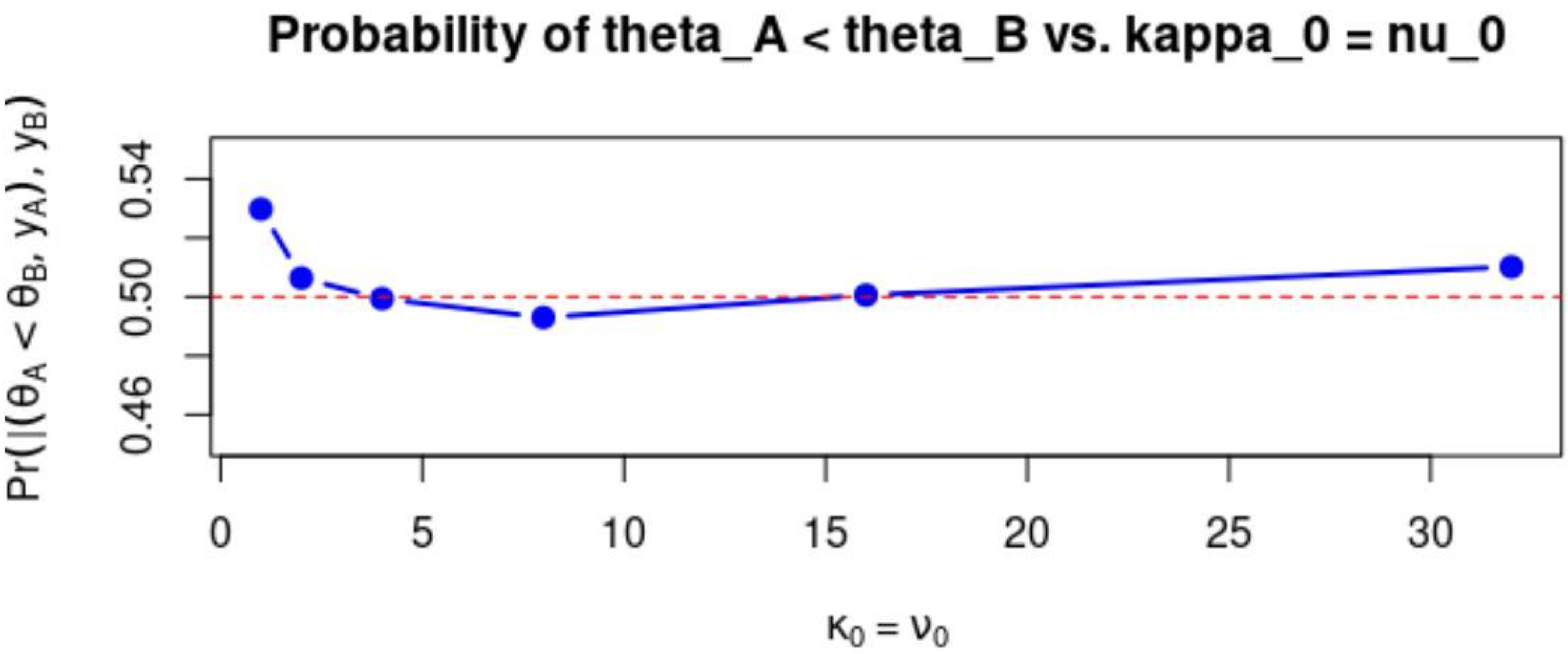


Figure 6.44: Our plot.

An insight from the chat" The plot suggests that while there's some evidence pointing towards $\theta_A < \theta_B$ for weaker priors, the evidence is not strong. The data doesn't provide a definitive answer about the superiority of one method over the other. Individuals should interpret the results in the context of their prior beliefs and other external information".

6.3 Marginal distributions

[38]

Given observations $Y_1, \ldots, Y_n \sim$ i.i.d. normal(θ, σ^2) and using the conjugate prior distribution for θ and σ^2, derive the formula for p$(\theta|y_1, \ldots, y_n)$, the marginal posterior distribution of θ, conditional on the data but marginal over σ^2. Check your work by comparing your formula to a Monte Carlo estimate of the marginal distribution, using some values of $Y_1, \ldots, Y_n, \mu_0, \sigma_0^2, \nu_0$ and κ_0, that you choose. Also derive p$(\tilde{\sigma}^2|y_1, \ldots, y_n)$, where $\tilde{\sigma}^2 = \frac{1}{\sigma^2}$ is the precision.

Solution

Well, the Analytical Solution would have taken a lot of time and effort to be presented analytically, so I 'll just outline the process that took me to that solution.

Spoiler Alert :

$$p(\theta|y_1, \ldots, y_n) = \frac{(\nu_n\sigma_n^2)^{\frac{\nu_n}{2}}}{\Gamma(\frac{\nu_n}{2})} \cdot \frac{2\sqrt{\frac{\kappa_n}{n}} \cdot \Gamma(\frac{\nu_n+1}{2})}{[(\kappa_n(\theta-\mu_n)^2 - \nu_n\sigma_n^2)]^{\frac{\nu_n+1}{2}}}$$

Because

$$p(\theta|y_1, \ldots, y_n) = \int p(\theta, \sigma^2|y_1, \ldots, y_n)\; d\sigma^2 = \int p(\theta|\sigma^2, y_1, \ldots, y_n) \times p(\sigma^2|y_1, \ldots, y_n)\; d\sigma^2$$

Then by replacing: $p(\theta|\sigma^2, y_1, \ldots, y_n)$ with the expression of the normal pdf with the corresponding parameters and $p(\sigma^2|y_1, \ldots, y_n)$ with the corresponding expression and parameters from the inverse-gamma we get:

$$p(\theta|y_1, \ldots, y_n) = \int \sqrt{\left(\frac{\kappa_n}{2\pi}\right)} \cdot C_{g_1} \times \left(\sigma^2\right)^{-\frac{\nu_n}{2}-1} \times \exp -\frac{1}{2\sigma^2} \cdot \left[\kappa_n\left(\theta-\mu_n\right)^2 + \nu_n\sigma_n^2\right]$$

where C_{g1} is the constant of the inv-gamma.
We then choose $\alpha = \frac{\nu_n+1}{2}$ and $\beta = \frac{1}{2}\left[\kappa_n\left(\theta-\mu_n\right)^2 + \nu_n\sigma_n^2\right]$. Supposing σ^2 follows an inv-gamma(α, β) with a pdf $p(\sigma^2|\alpha, \beta)$. We can produce the result I originally stated:

$$p(\theta|y_1, \ldots, y_n) = \sqrt{\left(\frac{\kappa_n}{2\pi}\right)} \cdot C_{g_1} \cdot (C_{g_2})^{-1} \times \int p(\sigma^2|\alpha, \beta)\; d\sigma^2$$

To derive $p(\sigma^2 \mid y_1, \ldots, y_n)$:

[38] [DHo09]pg.237 Exercise 5.3

$$
\begin{aligned}
p(\sigma^2 \mid y_1,\dots,y_n) &\propto p(\sigma^2)p(y_1,\dots,y_n \mid \sigma^2) \\
&= p(\sigma^2) \times \int p(y_1,\dots,y_n \mid \theta,\sigma^2)p(\theta \mid \sigma^2)\, d\theta \\
&= p(\sigma^2) \int \left[\prod_{i=1}^{n} \mathrm{dnorm}(y_i,\theta,\sigma^2)\right] \times \mathrm{dnorm}(\theta,\mu_0,\sigma^2/\kappa_0)\, d\theta \\
&= p(\sigma^2) \int \sqrt{\frac{1}{2\pi\sigma^2}}^{\,n} e^{-\frac{1}{2\sigma^2}\sum(y_i-\theta)^2} \times \sqrt{\frac{\kappa_0}{2\pi\sigma^2}} e^{-\frac{\kappa_0}{2\sigma^2}(\theta-\mu_0)^2}\, d\theta \\
&= \sqrt{\kappa_0}\,p(\sigma^2) \int \sqrt{\frac{1}{2\pi\sigma^2}}^{\,n+1} \times e^{-\frac{1}{2\sigma^2}\left[\kappa_0(\theta-\mu_0)^2+\sum(y_i-\theta)^2\right]}\, d\theta
\end{aligned}
$$

$$
= \sqrt{\kappa_0} \int \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma(\frac{\nu_0}{2})} \left(\sigma^2\right)^{-\frac{\nu_0}{2}-1} e^{-\frac{\frac{\nu_0\sigma_0^2}{2}}{\sigma^2}} \times \sqrt{\frac{1}{2\pi\sigma^2}}^{\,n+1} \times e^{-\frac{1}{2\sigma^2}\left[\kappa_0(\theta-\mu_0)^2+\sum(y_i-\theta)^2\right]}\, d\theta
$$

$$
= \sqrt{\kappa_0}\left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma(\frac{\nu_0}{2})} \int \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} \times e^{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2+\kappa_0(\theta-\mu_0)^2+\sum(y_i-\theta)^2\right]}\, d\theta
$$

$$
\begin{aligned}
&\propto \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} \int e^{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2+\kappa_0(\theta-\mu_0)^2+\sum(y_i-\theta)^2\right]}\, d\theta \\
&= \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} \int e^{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2+\kappa_0(\theta-\mu_0)^2+n(\bar{y}-\theta)^2+\sum(y_i-\bar{y})^2\right]}\, d\theta
\end{aligned}
$$

