Final Exam
of
STAT 7330 Bayesian Data Analysis

Due: 11:59 pm

December 8 (Friday), 2023

DON'T TURN THE PAGE UNTIL YOU'RE READY TO BEGIN THE EXAM.

Instructions:

1. You will have one week to complete this exam.
2. You must use LaTeX or related tools (e.g. `Rmarkdown`) to type your answers and submit the paper in pdf format.
3. Open text and open notes. NO INTERNET.
4. Statistical software packages are permitted.
5. All necessary code should be provided.
6. Discussion with anyone excluding the instrctor is forbidden
7. You must state and sign the UTD Honor Code.
8. Under the Honor Code, you must NOT DISCUSS this exam with any other person except the instructor.

The Pledge: As a Comet, I pledge honesty, integrity, and service in all that I do.

Signature: John Kenney

Printed name: John Kenney

Problem 1: Weibull model with conjugate prior

The Weibull distribution is a distribution that is often used for lifetimes of equipment/parts. It actually has two parameters but for the moment let's assume that one of them is fixed. The Weibull(2) density is

$$f(x|\theta) = 2\theta x \exp(-\theta x^2)$$

for $x > 0$. The parameter θ is something like the "inverse lifetime" parameter. Large θ means short lifetimes, while small θ means long lifetimes. The mean of the distribution is $0.886\theta^{-\frac{1}{2}}$. Suppose we observe data $x_1, \ldots, x_n$ as independent samples from the Weibull(2) distribution.

Question 1.1 (8 points)

Show that the Gamma distribution, i.e. $\text{Ga}(a, b)$, is the conjugate prior distribution and derive the posterior distribution. Use the p.d.f. $\pi(\theta) = \frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta)$.

$$\begin{aligned}
f(x_1, \ldots, x_n|\theta) &= \prod_{i=1}^{n} 2\theta x_i \exp(-\theta x_i^2) = 2^n \theta^n \left(\prod_{i=1}^{n} x_i\right) \exp\left(-\theta \sum_{i=1}^{n} x_i^2\right) \propto \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i^2\right) \\
\pi(\theta) &= \frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta) \propto \theta^{a-1}\exp(-b\theta) \\
\pi(\theta|x_1, \ldots, x_n) &\propto f(x_1, \ldots, x_n|\theta)\pi(\theta) \propto \theta^n \exp\left(-\theta \sum_{i=1}^{n} x_i^2\right)\theta^{a-1}\exp(-b\theta) = \theta^{a+n-1}\exp\left(-\theta\left(b + \sum_{i=1}^{n} x_i^2\right)\right)
\end{aligned}$$

$$\begin{aligned}
\theta &\sim Ga(a, b) \\
x_1, \ldots, x_n|\theta &\sim \text{Weibull}(2) \\
\theta|x_1, \ldots, x_n &\sim Ga(a + n, b + \sum_{i=1}^{n} x_i^2)
\end{aligned}$$

Question 1.2 (8 points)

Derive the marginal distribution $\pi(x_1, \ldots, x_n)$. (Hint: There are two related ways to obtain the marginal, through $\pi(x_1, \ldots, x_n) = \int f(x_1, \ldots, x_n|\theta)\pi(\theta)d\theta$ or $\pi(x_1, \ldots, x_n) = \frac{f(x_1,\ldots,x_n|\theta)}{\pi(\theta|x_1,\ldots,x_n)}$

$$
\begin{aligned}
\pi(x_1, \ldots, x_n) &= \int f(x_1, \ldots, x_n|\theta)\pi(\theta)d\theta \\
&= \int_0^\infty 2^n\theta^n\Big(\prod_{i=1}^n x_i\Big)\exp\Big(-\theta\sum_{i=1}^n x_i^2\Big)\frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta)d\theta \\
&= \frac{2^n b^a\Big(\prod_{i=1}^n x_i\Big)}{\Gamma(a)}\int_0^\infty \theta^{a+n-1}\exp\Big(-\theta\Big(b+\sum_{i=1}^n x_i^2\Big)\Big)d\theta \\
&= \frac{2^n b^a\Big(\prod_{i=1}^n x_i\Big)}{\Gamma(a)}\frac{\Gamma(a+n)}{\Big(b+\sum_{i=1}^n x_i^2\Big)^{a+n}}
\end{aligned}
$$

For this problem I could not think of a known distribution for the marginal distribution of the sample.

Question 1.3 (4 points)

In one application, the lifetime of a kind of gear was measured in $1,000$ hours. Let $x = 1$ means the part lasted $1,000$ hours and $x = 1/2$ means the part lasted 500 hours. Gears tend to last between 500 and $5,000$ hours, with $1,000$ being a typical lifetime. This suggests a Gamma prior distribution for θ with $a = 1.4$ and $b = 2$. Obtain a graph of this prior density and argue that this choice of prior is reasonable given the information provided.

```
# sim setting
T <- 1000
set.seed(7330)

# get density for the domain 0 to 5
prior.theta <- dgamma(seq(0,5,length.out = T),shape = 1.4,rate = 2)

library(ggplot2)
# plot prior
df <- data.frame(x = seq(0,5,length.out = T),prior.theta = prior.theta)
ggplot(data = df,aes(x = x,y = prior.theta)) + geom_line() + ylab("theta")
```

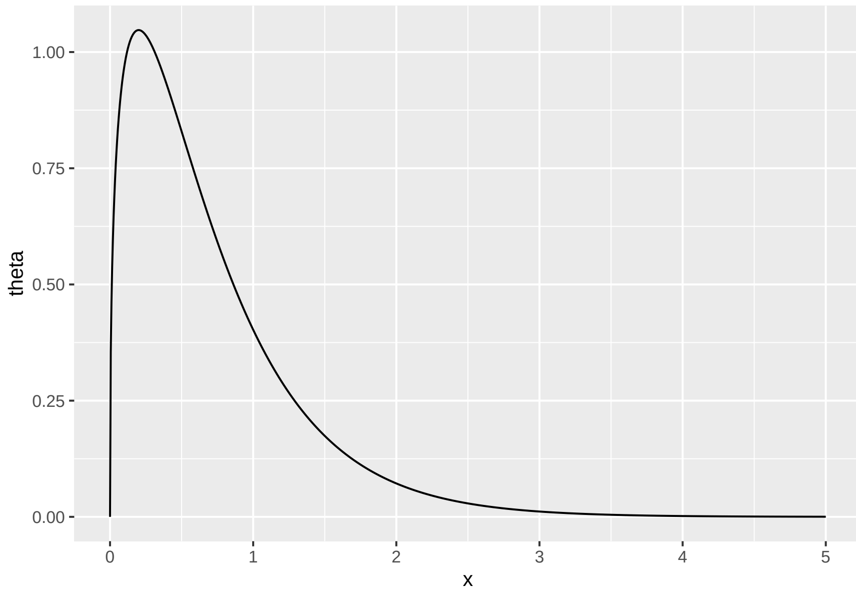


Figure 1: $\pi(\theta)$

The mean of the prior is 0.7, and if we plug this value into the mean of the weibull(2) $0.886(.7)^{-\frac{1}{2}} = 1.058973$ which is in the interval of expected interval. From this it seems that the gamma prior $a = 1.4$ and $b = 2$ works well into the expected behavior.

Question 1.4 (8 points)

Suppose that we observed $n = 10$ with

$$x = (0.25, 0.52, 0.60, 0.91, 0.97, 1.00, 1.07, 1.09, 1.18, 1.38).$$

Evidently this gear is fairly typical with lifetimes around $1,000$. Find and graph the posterior distribution. Give the posterior mean and variance. Give a 95% posterior interval for θ, using both analytical and numerical approaches.

```r
set.seed(7330)
# data stats
n <- 10
x <- c(0.25, 0.52, 0.60, 0.91, 0.97, 1.00, 1.07, 1.09, 1.18, 1.38)
x2 <- x^2
x2. <- sum(x2)

# prior settings
a <- 1.4
b <- 2

# number of samples
T <- 10000

# simulate posterior
post.theta <- rgamma(T,shape = (a+n),rate = (b + x2.))

# calc posterior summary statistics
post.mean <- mean(post.theta)
post.var <- var(post.theta)

lower.cl <- quantile(post.theta,probs = 0.025)
upper.cl <- quantile(post.theta,probs = 0.975)

# graphing posterior with mean and ci
df.post.theta <- data.frame(theta = post.theta,group = rep("post.theta",T))

ggplot(df.post.theta, aes(x = theta,fill = group,color = group)) +
  geom_density(alpha=0.2) + geom_vline(xintercept = c(lower.cl,upper.cl),color =
  ↪  "limegreen")+ geom_vline(xintercept = post.mean, color = "black")
```

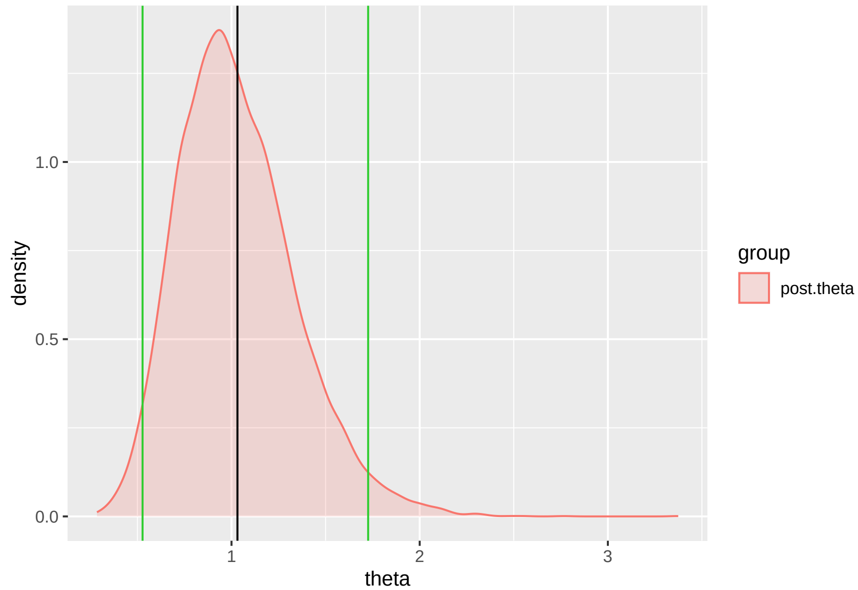


Figure 2: $\pi(\theta|x_1, \dots, x_n)$

```
# creating table for numerical and analytical results of the posterior dist
numerical <-
↪  c(round(post.mean,5),round(post.var,5),paste("(",round(lower.cl,5),",",round(upper.cl,5),")"))
analytical <- c(round((a+n)/(b + x2.),5),round((a+n)/(b +
↪  x2.)^2,5),paste("(",round(qgamma(0.025,shape = (a+n),rate = (b +
↪  x2.)),5),",",round(qgamma(0.975,shape = (a+n),rate = (b + x2.)),5),")"))

table.q1 <- as.data.frame(rbind(numerical,analytical))
row.names(table.q1) <- c("Numerical","Analytical")

library(kableExtra)
knitr::kable(table.q1,booktabs = T,col.names = c("Posterior Mean","Posterior Variance",
↪  "95% Credible interval"),caption = "Numerical vs Analytical approaches") %>%
 kable_styling(latex_options = c("hold_position"))
```