$$
= \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} e^{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2+\sum(y_i-\bar{y})^2\right]} \int e^{-\frac{1}{2\sigma^2}\left[(\kappa_0+n)\theta^2-2(\kappa_0\mu_0+n\bar{y})\theta+\kappa_0\mu_0^2+n\bar{y}^2\right]}\, d\theta
$$

$$
= \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} e^{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2+\sum(y_i-\bar{y})^2\right]} \int e^{-\frac{\kappa_0+n}{2\sigma^2}\left[\theta^2-2\frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\theta+\left(\frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\right)^2-\left(\frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\right)^2+\frac{\kappa_0\mu_0^2+n\bar{y}^2}{\kappa_0+n}\right]}\, d\theta
$$

$$
= \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} e^{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2+\sum(y_i-\bar{y})^2-\frac{(\kappa_0\mu_0+n\bar{y})^2}{\kappa_0+n}+\kappa_0\mu_0^2+n\bar{y}^2\right]} \int e^{-\frac{\kappa_0+n}{2\sigma^2}\left(\theta-\frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\right)^2}\, d\theta
$$

$$
= \left(\sigma^2\right)^{-\frac{\nu_0+n+1}{2}-1} e^{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2+\sum(y_i-\bar{y})^2-\frac{(\kappa_0\mu_0+n\bar{y})^2}{\kappa_0+n}+\kappa_0\mu_0^2+n\bar{y}^2\right]} \left(\frac{1}{\sqrt{2\pi\frac{\sigma^2}{\kappa_0+n}}}\right)^{-1} \times
$$

$$
\begin{aligned}
&\int \mathrm{dnorm}(\theta, \frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}, \frac{\sigma^2}{\kappa_0+n})\, d\theta \\
&\propto \left(\sigma^2\right)^{-\frac{\nu_0+n}{2}-1} e^{-\frac{1}{2\sigma^2}\left[\nu_0\sigma_0^2+\sum(y_i-\bar{y})^2-\frac{(\kappa_0\mu_0+n\bar{y})^2}{\kappa_0+n}+\kappa_0\mu_0^2+n\bar{y}^2\right]}
\end{aligned}
$$

$$
\propto \text{inv-gamma}\left(\sigma^2, \frac{\nu_0+n}{2}, \frac{\nu_0\sigma_0^2+\sum(y_i-\bar{y})^2-\frac{(\kappa_0\mu_0+n\bar{y})^2}{\kappa_0+n}+\kappa_0\mu_0^2+n\bar{y}^2}{2}\right)
$$

thus,

$$
\sigma^2 \sim \text{inv-gamma}(\frac{\nu_n}{2}, \frac{\nu_n\sigma_n^2}{2})
$$

6.4 Jeffreys' prior

[39]

For sampling models expressed in terms of a p-dimensional vector ψ, Jeffreys' prior (Exercise 3.11) is defined as $p_J(\psi) \propto \sqrt{|I(\psi)|}$, where $|I(\psi)|$ is the determinant of the p x p matrix $I(\psi)$ having entries $I(\psi)_{k,l} = -E\left[\partial^2 \log p(Y|\psi)/\partial\psi_k \partial\psi_l\right]$.
a) Show that Jeffreys' prior for the normal model is $p_J(\theta, \sigma^2) \propto (\sigma^2)^{-\frac{3}{2}}$.
b) Let $y = (y_1, \ldots, y_n)$ be the observed values of an i.i.d. sample from a normal(θ, σ^2) population. Find a probability density $pJ(\theta, \sigma^2|y)$ such that $p_J(\theta, \sigma^2|y) \propto p_J(\theta, \sigma^2) \cdot p(y|\theta, \sigma^2)$. It may be convenient to write this joint density as $p_J(\theta|\sigma^2, y) \times p_J(\sigma^2|y)$. Can this joint density be considered a posterior density?

Solution

a) Given a sampling model expressed in terms of a parameter vector $\boldsymbol{\psi}$, Jeffreys' prior is defined as $p_J(\boldsymbol{\psi}) \propto \sqrt{|\boldsymbol{I}(\boldsymbol{\psi})|}$, where $|\boldsymbol{I}(\boldsymbol{\psi})|$ is the determinant of the Fisher information matrix $\boldsymbol{I}(\boldsymbol{\psi})$.
We consider a normal model with parameters θ (mean) and σ^2 (variance). The log-likelihood function for a single observation Y is given by

$$\log p(Y|\theta, \sigma^2) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(Y-\theta)^2.$$

The Fisher information matrix elements are the expected negative second-order partial derivatives of the log-likelihood:

$$\begin{aligned} I(\theta, \sigma^2)_{\theta,\theta} &= E\left[\frac{\partial^2}{\partial\theta^2}\log p(Y|\theta, \sigma^2)\right], \\ I(\theta, \sigma^2)_{\sigma^2,\sigma^2} &= E\left[\frac{\partial^2}{\partial(\sigma^2)^2}\log p(Y|\theta, \sigma^2)\right], \\ I(\theta, \sigma^2)_{\theta,\sigma^2} &= E\left[\frac{\partial^2}{\partial\theta\partial\sigma^2}\log p(Y|\theta, \sigma^2)\right]. \end{aligned}$$

After calculating these derivatives, we find

$$\boldsymbol{I}(\theta, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}.$$

The determinant of $\boldsymbol{I}(\theta, \sigma^2)$ is

$$|\boldsymbol{I}(\theta, \sigma^2)| = \frac{1}{\sigma^2} \cdot \frac{1}{2\sigma^4} - 0 = \frac{1}{2\sigma^6}.$$

Therefore, Jeffreys' prior for the normal model is

[39] [DHo09]pg. 238 Exercise 5.4

$$p_J(\theta, \sigma^2) \propto \sqrt{|\boldsymbol{I}(\theta, \sigma^2)|} \propto \frac{1}{\sigma^3} \propto \left(\sigma^2\right)^{-\frac{3}{2}}.$$

b) Given observations $y = (y_1, \ldots, y_n)$ from an i.i.d. sample from a normal(θ, σ^2) population, we want to find a probability density $p_J(\theta, \sigma^2|y)$ such that $p_J(\theta, \sigma^2|y) \propto p_J(\theta, \sigma^2) \cdot p(y|\theta, \sigma^2)$. It may be convenient to write this joint density as $p_J(\theta|\sigma^2, y) \times p_J(\sigma^2|y)$. We will also discuss whether this joint density can be considered a posterior density.

First, we establish the likelihood and the prior:

1. *Likelihood*: The likelihood function is given by

$$p(y|\theta, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2\right).$$

2. *Prior*: From part (a), Jeffreys' prior for the normal model is

$$p_J(\theta, \sigma^2) \propto (\sigma^2)^{-\frac{3}{2}}.$$

Now, we find the joint posterior density $p_J(\theta, \sigma^2|y)$:

$$p_J(\theta, \sigma^2|y) \propto p(y|\theta, \sigma^2) \cdot p_J(\theta, \sigma^2).$$

Substituting in the likelihood and prior, we get

$$p_J(\theta, \sigma^2|y) \propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2\right) \cdot (\sigma^2)^{-\frac{3}{2}}.$$

This expression represents the joint posterior density for θ and σ^2 given the data, incorporating Jeffreys' prior [40].

This joint density can indeed be considered a posterior density, as it is derived from the product of the likelihood and the prior, following Bayes' theorem. However, it's important to note that Jeffreys' prior is a non-informative prior, designed to be invariant under parameter transformations, not to incorporate subjective prior information.

[40] We still have to express this joint density as $p_J(\theta|\sigma^2, y) \times p_J(\sigma^2|y)$

Task 7

Exercise 4.2

During the severe floods in the Midwest in the summer of 2008, the adjacent towns of Iowa City and Coralville in Johnson County, Iowa, were hit hard. Despite sustained efforts at sandbagging throughout the community, hundreds of homes, businesses, churches, and university buildings were damaged or destroyed. Major roads and bridges were closed for weeks. Less than a year later, with parts of both towns still recovering from the flood, a vote was held on a proposal to impose a local sales tax of one cent on the dollar to pay for flood-prevention and flood-mitigation projects. A few days before the actual vote, a local newspaper reported in its online edition, The Gazette Online (http://www.gazetteonline.com), on May 2, 2009:

> "The outcome of Tuesday's local-option sales tax election in Johnson County appears too close to call, based on results from a Gazette Communications poll of voters.
> The telephone survey of 327 registered voters in Johnson County, conducted April 27–29, shows 40% in favor of the 4-year 1% sales tax..."
> (Forty percent of 327 respondents is 131.)

A member of a local organization called "Ax the Tax" claims that this means that under half of all registered voters in the county support the local-option sales tax. She would like to use the sample survey data from the newspaper to test the two hypotheses:

$$H_0 : \pi \geq 0.5$$

$$H_a : \pi < 0.5$$

where π represents the proportion of all Johnson County registered voters who support the sales tax.

1. A frequentist method of testing these hypotheses is based on the p-value. The p-value is the probability of observing the sample result obtained, or something more extreme, if indeed exactly half of the registered voters in Johnson County supported the sales tax; that is,

$$\text{p-value} = \Pr(y \leq 131 | \pi = 0.5)$$

where y is a binomial random variable with sample size $n = 327$ and success probability $\pi = 0.5$. Compute the p-value for this example (the use of an R function will make this easy). If this probability is small, then one concludes that there is significant evidence in support of hypothesis $H_a : \pi < 0.5$.