Table 1: Numerical vs Analytical approaches

	Posterior Mean	Posterior Variance	95% Credible interval
Numerical	1.03152	0.09537	(0.52734 , 1.72616)
Analytical	1.0278	0.09266	(0.52051 , 1.70475)

Problem 2: Change-point model

Consider the following hierarchical change-point model for the number of occurrence Y_i of some event during time interval i:

$$Y_i \sim \begin{cases} \text{Poi}(\theta) & i = 1, \dots, k \\ \text{Poi}(\lambda) & i = k+1, \dots, n \end{cases},$$

with the following prior formulation: $\theta \sim Ga(a_1, \beta_1)$ and $\lambda \sim Ga(a_2, \beta_2)$ with θ and λ independent. We further assume $\beta_1 \sim IG(c_1, d_1)$ and $\beta_2 \sim IG(c_2, d_2)$. The p.d.f.'s of those priors are defined as below:

$$\begin{aligned}
\pi(\theta|a_1, \beta_1) &= \frac{1}{\beta_1^{a_1}\Gamma(a_1)}\theta^{a_1-1}\exp\Big(-\frac{\theta}{\beta_1}\Big) \\
\pi(\lambda|a_2, \beta_2) &= \frac{1}{\beta_2^{a_2}\Gamma(a_2)}\lambda^{a_2-1}\exp\Big(-\frac{\lambda}{\beta_2}\Big) \\
\pi(\beta_1|c_1, d_1) &= \frac{1}{d_1^{c_1}\Gamma(c_1)}\Big(\frac{1}{\beta_1}\Big)^{c_1+1}\exp\Big(-\frac{1}{d_1\beta_1}\Big) \\
\pi(\beta_2|c_2, d_2) &= \frac{1}{d_2^{c_2}\Gamma(c_2)}\Big(\frac{1}{\beta_2}\Big)^{c_2+1}\exp\Big(-\frac{1}{d_2\beta_2}\Big)
\end{aligned}$$

Question 2.1 (10 points)

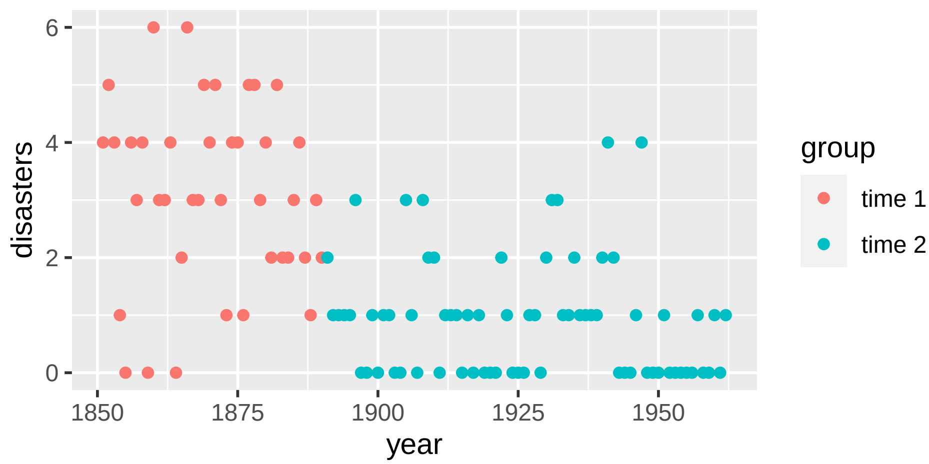
Fit this model the "coal-data.txt", which gives counts of coal mining disasters in Great Britain by year from 1851 to 1962, where disaster is defined as an accident resulting in the deaths of 10 or more minors. Set the hyperparameters to be reasonably non-informative to: $a_1 = a_2 = 0.5$, $c_1 = c_2 = 1$, and $d_1 = d_2 = 1$. Assume $k = 40$ (corresponding to year 1890). Derive full conditional distributions for θ and λ.

$$\begin{aligned}
\pi(\theta|a_1, \beta_1) &= \frac{1}{\beta_1^{a_1}\Gamma(a_1)}\theta^{a_1-1}\exp\Big(-\frac{\theta}{\beta_1}\Big) \\
\pi(\lambda|a_2, \beta_2) &= \frac{1}{\beta_2^{a_2}\Gamma(a_2)}\lambda^{a_2-1}\exp\Big(-\frac{\lambda}{\beta_2}\Big) \\
\pi(\beta_1|c_1, d_1) &= \frac{1}{d_1^{c_1}\Gamma(c_1)}\Big(\frac{1}{\beta_1}\Big)^{c_1+1}\exp\Big(-\frac{1}{d_1\beta_1}\Big) \\
\pi(\beta_2|c_2, d_2) &= \frac{1}{d_2^{c_2}\Gamma(c_2)}\Big(\frac{1}{\beta_2}\Big)^{c_2+1}\exp\Big(-\frac{1}{d_2\beta_2}\Big) \\
f(y_1, \dots, y_n|\theta, \lambda) &= \prod_{i=1}^{k}\frac{\theta^{y_i}\exp(-\theta)}{y_i!}\prod_{i=k+1}^{n}\frac{\lambda^{y_i}\exp(-\lambda)}{y_i!} \\
\pi(\theta|a_1, \beta_1, y_1, \dots, y_k) &\propto f(y_1, \dots, y_k|\theta)\pi(\theta|a_1, \beta_1) \\
&= \prod_{i=1}^{k}\frac{\theta^{y_i}\exp(-\theta)}{y_i!}\frac{1}{\beta_1^{a_1}\Gamma(a_1)}\theta^{a_1-1}\exp\Big(-\frac{\theta}{\beta_1}\Big) \\
&\propto \theta^{\sum_{i=1}^{k} y_i}\exp(-k\theta)\theta^{a_1-1}\exp\Big(-\frac{\theta}{\beta_1}\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \theta^{a_1+\sum_{i=1}^{k} y_i - 1}\exp\left(-\theta\left(\frac{k\beta_1+1}{\beta_1}\right)\right)\\
\theta|a_1,\beta_1,y_1,\ldots,y_k &\sim Ga\left(a_1+\sum_{i=1}^{k} y_i, \frac{k\beta_1+1}{\beta_1}\right)\\
\pi(\lambda|a_2,\beta_2,y_{k+1},\ldots,y_n) &\propto f(y_{k+1},\ldots,y_n|\lambda)\pi(\lambda|a_2,\beta_2)\\
&= \prod_{i=k+1}^{n}\frac{\lambda^{y_i}\exp(-\lambda)}{y_i!}\frac{1}{\beta_2^{a_2}\Gamma(a_2)}\lambda^{a_2-1}\exp\left(-\frac{\lambda}{\beta_2}\right)\\
&\propto \lambda^{\sum_{i=k+1}^{n} y_i}\exp(-(n-k)\lambda)\lambda^{a_2-1}\exp\left(-\frac{\lambda}{\beta_2}\right)\\
&= \lambda^{a_2+\sum_{i=k+1}^{n} y_i - 1}\exp\left(-\lambda\left(\frac{(n-k)\beta_2+1}{\beta_2}\right)\right)\\
\lambda|a_2,\beta_2,y_{k+1},\ldots,y_n &\sim Ga\left(a_2+\sum_{i=k+1}^{n} y_i, \frac{(n-k)\beta_2+1}{\beta_2}\right)\\
\pi(\beta_1|\theta,a_1,c_1,d_1) &\propto \pi(\theta|a_1,\beta_1)\pi(\beta_1|c_1,d_1)\\
&= \frac{1}{\beta_1^{a_1}\Gamma(a_1)}\theta^{a_1-1}\exp\left(-\frac{\theta}{\beta_1}\right)\frac{1}{d_1^{c_1}\Gamma(c_1)}\left(\frac{1}{\beta_1}\right)^{c_1+1}\exp\left(-\frac{1}{d_1\beta_1}\right)\\
&\propto \left(\frac{1}{\beta_1}\right)^{a_1}\exp\left(-\frac{\theta}{\beta_1}\right)\left(\frac{1}{\beta_1}\right)^{c_1+1}\exp\left(-\frac{1}{d_1\beta_1}\right)\\
&= \left(\frac{1}{\beta_1}\right)^{a_1+c_1+1}\exp\left(-\frac{1}{\beta_1}\left(\frac{d_1\theta+1}{d_1}\right)\right)\\
\beta_1|\theta,a_1,c_1,d_1 &\sim IG\left(a_1+c_1, \left(\frac{d_1\theta+1}{d_1}\right)\right)\\
\pi(\beta_2|\lambda,a_2,c_2,d_2) &\propto \pi(\lambda|a_2,\beta_2)\pi(\beta_2|c_2,d_2)\\
&= \frac{1}{\beta_2^{a_2}\Gamma(a_2)}\lambda^{a_2-1}\exp\left(-\frac{\lambda}{\beta_2}\right)\frac{1}{d_2^{c_2}\Gamma(c_2)}\left(\frac{1}{\beta_2}\right)^{c_2+1}\exp\left(-\frac{1}{d_2\beta_2}\right)\\
&\propto \left(\frac{1}{\beta_2}\right)^{a_2}\exp\left(-\frac{\lambda}{\beta_2}\right)\left(\frac{1}{\beta_2}\right)^{c_2+1}\exp\left(-\frac{1}{d_2\beta_2}\right)\\
&= \left(\frac{1}{\beta_2}\right)^{a_2+c_2+1}\exp\left(-\frac{1}{\beta_2}\left(\frac{d_2\lambda+1}{d_2}\right)\right)\\
\beta_2|\lambda,a_2,c_2,d_2 &\sim IG\left(a_2+c_2, \left(\frac{d_2\lambda+1}{d_2}\right)\right)
\end{aligned}
$$

```
coal.miners <-
↪  read.table("https://utdallas.box.com/shared/static/i49yk5bj1flvegj1mee3c89pnyvuzfyq.txt",
↪  col.names = c("year","disasters"))
#head(coal.miners)
```

```
library(ggplot2)
df <- cbind(coal.miners,group = c(rep("time 1",40),rep("time
↪  2",length(41:dim(coal.miners)[1]))))
ggplot(df, aes(x = year,y = disasters,color = group)) + geom_point() #+ geom_histogram()
```



Question 2.2 (10 points)

Implement a Gibbs sampler to derive the posterior distributions of θ, λ, and $R = \theta/\lambda$. Provide the histograms and kernel density estimates along with the posterior sample summaries: means, standard deviations, quantiles etc. What is meant by "convergence diagnosis"? Describe some tools you might use to assist in this regard. Present some evidence that the MCMC has reasonably converged.

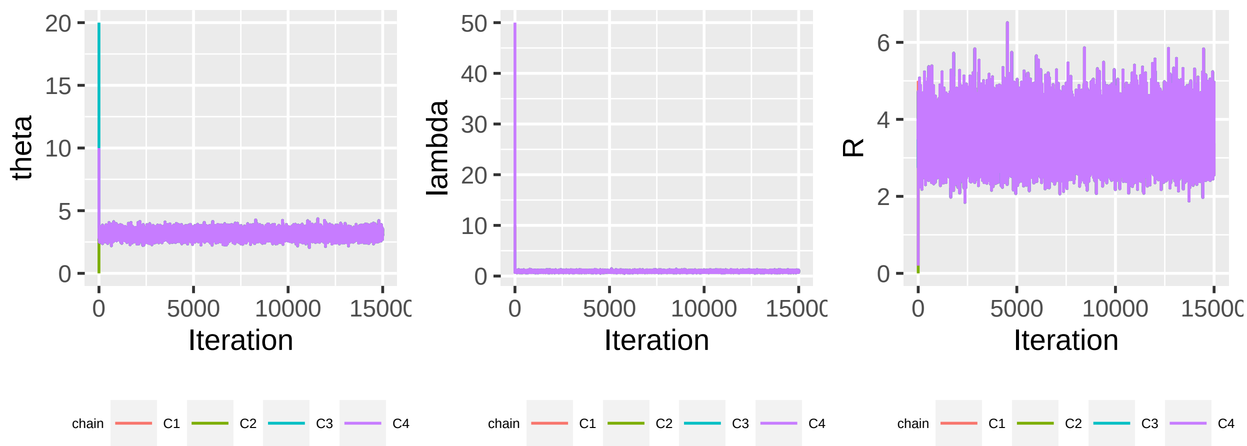
Convergence diagnostics are steps taken in analyzing the posterior samples and determining if the sequence is stable/stationary/converged to the "true value". Some options to determine whether sample estimates have not converged are:
look at a trace plot individually and jointly,
create multiple chains with different initial values and observe if all the chains converge to same value,
look at the autocorrelation if low autocorrelation then speed of mixing is fast and see if auto correlation drops over time if not this is an indicator that converge may be an issue or will be slow,
you can also use different statistical test such as Geweke statistic, Gelman-Rubin Statistic.

Below we can see that comparing the trace plots of four chains with different starting values that they all converge to around the same value, so this indicates that there is no obvious convergence issues. The joint trace plot of θ and λ also does not indicate that there are any convergence issues. The acf of each of the sequences drops really fast, so this indicates that there are no obvious mixing issues and fast movement in the parameter space. From each of these findings there seems to be no indication of convergence issues.

```
gibbs.k <- function(y,a1 = 0.5, a2 = 0.5, c1 = 1,c2 = 1,d1 = 1,d2 = 1,k = 40,T = 10000, B
↪  = T/2,theta_initial,lambda_initial,R_initial,beta1_initial,beta2_initial,seed = 7330)
↪  {
  # sim settings
  set.seed(seed)

  # data settings
  n <- length(y)
  n1 <- k
  n2 <- n-k
  y1. <- sum(y[1:k])
  y2. <- sum(y[(k+1):n])

  theta <- theta_initial
  lambda <- lambda_initial
  R <- R_initial
```

```r
  beta1 <- beta1_initial
  beta2 <- beta2_initial

  # initializing storage
  theta_store <- rep(NA,T+B)
  lambda_store <- rep(NA,T+B)
  R_store <- rep(NA,T+B)
  beta1_store <- rep(NA,T+B)
  beta2_store <- rep(NA,T+B)



  # gibbs sampler
  for(t in 1:(B+T)){
    # draw theta
    theta_shape <- a1 + y1.
    theta_rate <- (k*beta1 + 1)/beta1
    theta <- rgamma(1,shape = theta_shape,rate = theta_rate)

    # draw lambda
    lambda_shape <- a2 + y2.
    lambda_rate <- (n2*beta2 + 1)/beta2
    lambda <- rgamma(1,shape = lambda_shape,rate = lambda_rate)

    # draw beta1
    beta1_shape <- a1 + c1
    beta1_rate <- (d1*theta + 1)/d1
    beta1 <- 1/rgamma(1,shape = beta1_shape,rate = beta1_rate)

    # draw beta2
    beta2_shape <- a2 + c2
    beta2_rate <- (d2*lambda + 1)/d2
    beta2 <- 1/rgamma(1,shape = beta2_shape,rate = beta2_rate)



    # store results
    theta_store[t] <- theta
    lambda_store[t] <- lambda
    R_store[t] <- theta/lambda
    beta1_store[t] <- beta1
    beta2_store[t] <- beta2
  }
  return(list(theta_store = theta_store,
              lambda_store = lambda_store,
              R_store = R_store,
              beta1_store = beta1_store,
              beta2_store = beta2_store))
}
T <- 10000
B <- T/2

chain1 <- gibbs.k(y = coal.miners$disasters,theta_initial = 5,lambda_initial =
↪  1,R_initial = 5,beta1_initial  = 3,beta2_initial = 4)
```

```r
chain2 <- gibbs.k(y = coal.miners$disasters,theta_initial = 0,lambda_initial =
↪  10,R_initial = 0,beta1_initial  = 3,beta2_initial = 4)

chain3 <- gibbs.k(y = coal.miners$disasters,theta_initial = 20,lambda_initial =
↪  5,R_initial = 4,beta1_initial  = 3,beta2_initial = 4)

chain4 <- gibbs.k(y = coal.miners$disasters,theta_initial = 10,lambda_initial =
↪  50,R_initial = 0.2,beta1_initial  = 3,beta2_initial = 4)


df.theta.chain <- data.frame(theta = c(c(5,chain1$theta_store), c(0,chain2$theta_store),
↪  c(20,chain3$theta_store), c(10,chain4$theta_store)),
                             Iteration = c(0:(T+B),0:(T+B),0:(T+B),0:(T+B)),
                             chain =
                             ↪  c(rep("C1",B+T+1),rep("C2",B+T+1),rep("C3",B+T+1),rep("C4",B+T+1)))
df.lambda.chain <- data.frame(lambda = c(c(1,chain1$lambda_store),
↪  c(10,chain2$lambda_store), c(5,chain3$lambda_store), c(50,chain4$lambda_store)),
                             Iteration = c(0:(T+B),0:(T+B),0:(T+B),0:(T+B)),
                             chain =
                             ↪  c(rep("C1",B+T+1),rep("C2",B+T+1),rep("C3",B+T+1),rep("C4",B+T+1)))
df.R.chain <- data.frame(R = c(c(5,chain1$R_store), c(0,chain2$R_store),
↪  c(4,chain3$R_store), c(0.2,chain4$R_store)),
                             Iteration = c(0:(T+B),0:(T+B),0:(T+B),0:(T+B)),
                             chain =
                             ↪  c(rep("C1",B+T+1),rep("C2",B+T+1),rep("C3",B+T+1),rep("C4",B+T+1)))
library(ggplot2)
library(gridExtra)
grid.arrange(ggplot(df.theta.chain,aes(x = Iteration,y = theta,color = chain)) +
↪  geom_line() + theme(legend.position="bottom",legend.title = element_text(size = 5),
↪  legend.text = element_text(size = 5)) ,
             ggplot(df.lambda.chain,aes(x = Iteration,y = lambda,color = chain)) +
             ↪  geom_line() + theme(legend.position="bottom",legend.title =
             ↪  element_text(size = 5), legend.text = element_text(size = 5)) ,
             ggplot(df.R.chain,aes(x = Iteration,y = R,color = chain)) + geom_line() +
             ↪  theme(legend.position="bottom",legend.title = element_text(size = 5),
             ↪  legend.text = element_text(size = 5)) ,
             nrow = 1)
```




```r
# sim settings
set.seed(7330)
T <- 10000
B <- T/2

# prior settings
a1 <- 0.5
a2 <- 0.5
c1 <- 1
c2 <- 1
d1 <- 1
d2 <- 1
k <- 40

# data settings
y <- coal.miners$disasters
n <- length(y)
n1 <- k
n2 <- n-k
y1. <- sum(y[1:k])
y2. <- sum(y[(k+1):n])


# setting initial values
theta_initial <- 3
lambda_initial <- 1
R_initial <- theta_initial/lambda_initial
beta1_initial <- 1
beta2_initial <- 1

theta <- theta_initial
lambda <- lambda_initial
R <- R_initial
beta1 <- beta1_initial
beta2 <- beta2_initial

# initializing storage
theta_store <- rep(NA,T+B)
```