2. Now, consider a Bayesian approach to testing these hypotheses. Suppose that a uniform prior is assigned to π. Find the posterior distribution of π and use it to compute the posterior probabilities of H_0 and H_1.

Solution):

Context: The exercise provides a background about the severe floods in the Midwest in 2008, specifically in the towns of Iowa City and Coralville. A proposal was made to impose a local sales tax to fund flood-prevention and mitigation projects. A telephone survey was conducted to gauge the support for this tax, and the results showed that 40% of the 327 respondents (which is 131 respondents) were in favor.

Claim: A member of a local organization interprets this result to mean that less than half of all registered voters in the county support the tax. She wants to statistically test this claim.

Hypotheses: The null hypothesis H_0 is that half or more of the registered voters support the tax $\pi \geq 0.5$, and the alternative hypothesis H_a is that less than half support it $\pi < 0.5$.

Tasks:

1. Frequentist Approach: Using a frequentist method, we are asked to compute the p-value for the observed result. The p-value represents the probability of observing 131 or fewer voters in favor, assuming that exactly half of the registered voters support the tax. If the p-value is small, it would suggest evidence against the null hypothesis in favor of the alternative.

To find the p-value for the given scenario using the frequentist approach we need to calculate the probability of observing 131 or fewer successes (voters in favor of the tax) in a sample of size 327, assuming the true proportion of successes π is 0.5.

This probability is represented by the cumulative distribution function (CDF) of the binomial distribution.

$$\text{p-value} = P(Y \leq 131 | \pi = 0.5)$$

where Y is a binomial random variable representing the number of successes in the sample, $n = 327$ is the sample size and $\pi = 0.5$ is the assumed proportion of successes under the null hypothesis.
To compute the p-value, you can use the binomial CDF. In R:

```
p_value <- pbinom(131, size = 327, prob = 0.5)
print(p_value)
```

With output: [1] 0.0001927988

The magnitude of the p-value suggests that the observed result is unlikely to have occurred by random chance alone, given the null hypothesis. Thus, we should reject the null hypothesis in favor of the alternative.
2. Bayesian Approach: Using a Bayesian method with a uniform prior for π, we are asked to find the posterior distribution of π and use it to compute the posterior probabilities of the two hypotheses.
Given the prior is uniform and the binomial likelihood with $n = 327$ and $y = 131$ successes, the posterior distribution for π is proportional to the product of the prior and the likelihood:

$$p(\pi|y) \propto p(y|\pi) \times p(\pi)$$

Given that the prior is uniform, the posterior is directly proportional to the likelihood. Thus, the shape of the posterior will be similar to the binomial likelihood centered around $\frac{y}{n}$.
To compute the posterior probabilities of H_0 and H_a:

$$P(H_0|y) = \int_{0.5}^{1} p(\pi|y)\, d\pi P(H_a|y) = \int_{0}^{0.5} p(\pi|y)\, d\pi$$

In practice, you can use a computational method, such as Markov Chain Monte Carlo (MCMC), to sample from the posterior distribution and then estimate these probabilities. Alternatively, given the conjugacy of the problem, you can derive the posterior distribution analytically and then compute the integrals.
In R, given the posterior is Beta$(y + 1, n - y + 1)$ =Beta$(132, 197)$.

```
# Parameters for the posterior distribution
alpha_post <- 131 + 1
beta_post <- 327 - 131 + 1

# Compute posterior probabilities
P_H0 <- pbeta(0.5, alpha_post, beta_post, lower.tail = FALSE)
P_Ha <- pbeta(0.5, alpha_post, beta_post, lower.tail = TRUE)

print(P_H0)
print(P_Ha)
```

With output: [1] 0.0001590998 and [1] 0.9998409, respectivelly

5.4

[41]

Suppose that a trucking company owns a large fleet of well-maintained trucks and assume that breakdowns appear to occur at random times. The president of the company is interested in learning about the daily rate λ at which breakdowns occur. (Realistically, each truck would have a breakdown rate that depends possibly on its type, age, condition, driver, usage, etc. The breakdown rate for the whole company can be viewed as the sum of the breakdown rates of the individual trucks.) For a given value of the rate parameter λ, it is known that the number of breakdowns y on a particular day has a Poisson distribution with mean λ:

$$p(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots$$

1. Suppose that one observes the number of truck breakdowns for n consecutive days—denote these numbers by $y_1, \ldots, y_n$. If one assumes that these are exchangeable measurements (conditionally independent given λ), find the joint probability distribution of $y_1, \ldots, y_n$.

2. The numbers of breakdowns for 5 days are recorded to be 2, 5, 1, 0, and 3. Find the likelihood function $L(\lambda)$ of the rate parameter λ for these observations. Graph this function. (You may either use R or do it "by hand" by calculating the likelihood for the values $R = 0.1, 0.5, 1, 2, 4, 8$, and 16 and connecting the points with a smooth curve.)

3. Use calculus to find the mle of λ. Then use the `poisson.test` function in R to confirm the mle and to obtain a 95% frequentist confidence interval for λ.

Solution):

Context: The exercise sets up a context where a trucking company experiences breakdowns of its trucks. These breakdowns are assumed to occur at random times, and the number of breakdowns on any given day follows a Poisson distribution with a mean rate of λ.

Tasks:

1. Joint Probability Distribution: The first part asks for the joint probability distribution of the number of breakdowns observed over n consecutive days, given

[41] Trucking Company Breakdown Analysis

the breakdowns are exchangeable measurements (conditionally independent given λ).

Given that the number of breakdowns y on a particular day follows a Poisson distribution with mean λ, the probability mass function (PMF) for y is:

$$p(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$$

Now, if we assume that the observations $y_1, \ldots, y_n$ are exchangeable measurements (conditionally independent given λ), the joint probability distribution of these observations is simply the product of their individual probabilities:

$$p(y_1, \ldots, y_n|\lambda) = \prod_{i=1}^{n} p(y_i|\lambda) = e^{-n\lambda}\lambda^{\sum y_i}\prod_{i=1}^{n}\frac{1}{y_i!}$$

2. Likelihood Function: The second part provides specific data on the number of breakdowns over 5 days. Using this data, you're asked to compute the likelihood function for the rate parameter λ. Additionally, you're asked to graph this function for specific values of λ.

Given the observations $y_1 = 2, y_2 = 5, y_3 = 1, y_4 = 0$ and $y_5 = 3$, we want to find the likelihood function $L(\lambda) = p(y_1, \ldots, y_5|\lambda)$ of the rate parameter λ. Using the formula derived in the first part:

$$L(\lambda) = e^{-5\lambda}\lambda^{11}\frac{1}{1440}$$

To graph this function, you can plot $L(\lambda)$ against λ for the given values $R = 0.1, 0.5, 1, 2, 4, 8$ and 16. The shape of the graph will give you an idea of which values of λ are most supported by the data. Using the below code we plotted 7.45

```
# Define the likelihood function
likelihood <- function(lambda) {
  return(exp(-5*lambda) * lambda^11 / (factorial(2) * factorial(5) * fac
}

# Values of R for which we want to compute the likelihood
lambda_values <- c(0.1, 0.5, 1, 2, 4, 8, 16)

# Compute the likelihood for each value of R
likelihood_values <- sapply(lambda_values, likelihood)
```

```
# Plot the likelihood function
plot(lambda_values, likelihood_values, type="b", log="y",
     xlab=expression(lambda), ylab="Likelihood",
     main="Likelihood Function for Lambda")
```

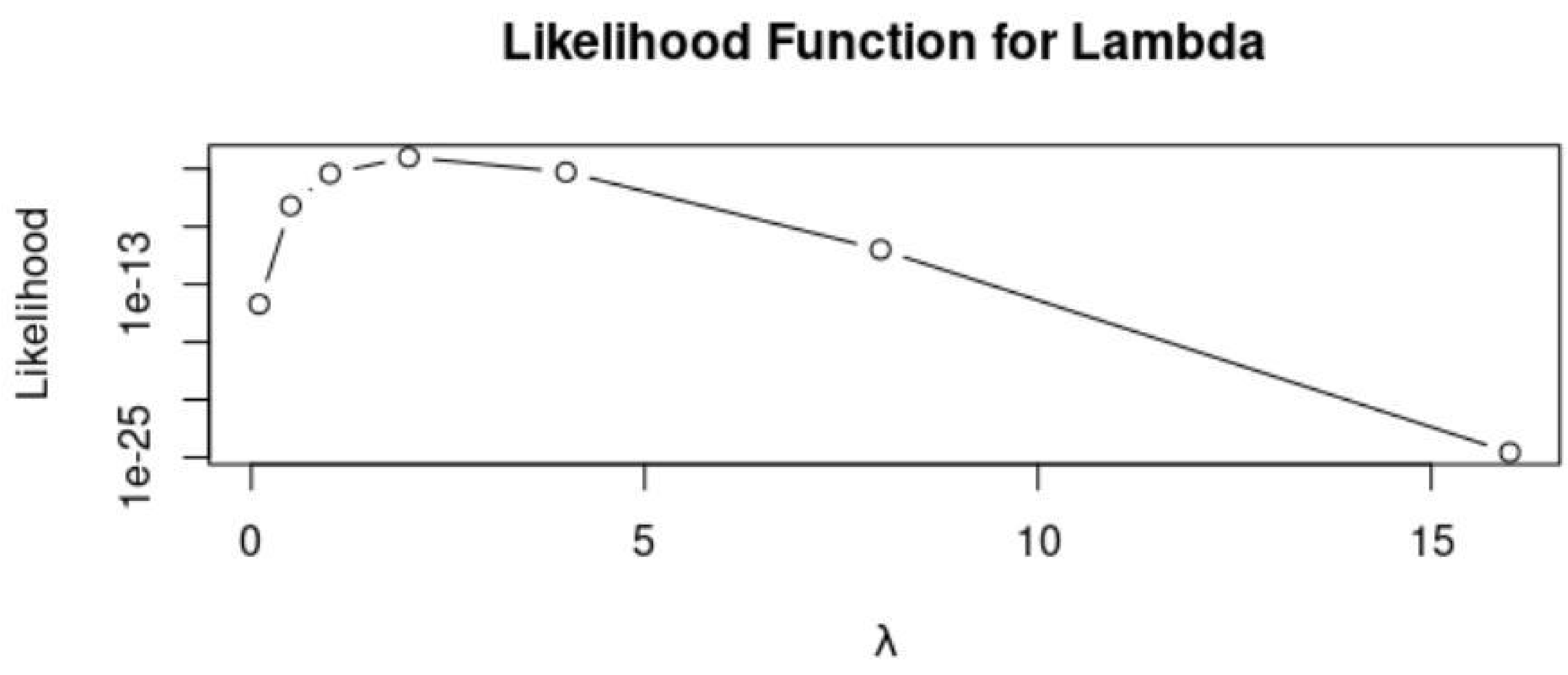


Figure 7.45: The plot for **a)**.

3. Maximum Likelihood Estimation (MLE): The third part asks us to determine the maximum likelihood estimate (MLE) of λ using calculus. After finding the MLE analytically, we're also asked to confirm this value using the poisson.test function in R and to compute a 95% confidence interval for λ.
The Maximum Likelihood Estimation (MLE) aims to find the value of λ that maximizes the likelihood function. To find the MLE of λ, we need to differentiate the likelihood function with respect to λ and set it to zero. This will give us the value of λ that maximizes the likelihood. Given the likelihood:

$$L(\lambda) = e^{-5\lambda}\lambda^{11}$$

we take the natural logarithm and differentiate with respect to λ to get the MLE:

$$\lambda = 2.2$$

To confirm this and obtain a 95% frequentist confidence interval for λ, you can use the poisson.test function in R.

```
# Given data
breakdowns <- c(2, 5, 1, 0, 3)

# Use the poisson test to confirm the MLE and obtain a 95% CI
```

```
test_result <- poisson.test(sum(breakdowns), T = 5)

# Print the results
print(test_result$estimate)
print(test_result$conf.int)
```

With output: $\lambda_{MLE} = 2.2$ and confidence interval[1] 1.0982323.936408

Exercise 5.5

The president of the company has some knowledge about the location of the Poisson rate parameter λ based on the observed number of breakdowns from previous years. His prior beliefs about λ are represented in the following:

$$p(\lambda) \propto \lambda^3 \exp(-2\lambda), \lambda > 0$$

1. Is this prior a member of a particular parametric family? If so, what family and what are the prior parameters?

2. Plot this prior density, either using software or by picking a few values at which to evaluate it as in the previous problem. Based on the plot, describe the president's prior beliefs about the rate parameter λ.

3. Write out the mathematical form of the unnormalized posterior density. Identify its parametric family and parameters.

4. Find the posterior mean and 95% central credible set for λ based on this posterior.

5. Was the president's prior from a conjugate family for the Poisson likelihood? How could you tell?

Solution:
Context: The exercise provides a prior distribution for the Poisson rate parameter λ, which is given by $p(\lambda) \propto \lambda^3 e^{-2\lambda}$ for $\lambda > 0$ and asks us to perform various tasks.
Tasks:
1. Determine family: Determine if the given prior belongs to a known parametric family of distributions. If so, identify the family and its parameters.
This form is reminiscent of the gamma distribution and comparing the given prior with the gamma distribution pdf, we can identify that it is in fact proportional to Gamma($\alpha = 4, \beta = 2$). Thus, the prior distribution is that gamma distribution.
2. Plot the prior & more: Plot the prior distribution and describe the president's beliefs about the rate parameter λ based on this plot.
To plot the prior density7.46, we'll use the gamma distribution with the parameters we just showed in R:

```
# Load necessary libraries
```

```
library(ggplot2)

# Define the parameters for the gamma distribution
alpha <- 4
beta <- 2

# Generate a sequence of lambda values for plotting
lambda_values <- seq(0, 10, 0.1)

# Compute the density values for each lambda
density_values <- dgamma(lambda_values, shape=alpha, rate=beta)

# Plot the prior density
ggplot(data.frame(lambda=lambda_values, density=density_values), aes(x=l
  geom_line() +
  theme_minimal() +
  labs(title="Prior Density for Lambda", x="Lambda", y="Density")
```

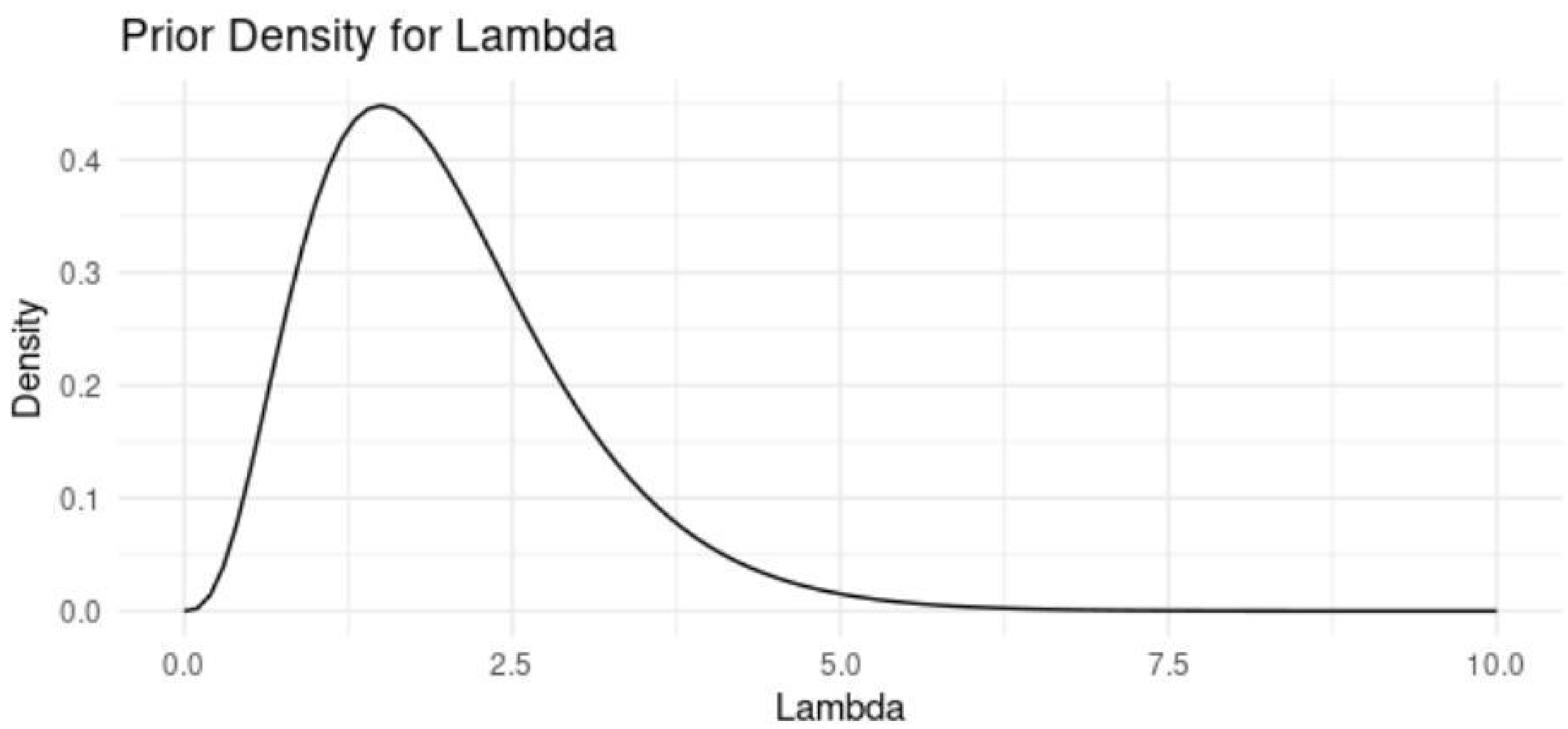


Figure 7.46: The plot for **a)**.

Description based on the plot the president's prior beliefs about the rate parameter λ are centered around the mode of the gamma distribution. The distribution's shape suggests that the president believes that values of λ around this mode are most probable. The spread of the distribution indicates the degree of uncertainty in these beliefs and as the distribution is narrow, it suggests a strong belief in that specific range of λ values, but all are relative.