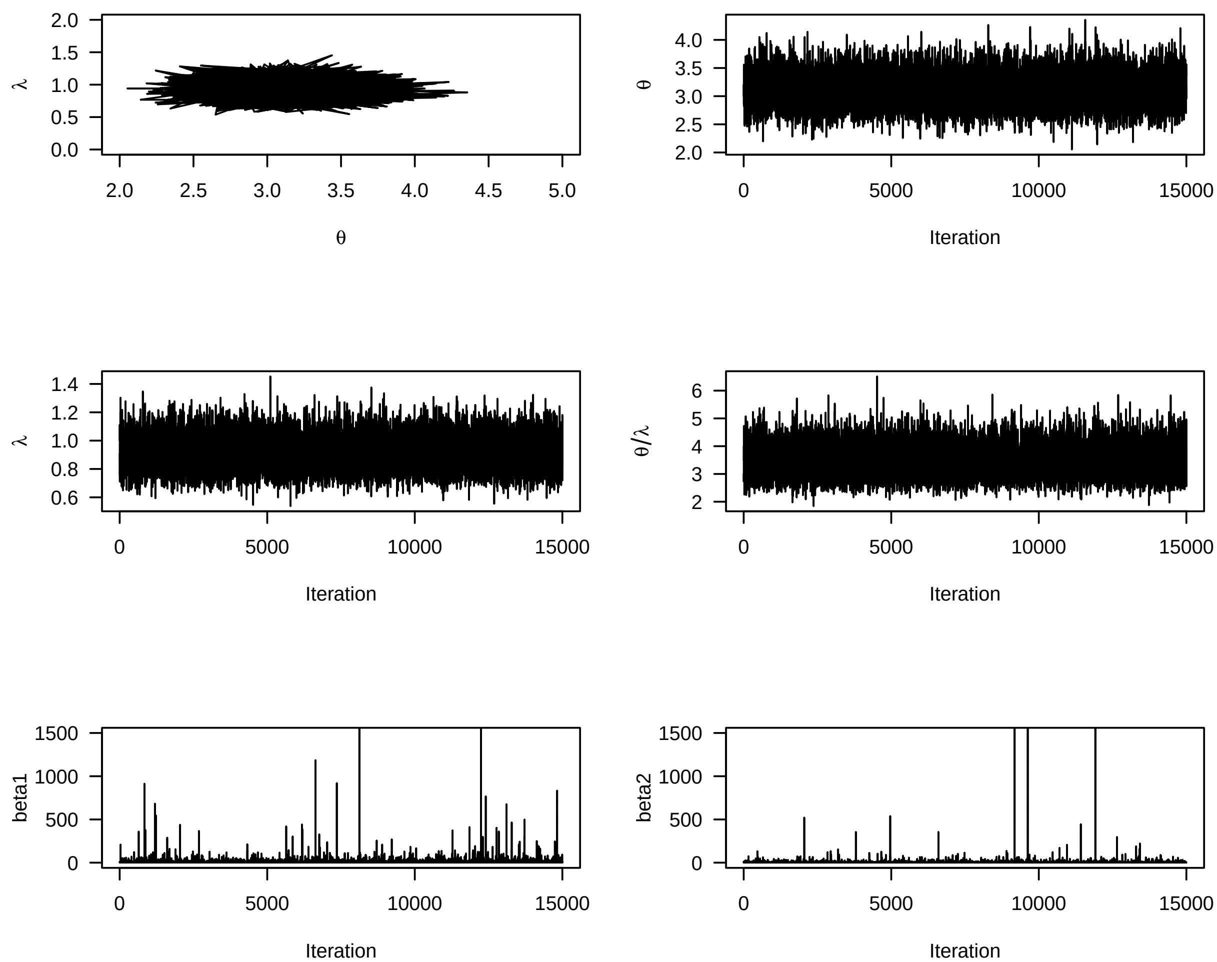
```r
lambda_store <- rep(NA,T+B)
R_store <- rep(NA,T+B)
beta1_store <- rep(NA,T+B)
beta2_store <- rep(NA,T+B)



# gibbs sampler
for(t in 1:(B+T)){
  # draw theta
  theta_shape <- a1 + y1.
  theta_rate <- (k*beta1 + 1)/beta1
  theta <- rgamma(1,shape = theta_shape,rate = theta_rate)

  # draw lambda
  lambda_shape <- a2 + y2.
  lambda_rate <- (n2*beta2 + 1)/beta2
  lambda <- rgamma(1,shape = lambda_shape,rate = lambda_rate)

  # draw beta1
  beta1_shape <- a1 + c1
  beta1_rate <- (d1*theta + 1)/d1
  beta1 <- 1/rgamma(1,shape = beta1_shape,rate = beta1_rate)

  # draw beta2
  beta2_shape <- a2 + c2
  beta2_rate <- (d2*lambda + 1)/d2
  beta2 <- 1/rgamma(1,shape = beta2_shape,rate = beta2_rate)



  # store results
  theta_store[t] <- theta
  lambda_store[t] <- lambda
  R_store[t] <- theta/lambda
  beta1_store[t] <- beta1
  beta2_store[t] <- beta2
}
```

```r
# calc posterior summary statistics theta
post.theta.meadian <- median(theta_store[(B+1):(B+T)])
post.theta.mean <- mean(theta_store[(B+1):(B+T)])
post.theta.sd <- sd(theta_store[(B+1):(B+T)])

post.theta.min <- min(theta_store[(B+1):(B+T)])
post.theta.max <- max(theta_store[(B+1):(B+T)])

post.theta.lower.cl <- quantile(theta_store[(B+1):(B+T)],probs = 0.025)
post.theta.upper.cl <- quantile(theta_store[(B+1):(B+T)],probs = 0.975)


# calc posterior summary statistics lambda
post.lambda.meadian <- median(lambda_store[(B+1):(B+T)])
```

```r
post.lambda.mean <- mean(lambda_store[(B+1):(B+T)])
post.lambda.sd <- sd(lambda_store[(B+1):(B+T)])

post.lambda.lower.cl <- quantile(lambda_store[(B+1):(B+T)],probs = 0.025)
post.lambda.upper.cl <- quantile(lambda_store[(B+1):(B+T)],probs = 0.975)

post.lambda.min <- min(lambda_store[(B+1):(B+T)])
post.lambda.max <- max(lambda_store[(B+1):(B+T)])


# calc posterior summary statistics R
post.R.meadian <- median(R_store[(B+1):(B+T)])
post.R.mean <- mean(R_store[(B+1):(B+T)])
post.R.sd <- sd(R_store[(B+1):(B+T)])

post.R.lower.cl <- quantile(R_store[(B+1):(B+T)],probs = 0.025)
post.R.upper.cl <- quantile(R_store[(B+1):(B+T)],probs = 0.975)

post.R.min <- min(R_store[(B+1):(B+T)])
post.R.max <- max(R_store[(B+1):(B+T)])


# calc posterior summary statistics beta1
post.beta1.meadian <- median(beta1_store[(B+1):(B+T)])
post.beta1.mean <- mean(beta1_store[(B+1):(B+T)])
post.beta1.sd <- sd(beta1_store[(B+1):(B+T)])

post.beta1.lower.cl <- quantile(beta1_store[(B+1):(B+T)],probs = 0.025)
post.beta1.upper.cl <- quantile(beta1_store[(B+1):(B+T)],probs = 0.975)

post.beta1.min <- min(beta1_store[(B+1):(B+T)])
post.beta1.max <- max(beta1_store[(B+1):(B+T)])

# calc posterior summary statistics beta2
post.beta2.meadian <- median(beta2_store[(B+1):(B+T)])
post.beta2.mean <- mean(beta2_store[(B+1):(B+T)])
post.beta2.sd <- sd(beta2_store[(B+1):(B+T)])

post.beta2.lower.cl <- quantile(beta2_store[(B+1):(B+T)],probs = 0.025)
post.beta2.upper.cl <- quantile(beta2_store[(B+1):(B+T)],probs = 0.975)

post.beta2.min <- min(beta2_store[(B+1):(B+T)])
post.beta2.max <- max(beta2_store[(B+1):(B+T)])
```

```r
# Monitor convergence
par(mfrow = c(3,2))
plot(cbind (c( theta_initial , theta_store ) , c( lambda_initial , lambda_store ) ) ,
↪  type = "l", xlim = c(2, 5) , ylim = c(0, 2),
      xlab = expression ( theta ) , ylab = expression ( lambda ) , las = 1)
plot(c( theta_initial , theta_store ) , type = "l",
      xlab = " Iteration ", ylab = expression ( theta ), las = 1)
plot(c( lambda_initial , lambda_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( lambda ) , las = 1)
```

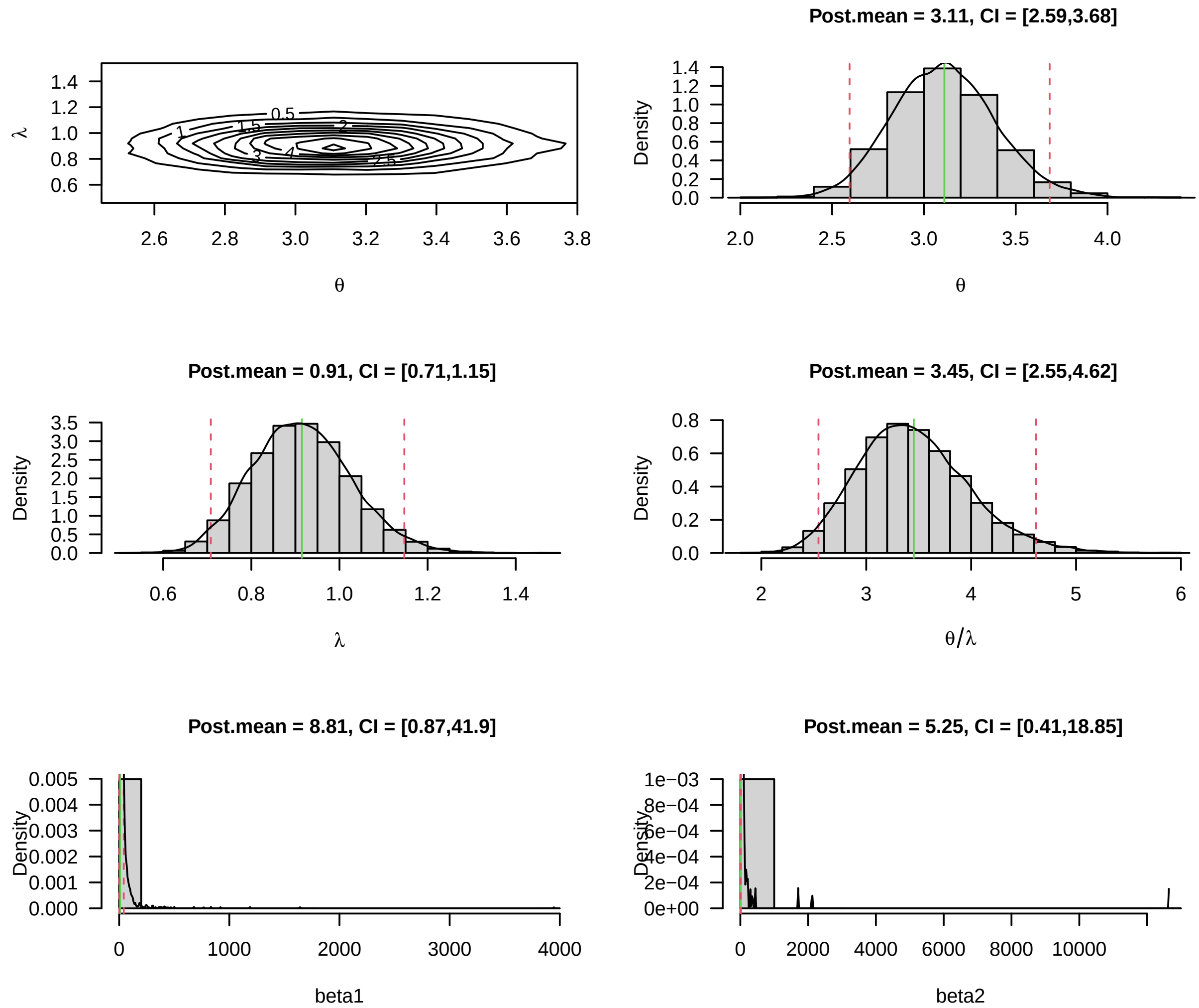
```r
plot(c( R_initial , R_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( R = theta/lambda ) , las = 1)

plot(c( beta1_initial , beta1_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( beta1 ) , las = 1,ylim = c(0,1500))
plot(c( beta2_initial , beta2_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( beta2 ) , las = 1,ylim = c(0,1500))
```



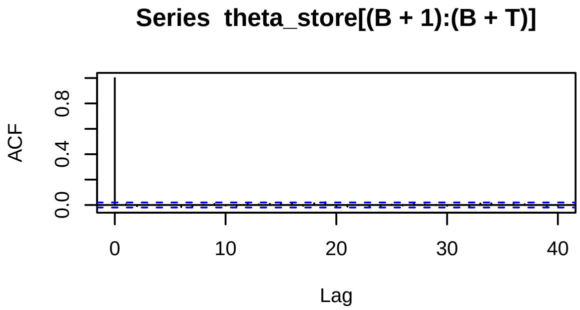

```r
par(mfrow = c(3,2))
# Visualize the joint and marginal posterior
require(MASS)
```