3. Derive the unnormalized posterior & more: Derive the unnormalized

posterior density for λ given the provided prior and the likelihood from a Poisson distribution. Identify the parametric family and parameters of this posterior.
Given that the conjugacy of our prior when used upon the Poisson model our posterior is a Gamma$(4 + \sum y_i, 3 + n)$.
4. mean and a 95% credible interval: Calculate the posterior mean and a 95% central credible interval for λ.
Given our posterior, the central credible intervals set is an interval that contains 95% of the posterior distribution and can be found using the quantile function. Specifically, we want to find values $\lambda_{0.025}$ and $\lambda_{0.975}$ s.t.

$$P(\lambda < \lambda_{0.025}) = 0.025 P(\lambda < \lambda_{0.975}) = 0.975$$

In R:

```r
# Given values for Y and n
# (You'll need to replace these with the actual values from your data)
Y <- ...  # Total number of breakdowns over n days
n <- ...  # Number of days

# Posterior parameters
alpha_post <- Y + 7
beta_post <- n + 2

# Posterior mean
posterior_mean <- alpha_post / beta_post

# 95% central credible set
lambda_lower <- qgamma(0.025, shape=alpha_post, rate=beta_post)
lambda_upper <- qgamma(0.975, shape=alpha_post, rate=beta_post)

print(posterior_mean)
print(c(lambda_lower, lambda_upper))
```

5. Determine conjugacy: Determine if the given prior is conjugate to the Poisson likelihood and explain the reasoning.
Conjugacy in Bayesian statistics refers to the situation where the posterior distribution belongs to the same family as the prior distribution when combined with a

particular likelihood function. In this context, we're dealing with a Poisson likelihood. The conjugate prior for the Poisson likelihood is the Gamma distribution.

7.1 Unit information prior

[42]

Obtain a unit information prior for the normal model as follows:

a) Reparameterize the normal model as $p(y|\theta, \psi)$, where $\psi = \frac{1}{\sigma^2}$. Write out the log likelihood $l(\theta, \psi|\boldsymbol{y}) = \sum \log p(y_i|\theta, \psi)$ in terms of θ and ψ.

b) Find a probability density $p_U(\theta, \psi)$ so that log $p_U(\theta, \psi) = \frac{l(\theta,\psi|\boldsymbol{y})}{n} + c$, where c is a constant that does not depend on θ or ψ. Hint: Write $\sum(y_i - \theta)^2$ as $\sum (y_i - \bar{y} + \bar{y} - \theta)^2 = \sum (yi - \bar{y})^2 + n(\theta - \bar{y})^2$, and recall that $p_U(\theta, \psi) = \log p_U(\theta|\psi) + \log p_U(\psi)$.

c) Find a probability density $p_U(\theta, \psi|\boldsymbol{y})$ that is proportional to $p_U(\theta, \psi) \times p(y_1, \ldots, y_n|\theta, \psi)$. It may be convenient to write this joint density as $p_U(\theta|\psi, y) \times p_U(\psi|y)$. Can this joint density be considered a posterior density?

Solution

a) The normal model density function:

$$p(y_i|\theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \times e^{-\frac{1}{2\sigma^2}(y_i-\theta)^2}$$

So the reparametrization $\psi = \frac{1}{\sigma^2}$ is derived as:

$$p(y_i|\theta, \psi) = \sqrt{\frac{\psi}{2\pi}} \times e^{-\frac{\psi}{2}(y_i-\theta)^2}$$

Thus the answer to **a)** is:

$$\begin{aligned} l(\theta, \psi|\boldsymbol{y}) &= \sum \log \sqrt{\frac{\psi}{2\pi}} \times e^{-\frac{\psi}{2}(y_i-\theta)^2} \\ &= \sum \left(\log \left(\frac{\psi}{2\pi}\right)^{\frac{1}{2}} - \frac{\psi}{2}(y_i - \theta)^2 \right) \\ &= \log \left(\frac{\psi}{2\pi}\right)^{\frac{n}{2}} - \frac{\psi}{2} \sum (y_i - \theta)^2 \end{aligned}$$

b) Given the relationship:

$$\begin{aligned} \log p_U(\theta, \psi) = \frac{l(\theta, \psi|\boldsymbol{y})}{n} + c &= \log \left(\frac{\psi}{2\pi}\right)^{\frac{1}{2}} - \frac{\psi}{2n} \sum (y_i - \theta)^2 + c \\ &= \log \left(\frac{\psi}{2\pi}\right)^{\frac{1}{2}} - \frac{\psi}{2n} \left[n(\theta - \bar{y})^2 + \sum (y_i - \bar{y})^2 \right] + c \\ &= \log \left[\left(\frac{\psi}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\psi}{2}(\theta - \bar{y})} \right] + \log \left[e^c e^{-\frac{\psi}{2n}\sum(y_i - \bar{y})^2} \right] \end{aligned}$$

[42][DHo09]pg.238 Exercise 5.5

Suppose e^c s.t $e^c = \frac{\sum(y_i-\bar{y})^2}{2n}$, then:

$$\begin{aligned}
\log p_U(\theta,\psi) &= \log\left[\left(\frac{\psi}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\psi}{2}(\theta-\bar{y})}\right] + \log\left[\frac{(y_i-\bar{y})^2}{2n} e^{-\frac{\psi}{2n}\sum(y_i-\bar{y})^2}\right] \\
&= \log\left[\text{dnormal}(\theta,\bar{y},\frac{1}{\psi})\right] + \log\left[\text{dgamma}(\psi,1,\frac{\sum(y_i-\bar{y})^2}{2n})\right] \\
&= \log\left[p_U(\theta|\psi)\right] + \log\left[p_U(\psi)\right]
\end{aligned}$$

thus,

$$p_U(\theta,\psi) = \left(\frac{\psi}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\psi}{2}(\theta-\bar{y})^2} \times \frac{(y_i-\bar{y})^2}{2n} e^{-\frac{\psi}{2n}\sum(y_i-\bar{y})^2}$$

c) To find a probability density $p_U(\theta,\psi|\boldsymbol{y})$ proportional to $p_U(\theta,\psi)\times p(y_1,\ldots,y_n|\theta,\psi)$:

$$\begin{aligned}
p_U(\theta,\psi|\boldsymbol{y}) &\propto p_U(\theta,\psi)\times p(y_1\ldots,y_n|\theta,\psi) \\
&\propto \left(\frac{\psi}{2\pi}\right)^{\frac{n+1}{2}} e^{-\frac{\psi}{2}\left[(\bar{y}-\theta)^2+\frac{1}{n}\sum(y_i-\bar{y})^2\right]} \times e^{-\frac{\psi}{2}\left[n(\bar{y}-\theta)^2+\sum(y_i-\bar{y})^2\right]} \\
&\propto p_U(\theta|\psi))\times p_U(\psi)\times\sqrt{\frac{n}{2\pi}}\psi^{\frac{1}{2}}e^{-\frac{\psi n}{2}(\theta-\bar{y})^2}\times\psi^{\frac{n-3}{2}-1}e^{-\frac{\psi}{2}\sum(y_i-\bar{y})^2} \\
&\propto p_U(\theta|\psi))\times p_U(\psi)\times\text{dnormal}\left(\theta,\bar{y},\frac{1}{n\psi}\right)\times\text{dgamma}\left(\psi,\frac{n-3}{2},\frac{\sum(y_i-\bar{y})^2}{2}\right)
\end{aligned}$$

Suppose I could have the following two things:

$$\begin{aligned}
&\textbf{1.}\{\boldsymbol{y},\psi|\theta\}\sim\text{normal}\left(\theta,\bar{y},\frac{1}{n\psi}\right) \\
&\textbf{2.}\{\boldsymbol{y}|\psi\}\sim\text{gamma}\left(\psi,\frac{n-3}{2},\frac{\sum(y_i-\bar{y})^2}{2}\right)
\end{aligned}$$

then,

$$\begin{aligned}
p_U(\theta,\psi|\boldsymbol{y}) &\propto p_U(\theta|\psi))\times p(\boldsymbol{y},\psi|\theta)\cdot p_U(\psi)\times p(\boldsymbol{y}|\psi) \\
&\propto p_U(\theta|\psi,\boldsymbol{y})\times p_U(\psi|\boldsymbol{y})
\end{aligned}$$

To conclude I believe this joint density can be considered as a posterior density as it's proportional to the joint of θ,ψ times the likelihood derived from our model. My only hang-up could be that based on some calculations I viewed that the $p_U(\theta,\psi|\boldsymbol{y})$ could be written as a function of $p_U(\theta,\psi)$

$$\begin{aligned}
p_U(\theta,\psi|\boldsymbol{y}) &= p_U(\theta,\psi)\times\sqrt{\frac{\psi}{2\pi}}^{\,n} e^{-\frac{\psi}{2}\sum(y_i-\theta)^2} \\
&= \sqrt{\frac{\psi}{2\pi}}^{\,n+1} e^{-[n+1]\frac{\psi}{2}\sum(y_i-\theta)^2} = p_U(\theta,\psi)^{n+1}
\end{aligned}$$

7.2 Poisson Population Comparisons

[43]

Reconsidering the number of children data from Exercise 4.8, we assume Poisson sampling models for the two groups, with a new parameterization: $\theta_A = \theta$ and $\theta_B = \theta \times \gamma$. Here, γ represents the relative rate $\frac{\theta_B}{\theta_A}$. The priors are $\theta \sim \text{gamma}(a_\theta, b_\theta)$ and $\gamma \sim \text{gamma}(a_\gamma, b_\gamma)$.**Let θ and γ be independent.**

(a) Are θ_A and θ_B independent or dependent under this prior distribution? In what situations is such a joint prior distribution justified?