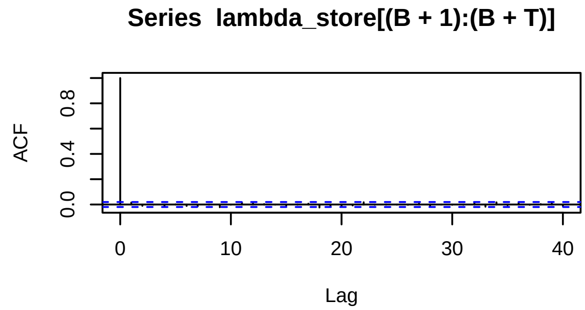
```
## Loading required package: MASS
```

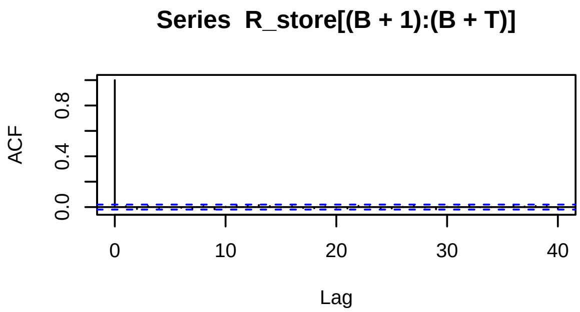
```r
# contour
z <- kde2d ( theta_store[(B+1):(B+T)] , lambda_store[(B+1):(B+T)])
contour(z , xlim = c(2.5, 3.75) , ylim = c(0.5, 1.5) , las = 1,
        xlab = expression ( theta ) , ylab =expression ( lambda ) )
# theta hist
hist(theta_store[(B+1):(B+T)], freq = FALSE , xlab = expression( theta ), las = 1,
     main = paste0(" Post.mean = ", round(post.theta.mean, 2), ", CI = [",
      ↪  round(post.theta.lower.cl ,2), ",",round(post.theta.upper.cl,2), "]") , cex.main
      ↪  = 1)
abline ( v = post.theta.mean , col = 3)
abline ( v = c(post.theta.lower.cl,post.theta.upper.cl) , col = 2, lty = 2)
lines(density(theta_store[(B+1):(B+T)]))


# lambda hist
hist(lambda_store[(B+1):(B+T)], freq = FALSE, xlab = expression( lambda ), las = 1,
     main = paste0(" Post.mean = ", round(post.lambda.mean, 2), ", CI = [",
      ↪  round(post.lambda.lower.cl ,2), ",",round(post.lambda.upper.cl,2), "]") ,
      ↪  cex.main = 1)
abline ( v = post.lambda.mean , col = 3)
abline ( v = c(post.lambda.lower.cl,post.lambda.upper.cl) , col = 2, lty = 2)
lines(density(lambda_store[(B+1):(B+T)]))

# R hist
hist(R_store[(B+1):(B+T)], freq = FALSE, xlab = expression( R = theta/ lambda ), las = 1,
 ↪
     main = paste0(" Post.mean = ", round(post.R.mean, 2), ", CI = [",
      ↪  round(post.R.lower.cl ,2), ",",round(post.R.upper.cl,2), "]") , cex.main = 1)
abline ( v = post.R.mean , col = 3)
abline ( v = c(post.R.lower.cl,post.R.upper.cl) , col = 2, lty = 2)
lines(density(R_store[(B+1):(B+T)]))

# beta1 hist
hist(beta1_store[(B+1):(B+T)], freq = FALSE, xlab = "beta1", las = 1,
     main = paste0(" Post.mean = ", round(post.beta1.mean, 2), ", CI = [",
      ↪  round(post.beta1.lower.cl ,2), ",",round(post.beta1.upper.cl,2), "]") , cex.main
      ↪  = 1)
abline ( v = post.beta1.mean , col = 3)
abline ( v = c(post.beta1.lower.cl,post.beta1.upper.cl) , col = 2, lty = 2)
lines(density(beta1_store[(B+1):(B+T)]))

# beta2 hist
hist(beta2_store[(B+1):(B+T)], freq = FALSE, xlab = "beta2", las = 1,
     main = paste0(" Post.mean = ", round(post.beta2.mean, 2), ", CI = [",
      ↪  round(post.beta2.lower.cl ,2), ",",round(post.beta2.upper.cl,2), "]") , cex.main
      ↪  = 1)
abline ( v = post.beta2.mean , col = 3)
abline ( v = c(post.beta2.lower.cl,post.beta2.upper.cl) , col = 2, lty = 2)
lines(density(beta2_store[(B+1):(B+T)]))
```

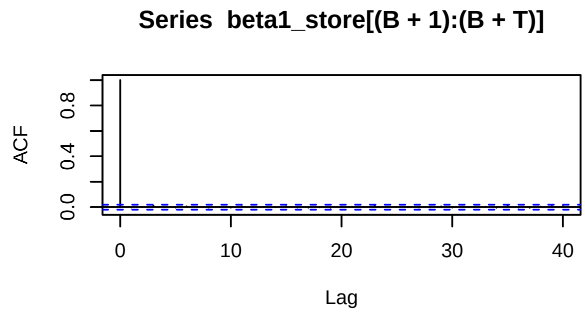


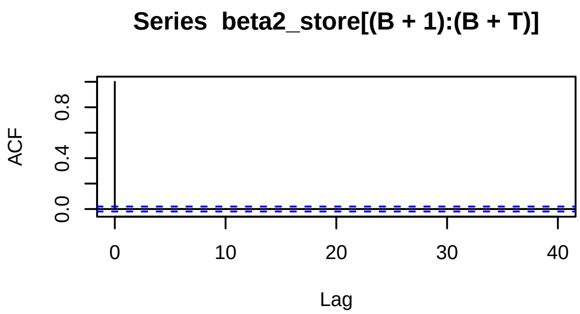

```
par(mfrow = c(3,2))
# acf plots

acf(theta_store[(B+1):(B+T)])

acf(lambda_store[(B+1):(B+T)])

acf(R_store[(B+1):(B+T)])

acf(beta1_store[(B+1):(B+T)])

acf(beta2_store[(B+1):(B+T)])
```












```
# summary table
post.theta.summary <- c(round(post.theta.min,5),round(post.theta.meadian,5),
                        round(post.theta.mean,5),round(post.theta.sd,5),
                        paste("(",round(post.theta.lower.cl,5),",",
                              round(post.theta.upper.cl,5),")"),
                        round(post.theta.max,5))
post.lambda.summary <- c(round(post.lambda.min,5),round(post.lambda.meadian,5),
                         round(post.lambda.mean,5), round(post.lambda.sd,5),
                         paste("(",round(post.lambda.lower.cl,5),",",
                               round(post.lambda.upper.cl,5),")"),
                         round(post.lambda.max,5))
post.R.summary <- c(round(post.R.min,5),round(post.R.meadian,5),round(post.R.mean,5),
↪   round(post.R.sd,5),
                    paste("(",round(post.R.lower.cl,5),",",round(post.R.upper.cl,5),")"),
                    round(post.R.max,5))
post.beta1.summary <-
↪   c(round(post.beta1.min,5),round(post.beta1.meadian,5),round(post.beta1.mean,5),round(post.beta1.sd,5

                    ↪   paste("(",round(post.beta1.lower.cl,5),",",round(post.beta1.upper.cl,5),")"),
```

```
                    round(post.beta1.max,5))
post.beta2.summary <-
↪  c(round(post.beta2.min,5),round(post.beta2.meadian,5),round(post.beta2.mean,5),round(post.beta2.sd,5

                    ↪  paste("(",round(post.beta2.lower.cl,5),",",round(post.beta2.upper.cl,5),")"),
                    round(post.beta2.max,5))

table.q22 <-
↪  as.data.frame(rbind(post.theta.summary,post.lambda.summary,post.R.summary,post.beta1.summary,post.be
row.names(table.q22) <-
↪  c("theta","lambda","R","beta1","beta2")#paste("$",c("\\theta","\\lambda","R"),"$",sep
↪  = "")
  #c("$\\theta$","$\\lambda$","$R = \\theta/\\lambda$")

library(kableExtra)
knitr::kable(table.q22,booktabs = T,col.names = c("Posterior Min","Posterior
↪  Median","Posterior Mean","Posterior sd", "95% Credible interval","Posterior
↪  Max"),caption = "Posterior Summary Statistics of $\\theta,\\lambda,R = \\theta /
↪  \\lambda$, $\\beta_1$, $\\beta_2$", align = "cccccc") %>%
  kable_styling(latex_options = c("hold_position","scale_down"))
```

Table 2: Posterior Summary Statistics of $\theta, \lambda, R = \theta/\lambda$, β_1, β_2

	Posterior Min	Posterior Median	Posterior Mean	Posterior sd	95% Credible interval	Posterior Max
theta	2.05241	3.10642	3.11134	0.2781	(2.59497 , 3.68466)	4.356
lambda	0.53804	0.91064	0.91461	0.11181	(0.708 , 1.14707)	1.45332
R	1.87774	3.41033	3.45308	0.52658	(2.54503 , 4.61844)	5.85961
beta1	0.36172	3.46552	8.81313	51.10141	(0.87443 , 41.89753)	3945.77613
beta2	0.16001	1.63836	5.25214	129.67036	(0.40776 , 18.85458)	12642.94199

Question 2.3 (10 points)

Now assume that k is unknown and adopt the following prior:

$$k \sim \text{Discrete-Unif}(1, n),$$

which should be independent of θ and λ. Now add k into the sampling chain, and obtain the marginal posterior density estimate for it. What is the effect on the posterior for $R = \theta/\lambda$?

$$\begin{aligned}
\pi(\theta|a_1,\beta_1) &= \frac{1}{\beta_1^{a_1}\Gamma(a_1)}\theta^{a_1-1}\exp\Big(-\frac{\theta}{\beta_1}\Big)\\
\pi(\lambda|a_2,\beta_2) &= \frac{1}{\beta_2^{a_2}\Gamma(a_2)}\lambda^{a_2-1}\exp\Big(-\frac{\lambda}{\beta_2}\Big)\\
\pi(\beta_1|c_1,d_1) &= \frac{1}{d_1^{c_1}\Gamma(c_1)}\Big(\frac{1}{\beta_1}\Big)^{c_1+1}\exp\Big(-\frac{1}{d_1\beta_1}\Big)\\
\pi(\beta_2|c_2,d_2) &= \frac{1}{d_2^{c_2}\Gamma(c_2)}\Big(\frac{1}{\beta_2}\Big)^{c_2+1}\exp\Big(-\frac{1}{d_2\beta_2}\Big)\\
\pi(k|n) &= \frac{1}{n}\\
f(y_1,\ldots,y_n|\theta,\lambda,k) &= \prod_{i=1}^{k}\frac{\theta^{y_i}\exp(-\theta)}{y_i!}\prod_{i=k+1}^{n}\frac{\lambda^{y_i}\exp(-\lambda)}{y_i!} = \frac{1}{\prod_{i=1}^{n}y_i!}\theta^{\sum_{i=1}^{k}y_i}\exp(-k\theta)\lambda^{\sum_{i=k+1}^{n}y_i}\exp(-(n-k)\lambda)\\
\pi(k|y_1,\ldots,y_n,\theta,\lambda) &\propto f(y_1,\ldots,y_n|\theta,\lambda,k)\pi(k|n)\\
&\propto \frac{1}{n}\frac{1}{\prod_{i=1}^{n}y_i!}\theta^{\sum_{i=1}^{k}y_i}\exp(-k\theta)\lambda^{\sum_{i=k+1}^{n}y_i}\exp(-(n-k)\lambda)\\
&\propto \theta^{\sum_{i=1}^{k}y_i}\exp(-k\theta)\lambda^{\sum_{i=k+1}^{n}y_i}\exp(k\lambda)
\end{aligned}$$

$$\begin{aligned}
p(\beta_1|\theta) &= \pi(\theta|a_1,\beta_1)\pi(\beta_1|c_1,d_1) \propto \theta^{a_1-1}IG\Big(\beta_1; a_1+c_1, \Big(\frac{d_1\theta+1}{d_1}\Big)\Big)\\
\int p(\beta_1|\theta)d\beta_1 &\propto \theta^{a_1-1}\int IG\Big(\beta_1; a_1+c_1, \Big(\frac{d_1\theta+1}{d_1}\Big)\Big)d\beta_1 = \theta^{a_1-1}\\
p(\beta_2|\lambda) &= \pi(\lambda|a_2,\beta_2)\pi(\beta_2|c_2,d_2) \propto \lambda^{a_2-1}IG\Big(\beta_2; a_2+c_2, \Big(\frac{d_2\lambda+1}{d_2}\Big)\Big)\\
\int p(\beta_2|\lambda)d\beta_2 &\propto \lambda^{a_2-1}\int IG\Big(\beta_2; a_2+c_2, \Big(\frac{d_2\lambda+1}{d_2}\Big)\Big)d\beta_2 = \lambda^{a_2-1}
\end{aligned}$$

$$\begin{aligned}
\pi(k|y_1,\ldots,y_n) &\propto \int \pi(k,\theta,\lambda,\beta_1,\beta_2|y_1,\ldots,y_n)d\theta d\lambda d\beta_1 d\beta_2\\
&= \int f(y_1,\ldots,y_n|\theta,\lambda)\pi(\theta|a_1,\beta_1)\pi(\lambda|a_2,\beta_2)\pi(\beta_1|c_1,d_1)\pi(\beta_2|c_2,d_2)\pi(k|n)d\theta d\lambda d\beta_1 d\beta_2\\
&\propto \int \frac{1}{n}\frac{\theta^{\sum_{i=1}^{k}y_i}\exp(-k\theta)}{\prod_{i=1}^{k}y_i!}\frac{\lambda^{\sum_{i=k+1}^{n}y_i}\exp(-(n-k)\lambda)}{\prod_{i=k+1}^{n}y_i!}\theta^{a_1-1}\lambda^{a_2-1}d\theta d\lambda\\
&= \frac{1}{n\prod_{i=1}^{n}y_i!}\int \theta^{a_1+\sum_{i=1}^{k}y_i-1}\exp(-k\theta)d\theta\int \lambda^{a_2+\sum_{i=k+1}^{n}y_i-1}\exp(-(n-k)\lambda)d\lambda
\end{aligned}$$

$$= \frac{1}{n\prod_{i=1}^{n} y_i!} \frac{\Gamma(a_1 + \sum_{i=1}^{k} y_i)}{k^{a_1 + \sum_{i=1}^{k} y_i}} \frac{\Gamma(a_2 + \sum_{i=k+1}^{n} y_i)}{(n-k)^{a_2 + \sum_{i=k+1}^{n} y_i}}$$

I could not find a common distribution for this marginal distribution of k.
To implement this prior on k I implemented a metropolis hastings step to draw k. I used the discrete uniform distribution as the proposal distribution. The acceptance rate is not the best around 5%, but it seems to work well.

I did not notice too much difference in the distribution/histograms of R after introduction k as unknown. The posterior summaries also do not change by any very noticeable difference, so the introduction as k as an unknown parameter did not effect the posterior of R very much. (Table 2 and Table 3)

```
# sim settings
set.seed(7330)
T <- 10000
B <- T/2

# prior settings
a1 <- 0.5
a2 <- 0.5
c1 <- 1
c2 <- 1
d1 <- 1
d2 <- 1

# data settings
y <- coal.miners$disasters
n <- length(y)


# setting initial values
theta_initial <- 3
lambda_initial <- 1
R_initial <- theta_initial/lambda_initial
beta1_initial <- 1
beta2_initial <- 1
k_initial <- n
acceptk <- 0

theta <- theta_initial
lambda <- lambda_initial
R <- R_initial
beta1 <- beta1_initial
beta2 <- beta2_initial
k <- k_initial

# initializing storage
theta_store <- rep(NA,T+B)
lambda_store <- rep(NA,T+B)
R_store <- rep(NA,T+B)
beta1_store <- rep(NA,T+B)
beta2_store <- rep(NA,T+B)
```

```r
k_store <- rep(NA,T+B)

klike <- function(y,k,theta,lambda) {
  # special cases of k
  if(k == length(y)) {
    return(theta^(sum(y[1:k])) * exp(-k*theta) )
  }
  else if(k == 1) {
    return( lambda^(sum(y[(k):n])) * exp(k*lambda) )
  }
  else {
    return(theta^(sum(y[1:k])) * exp(-k*theta) * lambda^(sum(y[(k+1):n])) * exp(k*lambda)
     ↪  )
  }
}

# gibbs sampler
for(t in 1:(B+T)){
  # special cases for k
  if(k == n) {
    n1 <- k
    n2 <- 0
    y1. <- sum(y[1:k])
    y2. <- 0
  }
  else if (k == 1) {
    n1 <- 0
    n2 <- n
    y1. <- 0
    y2. <- sum(y[1:n])
  }
  else {
    n1 <- k
    n2 <- n-k
    y1. <- sum(y[1:k])
    y2. <- sum(y[(k+1):n])
  }


  # draw theta
  theta_shape <- a1 + y1.
  theta_rate <- (k*beta1 + 1)/beta1
  theta <- rgamma(1,shape = theta_shape,rate = theta_rate)

  # draw lambda
  lambda_shape <- a2 + y2.
  lambda_rate <- (n2*beta2 + 1)/beta2
  lambda <- rgamma(1,shape = lambda_shape,rate = lambda_rate)

  # draw beta1
  beta1_shape <- a1 + c1
  beta1_rate <- (d1*theta + 1)/d1
  beta1 <- 1/rgamma(1,shape = beta1_shape,rate = beta1_rate)
```

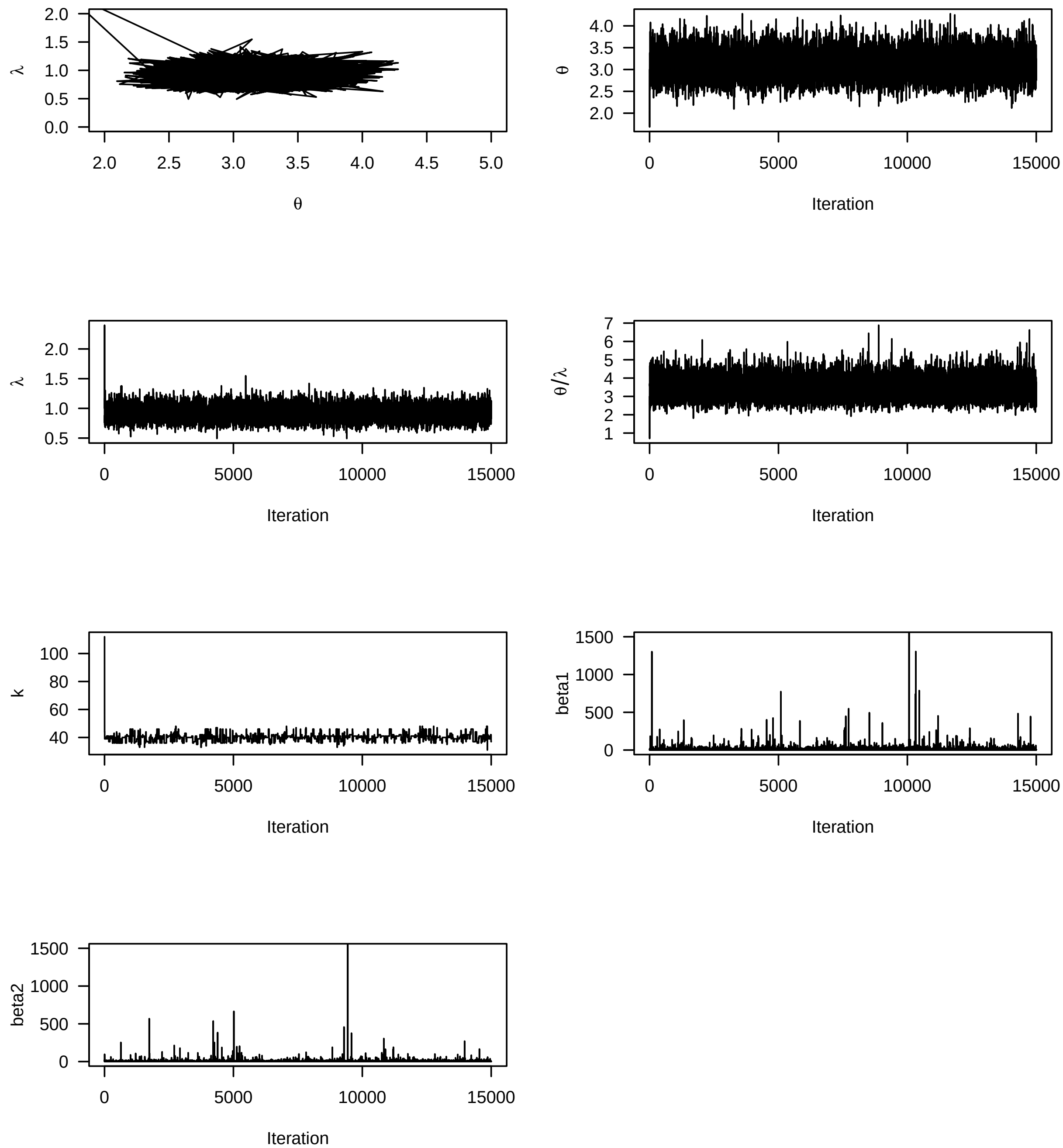
```r
  # draw beta2
  beta2_shape <- a2 + c2
  beta2_rate <- (d2*lambda + 1)/d2
  beta2 <- 1/rgamma(1,shape = beta2_shape,rate = beta2_rate)

  # draw k by mH
  # proposal dist is discrete uniform
  k_star <- sample(1:n,1,replace = T)
  rk <- klike(y,k_star,theta,lambda) / klike(y,k,theta,lambda)

  u <- runif(1)

  if(rk >= u) {
    k <- k_star
    acceptk <- acceptk+1
  }


  # store results
  theta_store[t] <- theta
  lambda_store[t] <- lambda
  R_store[t] <- theta/lambda
  beta1_store[t] <- beta1
  beta2_store[t] <- beta2
  k_store[t] <- k
}
```

```r
# calc posterior summary statistics theta
post.theta.meadian <- median(theta_store[(B+1):(B+T)])
post.theta.mean <- mean(theta_store[(B+1):(B+T)])
post.theta.sd <- sd(theta_store[(B+1):(B+T)])

post.theta.min <- min(theta_store[(B+1):(B+T)])
post.theta.max <- max(theta_store[(B+1):(B+T)])

post.theta.lower.cl <- quantile(theta_store[(B+1):(B+T)],probs = 0.025)
post.theta.upper.cl <- quantile(theta_store[(B+1):(B+T)],probs = 0.975)


# calc posterior summary statistics lambda
post.lambda.meadian <- median(lambda_store[(B+1):(B+T)])
post.lambda.mean <- mean(lambda_store[(B+1):(B+T)])
post.lambda.sd <- sd(lambda_store[(B+1):(B+T)])

post.lambda.lower.cl <- quantile(lambda_store[(B+1):(B+T)],probs = 0.025)
post.lambda.upper.cl <- quantile(lambda_store[(B+1):(B+T)],probs = 0.975)

post.lambda.min <- min(lambda_store[(B+1):(B+T)])
post.lambda.max <- max(lambda_store[(B+1):(B+T)])


# calc posterior summary statistics R
```

```
post.R.meadian <- median(R_store[(B+1):(B+T)])
post.R.mean <- mean(R_store[(B+1):(B+T)])
post.R.sd <- sd(R_store[(B+1):(B+T)])

post.R.lower.cl <- quantile(R_store[(B+1):(B+T)],probs = 0.025)
post.R.upper.cl <- quantile(R_store[(B+1):(B+T)],probs = 0.975)

post.R.min <- min(R_store[(B+1):(B+T)])
post.R.max <- max(R_store[(B+1):(B+T)])

# calc posterior summary statistics k
post.k.meadian <- median(k_store[(B+1):(B+T)])
post.k.mean <- mean(k_store[(B+1):(B+T)])
post.k.sd <- sd(k_store[(B+1):(B+T)])

post.k.lower.cl <- quantile(k_store[(B+1):(B+T)],probs = 0.025)
post.k.upper.cl <- quantile(k_store[(B+1):(B+T)],probs = 0.975)

post.k.min <- min(k_store[(B+1):(B+T)])
post.k.max <- max(k_store[(B+1):(B+T)])

# calc posterior summary statistics beta1
post.beta1.meadian <- median(beta1_store[(B+1):(B+T)])
post.beta1.mean <- mean(beta1_store[(B+1):(B+T)])
post.beta1.sd <- sd(beta1_store[(B+1):(B+T)])

post.beta1.lower.cl <- quantile(beta1_store[(B+1):(B+T)],probs = 0.025)
post.beta1.upper.cl <- quantile(beta1_store[(B+1):(B+T)],probs = 0.975)

post.beta1.min <- min(beta1_store[(B+1):(B+T)])
post.beta1.max <- max(beta1_store[(B+1):(B+T)])

# calc posterior summary statistics beta2
post.beta2.meadian <- median(beta2_store[(B+1):(B+T)])
post.beta2.mean <- mean(beta2_store[(B+1):(B+T)])
post.beta2.sd <- sd(beta2_store[(B+1):(B+T)])

post.beta2.lower.cl <- quantile(beta2_store[(B+1):(B+T)],probs = 0.025)
post.beta2.upper.cl <- quantile(beta2_store[(B+1):(B+T)],probs = 0.975)

post.beta2.min <- min(beta2_store[(B+1):(B+T)])
post.beta2.max <- max(beta2_store[(B+1):(B+T)])
```