(b) Obtain the form of the full conditional distribution of θ given y_A, y_B, and γ.

(c) Obtain the form of the full conditional distribution of γ given y_A, y_B, and θ.

(d) Set $a_\theta = 2$ and $b_\theta = 1$. Let $a_\gamma = b_\gamma \in \{8, 16, 32, 64, 128\}$. For each of these five values, run a Gibbs sampler for at least 5,000 iterations and obtain $E[\theta_B - \theta_A | y_A, y_B]$. Describe the effects of the prior distribution for γ on the results.

Solution

a) θ_A and θ_B are dependent under this prior distribution because θ_B is directly defined as a function of θ_A ($\theta_B = \theta_A \times \gamma$). As:

$$p(\theta_A, \theta_B) = p(\theta_B|\theta_A) \times p(\theta_A) = \frac{b_\gamma^{a_\gamma}}{\Gamma(a_\gamma)}(\theta\gamma)^{a_\gamma - 1}e^{-b_\gamma(\theta\gamma)} \times p(\theta) \neq p(\theta_B) \times p(\theta_A)$$

as θ, is independent from γ:

$$p(\theta_B) \times p(\theta_A) = p(\theta)^2 p(\gamma)$$

b) Given that $\theta_A = \theta$ and $\theta_B = \theta \times \gamma$, and the data $\boldsymbol{y_A}$ and $\boldsymbol{y_B}$ are assumed to follow Poisson distributions with means θ_A and θ_B respectively, the likelihood functions for $\boldsymbol{y_A}$ and $\boldsymbol{y_B}$ are:

$$p(\boldsymbol{y_A}|\theta, \gamma) = p(\boldsymbol{y_A}|\theta) = \prod_{i=1}^{n_A} \frac{e^{-\theta}\theta^{y_{A_i}}}{y_{A_i}!}$$

and

[43] [DHo09]pg.238 Exercise 6.1

$$p(\boldsymbol{y_B}|\theta,\gamma) = \prod_{j=1}^{n_B} \frac{e^{-\theta\gamma}(\theta\gamma)^{y_{B_j}}}{y_{B_j}!}$$

The joint likelihood is the product of the individual likelihoods because y_A and y_B are independent:

$$p(\boldsymbol{y_A},\boldsymbol{y_B}|\theta,\gamma) = \prod_{i=1}^{n_A} \frac{e^{-\theta}\theta^{y_{A_i}}}{y_{A_i}!} \cdot \prod_{j=1}^{n_B} \frac{e^{-\theta\gamma}(\theta\gamma)^{y_{B_j}}}{y_{B_j}!}$$

Given the priors $\theta \sim \text{gamma}(a_\theta, b_\theta)$ and $\gamma \sim \text{gamma}(a_\gamma, b_\gamma)$, and the fact that θ and γ are independent, the joint prior distribution is the product of the individual priors:

$$p(\theta,\gamma) = \text{gamma}(\theta; a_\theta, b_\theta) \cdot \text{gamma}(\gamma; a_\gamma, b_\gamma)$$

The posterior distribution is:

$$\begin{aligned}
p(\theta,\gamma \mid \boldsymbol{y}_A,\boldsymbol{y}_B) &\propto p(\theta,\gamma) \times p(\boldsymbol{y}_A,\boldsymbol{y}_B \mid \theta,\gamma) \\
&= p(\theta) \times p(\gamma) \times p(\boldsymbol{y}_A \mid \theta,\gamma) \times p(\boldsymbol{y}_B \mid \theta,\gamma) \\
&\propto \left(\theta^{a_\theta-1}e^{-b_\theta\theta}\right)\left(\gamma^{a_\gamma-1}e^{-b_\gamma\gamma}\right)\left(\prod_{i=1}^{n_A}\theta^{y_{A_i}}e^{-\theta}\right)\left(\prod_{i=1}^{n_B}(\gamma\theta)^{y_{B_i}}e^{-\gamma\theta}\right) \\
&= \left(\theta^{a_\theta-1}e^{-b_\theta\theta}\right)\left(\gamma^{a_\gamma-1}e^{-b_\gamma\gamma}\right)\left(\theta^{\sum_{i=1}^{n_A}y_{A_i}}e^{-n_A\theta}\right)\left((\gamma\theta)^{\sum_{i=1}^{n_B}y_{B_i}}e^{-n_B\gamma\theta}\right) \\
&= \left(\theta^{a_\theta-1}e^{-b_\theta\theta}\right)\left(\gamma^{a_\gamma-1}e^{-b_\gamma\gamma}\right)\left(\theta^{n_A\bar{y}_A}e^{-n_A\theta}\right)\left((\gamma\theta)^{n_B\bar{y}_B}e^{-n_B\gamma\theta}\right)
\end{aligned}$$

The full conditional distribution of θ given y_A, y_B, and γ is proportional to the integral of the posterior over γ:

$$\begin{aligned}
p(\theta|\boldsymbol{y_A},\boldsymbol{y_B},\gamma) &\propto \int p(\theta,\gamma \mid \boldsymbol{y}_A,\boldsymbol{y}_B)\, d\gamma \\
&= p(\theta) \times p(\boldsymbol{y_A}|\theta) \times \theta^{n_B\bar{y}_B} \times \int \gamma^{n_B\bar{y}_A+a_\gamma-1}e^{-\gamma(n_B\theta+b_\gamma)}\, d\gamma \\
&\propto p(\theta) \times p(\boldsymbol{y_A}|\theta) \times \theta^{n_B\bar{y}_B} \times \int \text{dgamma}(\gamma, n_B\bar{y}_A + a_\gamma, n_B\theta + b_\gamma)\, d\gamma \\
&\propto \prod_{i=1}^{n_A} \frac{e^{-\theta}\theta^{y_{A_i}}}{y_{A_i}!}\theta^{n_B\bar{y}_B}\theta^{\alpha_\theta-1}e^{-\theta\beta_\theta} \\
&\propto \theta^{n_A\bar{y}_A+n_B\bar{y}_B+\alpha_\theta-1}e^{-\theta(n_A\bar{y}_A+\beta_\theta)} \\
&\propto \text{dgamma}(\theta, n_A\bar{y}_A + n_B\bar{y}_B + \alpha_\theta, n_A\bar{y}_A + \beta_\theta)
\end{aligned}$$

c) Given the model and the priors, the full conditional for γ is derived from the joint distribution of y_A, y_B, θ, and γ, focusing on the terms that involve γ. As we did for θ.
The full conditional distribution of γ is proportional to the integral of the posterior over θ:

$$\begin{aligned}
p(\gamma|y_A, y_B, \theta) &\propto \int p(\theta, \gamma \mid \boldsymbol{y}_A, \boldsymbol{y}_B)\, d\gamma \\
&= p(\gamma) \times \gamma^{n_B \bar{y}_B} \int \theta^{n_A \bar{y}_A + n_B \bar{y}_B + \alpha_\theta - 1} \times e^{-\theta(n_B \gamma + n_A + \beta_\theta)}\, d\theta \\
&\propto \gamma^{n_B \bar{y}_B + \alpha_\gamma - 1} e^{-\gamma \beta_\gamma} \times \int \text{dgamma}(\theta, n_A \bar{y}_A + n_B \bar{y}_B + \alpha_\theta, n_B \gamma + n_A + \beta_\theta)\, d\theta \\
&\propto \text{dgamma}(\gamma, n_B \bar{y}_B + \alpha_\gamma, \beta_\gamma)
\end{aligned}$$

d) In a Gamma distribution, the shape and rate parameters influence the distribution's mean and variance. Specifically, the mean of a Gamma distribution is a/b, and the variance is a/b^2.

In our study, if $a_\gamma = b_\gamma$, and both are greater than 1, the prior distribution for γ is centered around 1 and is more concentrated the larger these parameters are. This could be interpreted as a stronger belief that γ is close to 1, because values far from 1 are less likely under the prior.

Thus, it affects the posterior difference as our belief in $\gamma = 1$ strengthens, and the mean posterior difference between θ_B and θ_A decreases. This makes sense because γ is interpreted as the ratio θ_B/θ_A, and a strong belief that this ratio is 1 would imply θ_B and θ_A are close to each other.