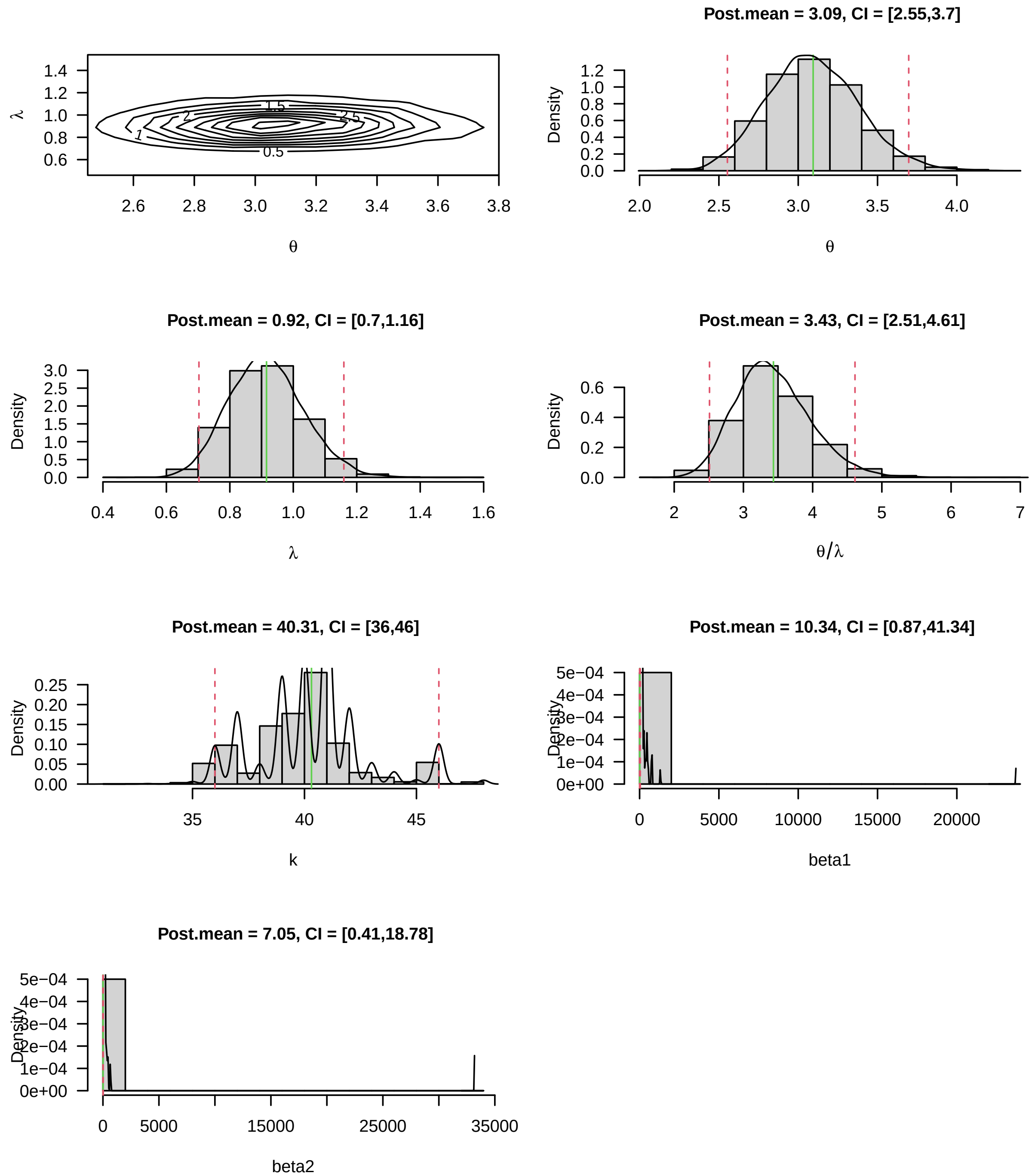
```
# Monitor convergence
par(mfrow = c(4,2))
plot(cbind (c( theta_initial , theta_store ) , c( lambda_initial , lambda_store ) ) ,
↪  type = "l", xlim = c(2, 5) , ylim = c(0, 2),
      xlab = expression ( theta ) , ylab = expression ( lambda ) , las = 1)
plot(c( theta_initial , theta_store ) , type = "l",
      xlab = " Iteration ", ylab = expression ( theta ), las = 1)
plot(c( lambda_initial , lambda_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( lambda ) , las = 1)
```

```
plot(c( R_initial , R_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( R = theta/lambda ) , las = 1)
plot(c( k_initial , k_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( k ) , las = 1)
plot(c( beta1_initial , beta1_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( beta1 ) , las = 1,ylim = c(0,1500))
plot(c( beta2_initial , beta2_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( beta2 ) , las = 1,ylim = c(0,1500))
```




```r
par(mfrow = c(4,2))
# Visualize the joint and marginal posterior
require(MASS)
# contour
z <- kde2d ( theta_store[(B+1):(B+T)] , lambda_store[(B+1):(B+T)])
contour(z , xlim = c(2.5, 3.75) , ylim = c(0.5, 1.5) , las = 1,
        xlab = expression ( theta ) , ylab =expression ( lambda ) )
# theta hist
```

```
hist(theta_store[(B+1):(B+T)], freq = FALSE , xlab = expression( theta ), las = 1,
     main = paste0(" Post.mean = ", round(post.theta.mean, 2), ", CI = [",
      ↪  round(post.theta.lower.cl ,2), ",",round(post.theta.upper.cl,2), "]") , cex.main
      ↪  = 1)
abline ( v = post.theta.mean , col = 3)
abline ( v = c(post.theta.lower.cl,post.theta.upper.cl) , col = 2, lty = 2)
lines(density(theta_store[(B+1):(B+T)]))


# lambda hist
hist(lambda_store[(B+1):(B+T)], freq = FALSE, xlab = expression( lambda ), las = 1,
     main = paste0(" Post.mean = ", round(post.lambda.mean, 2), ", CI = [",
      ↪  round(post.lambda.lower.cl ,2), ",",round(post.lambda.upper.cl,2), "]") ,
      ↪  cex.main = 1)
abline ( v = post.lambda.mean , col = 3)
abline ( v = c(post.lambda.lower.cl,post.lambda.upper.cl) , col = 2, lty = 2)
lines(density(lambda_store[(B+1):(B+T)]))

# R hist
hist(R_store[(B+1):(B+T)], freq = FALSE, xlab = expression( R = theta/ lambda ), las = 1,
 ↪
     main = paste0(" Post.mean = ", round(post.R.mean, 2), ", CI = [",
      ↪  round(post.R.lower.cl ,2), ",",round(post.R.upper.cl,2), "]") , cex.main = 1)
abline ( v = post.R.mean , col = 3)
abline ( v = c(post.R.lower.cl,post.R.upper.cl) , col = 2, lty = 2)
lines(density(R_store[(B+1):(B+T)]))

# k hist
hist(k_store[(B+1):(B+T)], freq = FALSE, xlab = "k", las = 1,
     main = paste0(" Post.mean = ", round(post.k.mean, 2), ", CI = [",
      ↪  round(post.k.lower.cl ,2), ",",round(post.k.upper.cl,2), "]") , cex.main = 1)
abline ( v = post.k.mean , col = 3)
abline ( v = c(post.k.lower.cl,post.k.upper.cl) , col = 2, lty = 2)
lines(density(k_store[(B+1):(B+T)]))

# beta1 hist
hist(beta1_store[(B+1):(B+T)], freq = FALSE, xlab = "beta1", las = 1,
     main = paste0(" Post.mean = ", round(post.beta1.mean, 2), ", CI = [",
      ↪  round(post.beta1.lower.cl ,2), ",",round(post.beta1.upper.cl,2), "]") , cex.main
      ↪  = 1)
abline ( v = post.beta1.mean , col = 3)
abline ( v = c(post.beta1.lower.cl,post.beta1.upper.cl) , col = 2, lty = 2)
lines(density(beta1_store[(B+1):(B+T)]))

# beta2 hist
hist(beta2_store[(B+1):(B+T)], freq = FALSE, xlab = "beta2", las = 1,
     main = paste0(" Post.mean = ", round(post.beta2.mean, 2), ", CI = [",
      ↪  round(post.beta2.lower.cl ,2), ",",round(post.beta2.upper.cl,2), "]") , cex.main
      ↪  = 1)
abline ( v = post.beta2.mean , col = 3)
abline ( v = c(post.beta2.lower.cl,post.beta2.upper.cl) , col = 2, lty = 2)
lines(density(beta2_store[(B+1):(B+T)]))
```




```
# summary table
post.theta.summary <- c(round(post.theta.min,5),round(post.theta.meadian,5),
                        round(post.theta.mean,5),round(post.theta.sd,5),
                        paste("(",round(post.theta.lower.cl,5),",",
                              round(post.theta.upper.cl,5),")"),
                        round(post.theta.max,5))
post.lambda.summary <- c(round(post.lambda.min,5),round(post.lambda.meadian,5),
```

```r
                         round(post.lambda.mean,5), round(post.lambda.sd,5),
                         paste("(",round(post.lambda.lower.cl,5),",",
                               round(post.lambda.upper.cl,5),")"),
                         round(post.lambda.max,5))
post.R.summary <- c(round(post.R.min,5),round(post.R.meadian,5),round(post.R.mean,5),
↪  round(post.R.sd,5),
                    paste("(",round(post.R.lower.cl,5),",",round(post.R.upper.cl,5),")"),
                    round(post.R.max,5))
post.k.summary <-
↪  c(round(post.k.min,5),round(post.k.meadian,5),round(post.k.mean,5),round(post.k.sd,5),
                    paste("(",round(post.k.lower.cl,5),",",round(post.k.upper.cl,5),")"),
                    round(post.k.max,5))
post.beta1.summary <-
↪  c(round(post.beta1.min,5),round(post.beta1.meadian,5),round(post.beta1.mean,5),round(post.beta1.sd,5

                    ↪  paste("(",round(post.beta1.lower.cl,5),",",round(post.beta1.upper.cl,5),")"),
                    round(post.beta1.max,5))
post.beta2.summary <-
↪  c(round(post.beta2.min,5),round(post.beta2.meadian,5),round(post.beta2.mean,5),round(post.beta2.sd,5

                    ↪  paste("(",round(post.beta2.lower.cl,5),",",round(post.beta2.upper.cl,5),")"),
                    round(post.beta2.max,5))

table.q22 <-
↪  as.data.frame(rbind(post.theta.summary,post.lambda.summary,post.R.summary,post.k.summary,post.beta1
row.names(table.q22) <-
↪  c("theta","lambda","R","k","beta1","beta2")#paste("$",c("\\theta","\\lambda","R"),"$",sep
↪  = "")
  #c("$\\theta$","$\\lambda$","$R = \\theta/\\lambda$")

library(kableExtra)
knitr::kable(table.q22,booktabs = T,col.names = c("Posterior Min","Posterior
↪  Median","Posterior Mean","Posterior sd", "95% Credible interval","Posterior
↪  Max"),caption = "Posterior Summary Statistics of $\\theta,\\lambda,R = \\theta /
↪  \\lambda$, k, $\\beta_1$, $\\beta_2$", align = "cccccc") %>%
  kable_styling(latex_options = c("hold_position","scale_down"))
```