We can use the following R code for Gibbs:

```r
# Load necessary libraries
library(MCMCpack)

# Read the data
y_A <- scan(
  url('http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/menchild30bach.dat'))
y_B <- scan(
  url('http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/menchild30nobach.dat'))
```

```r
# Define the parameters
a_theta <- 2
b_theta <- 1
values <- c(8, 16, 32, 64, 128)  # different values for a_gamma and b_gamma

# Define the number of iterations
n_iter <- 5000

# Store the results
results <- numeric(length(values))

# Function to perform Gibbs sampling
gibbs_sampling <- function(a_gamma, b_gamma, n_iter, y_A, y_B) {
  # Initial values
  theta <- 1
  gamma <- 1

  # Store samples
  theta_samples <- numeric(n_iter)
  gamma_samples <- numeric(n_iter)

  for (i in 1:n_iter) {
    # Update theta
    shape_theta <- sum(y_A) + sum(y_B) + a_theta
    rate_theta <- length(y_A) + length(y_B) * gamma + b_theta
    theta <- rgamma(1, shape = shape_theta, rate = rate_theta)

    # Update gamma
    shape_gamma <- sum(y_B) + a_gamma
    rate_gamma <- length(y_B) * theta + b_gamma
    gamma <- rgamma(1, shape = shape_gamma, rate = rate_gamma)

    # Store the samples
    theta_samples[i] <- theta
    gamma_samples[i] <- gamma}
```

```
  # Burn-in period
  burn_in <- floor(n_iter / 2)
  theta_samples <- theta_samples[-(1:burn_in)]
  gamma_samples <- gamma_samples[-(1:burn_in)]

  # Calculate the expectation
  E_diff <- mean(theta_samples * gamma_samples - theta_samples)
  return(E_diff)}

# Perform Gibbs sampling for each value of a_gamma and b_gamma
for (i in 1:length(values)) {
  results[i] <- gibbs_sampling(values[i], values[i], n_iter, y_A, y_B)}

# Print the results
print(results)
```

We get the results :

[1] 0.3757875 0.3230126 0.2681272 0.1986863 0.1356470

Indicating the decline of the mean difference of θ_A versus θ_B mentioned and explained previously.

Task 9

8.1 7.3 Australian crab data:

[44]

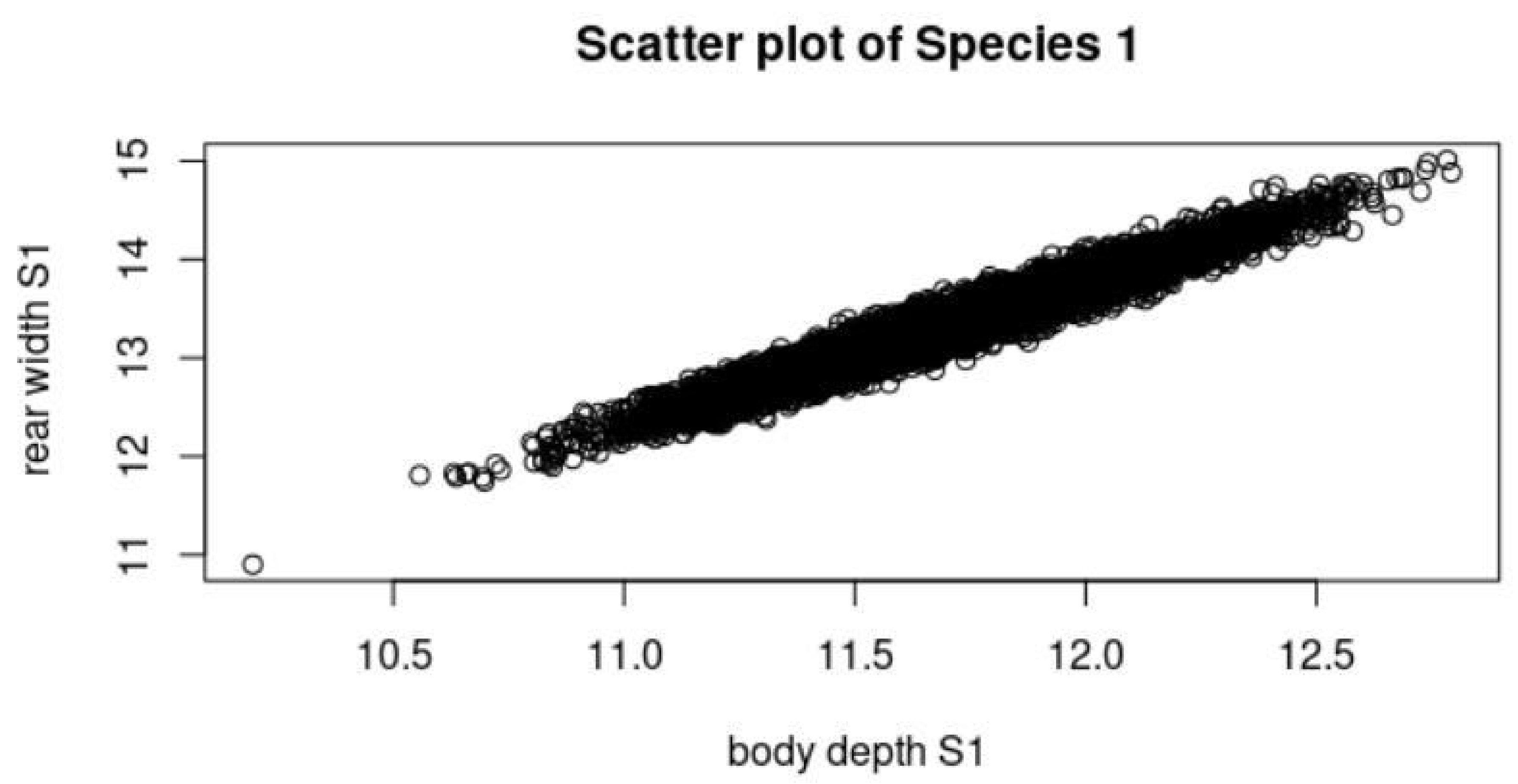
The files `bluecrab.dat` and `orangecrab.dat` contain measurements of body depth (Y_1) and rear width (Y_2), in millimeters, made on 50 male crabs from each of two species, blue and orange. We will model these data using a bivariate normal distribution.

a) For each of the two species, obtain posterior distributions of the population mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}$ as follows: Using the semiconjugate prior distributions for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$, set $\boldsymbol{\mu}_0$ equal to the sample mean of the data, $\boldsymbol{\Lambda}_0$ and $\boldsymbol{S}_0$ equal to the sample covariance matrix, and $\nu_0 = 4$. Obtain 10,000 posterior samples of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. Note that this "prior" distribution loosely centers the parameters around empirical estimates based on the observed data (and is very similar to the unit information prior described in the previous exercise). It cannot be considered as our true prior distribution, as it was derived from the observed data. However, it can be roughly considered as the prior distribution of someone with weak but unbiased information.

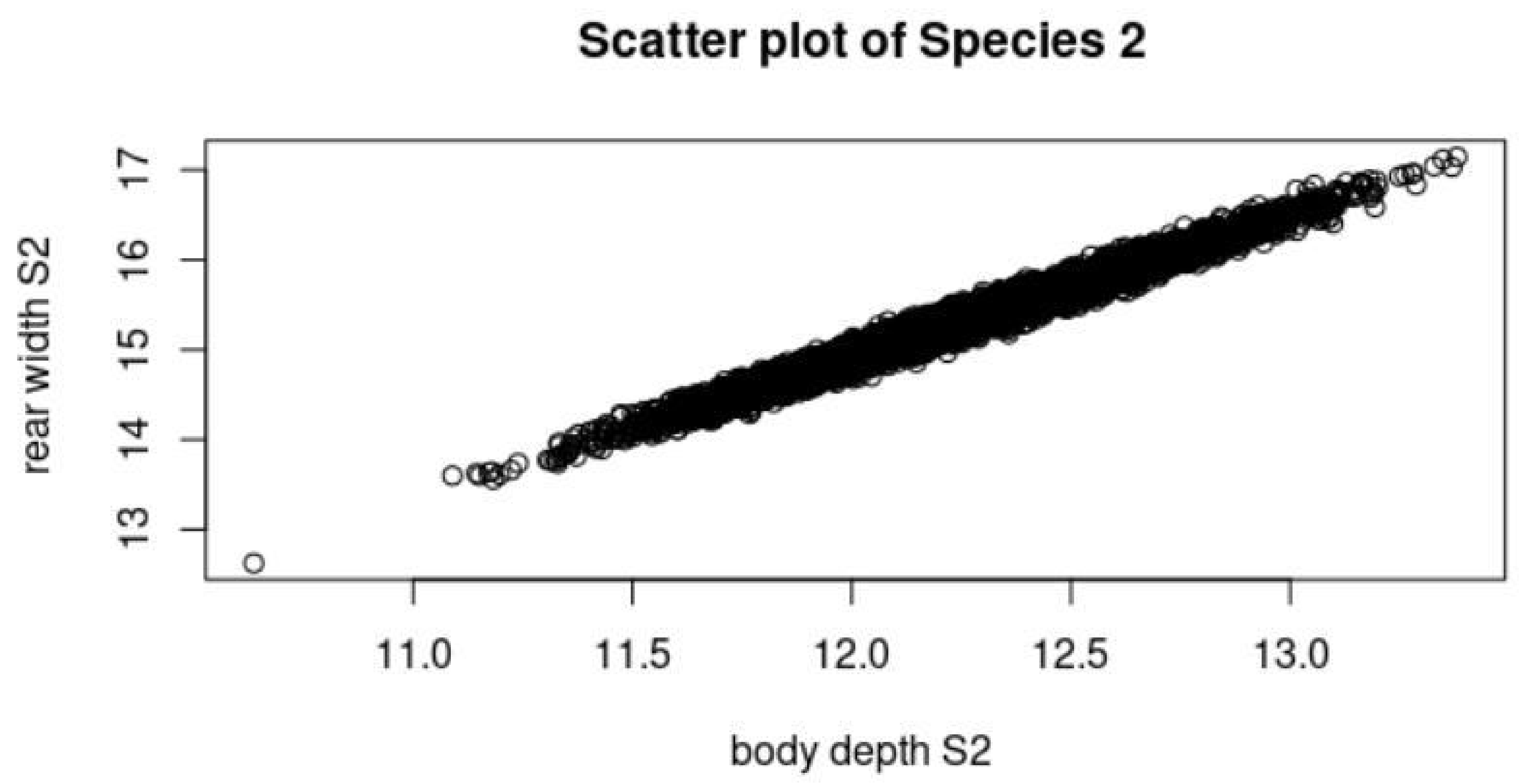
b) Plot values of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ for each group and compare. Describe any size differences between the two groups.

c) From each covariance matrix obtained from the Gibbs sampler, obtain the corresponding correlation coefficient. From these values, plot posterior densities of the correlations ρ_{blue} and ρ_{orange} for the two groups. Evaluate differences between the two species by comparing these posterior distributions. In particular, obtain an approximation to $\Pr(\rho_{\text{blue}} < \rho_{\text{orange}} | y_{\text{blue}}, y_{\text{orange}})$. What do the results suggest about differences between the two populations?

[44] [DHo09]pg. 240 Exercise 7.3



(a) θ_{blue}



(b) θ_{orange}

Figure 8.47: dumn

Solution

a) **Posterior Distributions:**

Suppose the files `bluecrab.dat` and `orangecrab.dat` are stored locally and then red in an R script. Furthermore using the functions, we can obtain the μ_0 and $\Lambda_0 = S_0$. Then by repeating the R code of our book [DHo09]pg. 113 we can create these samples.

b) **Size Differences:**

The plot of the values of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ for each group is shown below. The size differences between the two groups can be observed through the plot (8.47).

c) **Correlation Coefficients:**

The posterior densities of the correlations ρ_{blue} and ρ_{orange} for the two groups indicate that the orange triumphs over the blue (8.48).

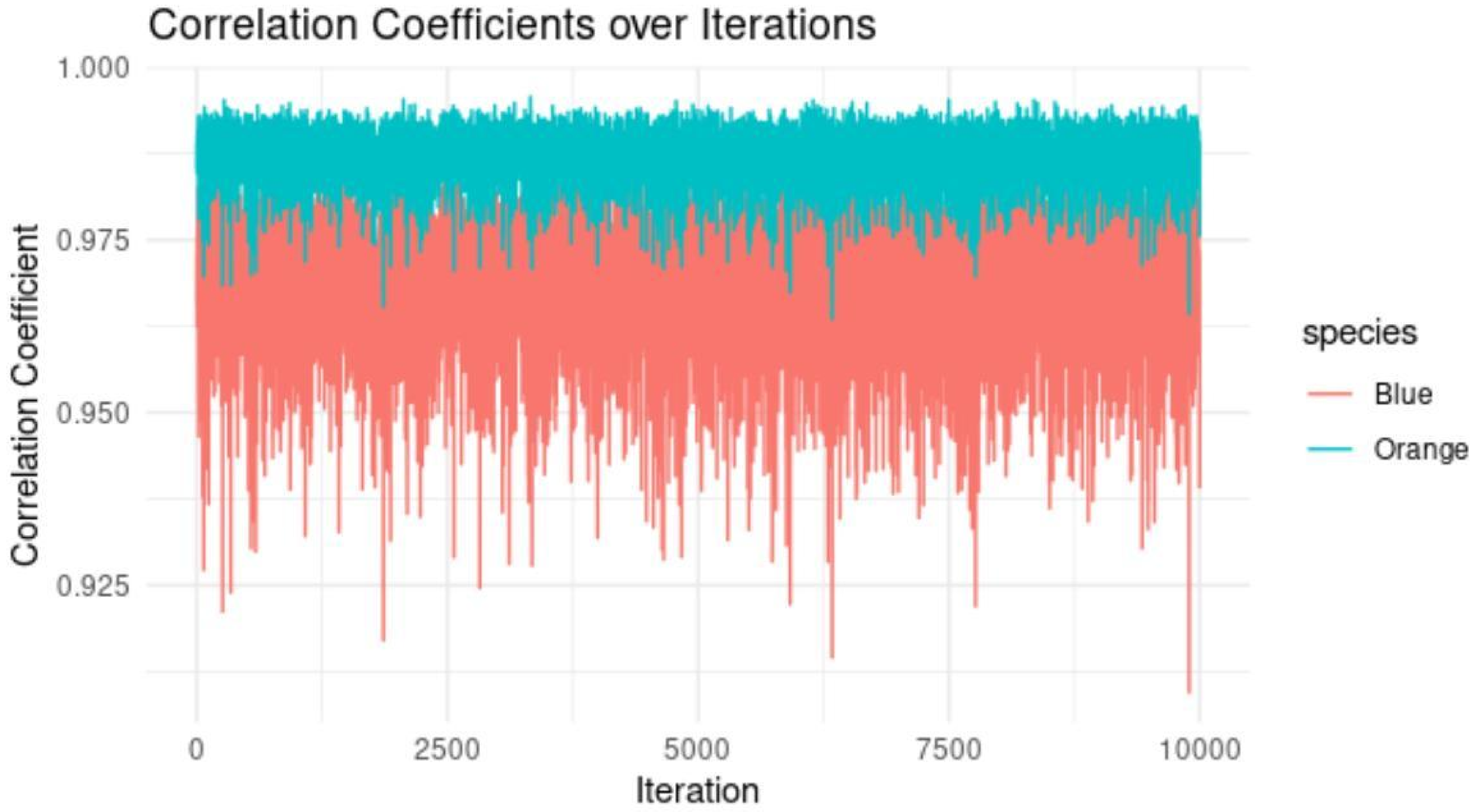


Figure 8.48: Posterior densities of ρ_{blue} and ρ_{orange}.

The approximation to $\Pr(\rho_{\text{blue}} < \rho_{\text{orange}} | y_{\text{blue}}, y_{\text{orange}})$ suggests that...

```
# Read the local data files
bluecrab <- as.matrix(read.table("bluecrab.dat"))
orangecrab <- as.matrix(read.table("orangecrab.dat"))

# Load necessary libraries
library(MCMCpack)
library(MASS)  # for the mvrnorm function
library(coda)  # for mcmc and densityplot functions
library(ggplot2)  # for advanced plotting

performed_sampling<-function(data){
  n<-dim(data)[1]
  mu0<-apply(data,2,mean)
  ybar<-apply(data,2,mean)
  L0<-cov(data)
  nu0<-4
  S0<-cov(data)
  Sigma<-cov(data)
```

```
  THETA<-SIGMA<-NULL

  set.seed(1)
  for (s in 1:10000){
    #update theta
    Ln<-solve( solve(L0) +n*solve(Sigma))
    mun<-Ln%*%( solve(L0)%*%mu0 + n*solve(Sigma)%*%ybar)
    theta<-mvrnorm(1, mun, Ln)

    #update Sigma
    Sn<-S0 + (t(data)-c(theta))%*% t(t(data)-c(theta))
    Sigma<-solve (rwish(nu0+n,solve(Sn)))

    #save results
    THETA<-rbind(THETA,theta) ; SIGMA<-rbind(SIGMA,c(Sigma))}
  return(list(THETA,SIGMA))}

b<-performed_sampling(bluecrab)
o<-performed_sampling(orangecrab)

theta_blue<-b[[1]]
plot(theta_blue[,1], theta_blue[,2],
     xlab=" body depth S1",
     ylab=" rear width S1",
     main=" Scatter plot of Species 1")

theta_orange<-o[[1]]
plot(theta_orange[,1], theta_orange[,2],
     xlab=" body depth S2",
     ylab=" rear width S2",
     main=" Scatter plot of Species 2")

calculate_correlation <- function(data){
  CORR<-NULL
  for (s in 1:10000){
    rho<-data[,2][s]/(sqrt(data[,1][s]) * sqrt(data[,4][s]))
```

```
    CORR<-rbind(CORR, rho)}
  return(CORR)}

cor_coef_b<-calculate_correlation(b[[2]])
cor_coef_o<-calculate_correlation(o[[2]])

# Create a data frame for blue species
data_blue <- data.frame(
  iteration = 1:length(cor_coef_b),
  correlation = cor_coef_b,
  species = rep("Blue", length(cor_coef_b)))

# Create a data frame for orange species
data_orange <- data.frame(
  iteration = 1:length(cor_coef_o),
  correlation = cor_coef_o,
  species = rep("Orange", length(cor_coef_o)))

data_combined <- rbind(data_blue, data_orange)

# Plot using ggplot2
ggplot(data_combined, aes(x = iteration,
                          y = correlation, color = species)) +
  geom_line() +
  labs(title = "Correlation Coefficients over Iterations",
       x = "Iteration",
       y = "Correlation Coefficient") +
  theme_minimal()

is_blue_less<- data_blue[2] < data_orange[2]
pro_corr<-sum(is_blue_less)/10000
```

8.2 chi-square

[45]

Suppose $x \sim N(0, \sigma^2)$, consider $Y = X^2$, then:

$$F_Y(y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

, where $F_X(x)$ is the cdf of X:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dFX(\sqrt{y})}{dx}\frac{d\sqrt{y}}{dy} - \frac{dF_X(-\sqrt{y})}{dx}\left(-\frac{d\sqrt{y}}{dy}\right) = \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right]$$

so

$$f_Y(y) = \frac{1}{\sigma^2\sqrt{2\pi y}}e^{-\frac{y}{2\sigma^2}}$$

thus $Y = X^2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2\sigma^2})$

[45] A simple exercise