Table 3: Posterior Summary Statistics of $\theta, \lambda, R = \theta/\lambda$, k, β_1, β_2

	Posterior Min	Posterior Median	Posterior Mean	Posterior sd	95% Credible interval	Posterior Max
theta	2.11694	3.08392	3.09412	0.2894	(2.55353 , 3.69666)	4.27808
lambda	0.49245	0.91157	0.91559	0.11678	(0.7026 , 1.15938)	1.54862
R	1.91515	3.38284	3.4337	0.54097	(2.5102 , 4.61203)	6.88399
k	31	40	40.3128	2.38561	(36 , 46)	48
beta1	0.3702	3.49433	10.33666	238.62555	(0.86817 , 41.33843)	23722.17797
beta2	0.17642	1.60593	7.05171	332.05212	(0.41084 , 18.77722)	33184.11718

Question 2.4 (10 points)

Now replace the third-stage priors given above with, $\beta_1 \sim Ga(c_1, d_1)$ and $\beta_2 \sim Ga(c_2, d_2)$ with β_1 and β_2 being independent, thus destroying the conjugacy for these two conditionals. Resort to a Metropolis Hastings subsampling for these two components instead, using the following hyperparameter values $c_1 = c_2 = 1$ and $d_1 = d_2 = 100$. What is the effect on the posterior distribution for β_1 and β_2. For $R = \theta/\lambda$? For k?

$$
\begin{aligned}
\pi(\theta|a_1,\beta_1) &= \frac{1}{\beta_1^{a_1}\Gamma(a_1)}\theta^{a_1-1}\exp\left(-\frac{\theta}{\beta_1}\right)\\
\pi(\lambda|a_2,\beta_2) &= \frac{1}{\beta_2^{a_2}\Gamma(a_2)}\lambda^{a_2-1}\exp\left(-\frac{\lambda}{\beta_2}\right)\\
\pi(\beta_1|c_1,d_1) &= \frac{1}{d_1^{c_1}\Gamma(c_1)}\beta_1^{c_1-1}\exp\left(-\frac{\beta_1}{d_1}\right)\\
\pi(\beta_2|c_2,d_2) &= \frac{1}{d_2^{c_2}\Gamma(c_2)}\beta_2^{c_2-1}\exp\left(-\frac{\beta_2}{d_2}\right)\\
\pi(k|n) &= \frac{1}{n}\\
f(y_1,\dots,y_n|\theta,\lambda,k) &= \prod_{i=1}^{k}\frac{\theta^{y_i}\exp(-\theta)}{y_i!}\prod_{i=k+1}^{n}\frac{\lambda^{y_i}\exp(-\lambda)}{y_i!}
\end{aligned}
$$

$$
\begin{aligned}
\pi(\beta_1|\theta) &\propto \pi(\theta|a_1,\beta_1)\pi(\beta_1|c_1,d_1)\\
&= \frac{1}{\beta_1^{a_1}\Gamma(a_1)}\theta^{a_1-1}\exp\left(-\frac{\theta}{\beta_1}\right)\frac{1}{d_1^{c_1}\Gamma(c_1)}\beta_1^{c_1-1}\exp\left(-\frac{\beta_1}{d_1}\right)\\
&\propto \frac{1}{\beta_1^{a_1}}\exp\left(-\frac{\theta}{\beta_1}\right)\beta_1^{c_1-1}\exp\left(-\frac{\beta_1}{d_1}\right)\\
&= \beta_1^{c_1-a_1-1}\exp\left(-\frac{\theta}{\beta_1}-\frac{\beta_1}{d_1}\right)
\end{aligned}
$$

$$
\begin{aligned}
\pi(\beta_2|\lambda) &\propto \pi(\lambda|a_2,\beta_2)\pi(\beta_2|c_2,d_2)\\
&= \frac{1}{\beta_2^{a_2}\Gamma(a_2)}\lambda^{a_2-1}\exp\left(-\frac{\lambda}{\beta_2}\right)\frac{1}{d_2^{c_2}\Gamma(c_2)}\beta_2^{c_2-1}\exp\left(-\frac{\beta_2}{d_2}\right)\\
&\propto \beta_2^{-a_2}\exp\left(-\frac{\lambda}{\beta_2}\right)\beta_2^{c_2-1}\exp\left(-\frac{\beta_2}{d_2}\right)\\
&= \beta_2^{c_2-a_2-1}\exp\left(-\frac{\lambda}{\beta_2}-\frac{\beta_2}{d_2}\right)
\end{aligned}
$$

$$
\begin{aligned}
\pi(\beta_1,\beta_2|\theta,\lambda) &\propto \beta_1^{c_1-a_1-1}\exp\left(-\frac{\theta}{\beta_1}-\frac{\beta_1}{d_1}\right)\beta_2^{c_2-a_2-1}\exp\left(-\frac{\lambda}{\beta_2}-\frac{\beta_2}{d_2}\right)\\
&= \beta_1^{c_1-a_1-1}\beta_2^{c_2-a_2-1}\exp\left(-\frac{\theta}{\beta_1}-\frac{\beta_1}{d_1}-\frac{\lambda}{\beta_2}-\frac{\beta_2}{d_2}\right)
\end{aligned}
$$

proposal dist:

$$
J((\beta_1^*,\beta_2^*)|(\beta_1^{(t-1)},\beta_2^{(t-1)})) \sim MN\left(((\beta_1^*,\beta_2^*));(\beta_1^{(t-1)},\beta_2^{(t-1)}),\begin{pmatrix}\psi_{\beta_1}^2 & 0\\ 0 & \psi_{\beta_2}^2\end{pmatrix}\right)
$$

The posterior distributions of β_1 and β_2 are both much narrower that in the previous sampling setting, and noticeably lower parameter estimates. One thing to note is that to me it looks like both β_1 and β_2 are not converging, so this may impact any inferences done on these two parameters. I tried different proposal distributions and different sampling settings, but could not achieve any noticeable convergence. This may indicate that I have derived my posterior incorrectly, but I could not figure out where the issue is. The histograms are more bell shaped than before using the inverse gamma prior.

For the effect on $R = \theta/\lambda$ the coverage is slightly narrower, but other than that there is not too much change from table 3 to table 4 and the histograms. This may be a part of the convergence issue of β_1 and β_2 or this may indicate that the prior choice of β_1 and β_2 has minimal change in the posterior estimates of R.

For the effect on k the range is slightly larger, but the median and coverage does not change. There is a slightly lower posterior mean for k. Overall, there seems to be little effect on the posterior of k using that improper prior Gamma for β_1 and β_2 again there may be some issues because of the convergence problem of β_1 and β_2.

```r
# sim settings
set.seed(7330)
T <- 10000
B <- T/2

# prior settings
a1 <- 0.5
a2 <- 0.5
c1 <- 1
c2 <- 1
d1 <- 100
d2 <- 100
psi_beta1 <- 0.1
psi_beta2 <- 0.1

# data settings
y <- coal.miners$disasters
n <- length(y)


# setting initial values
theta_initial <- 3
lambda_initial <- 1
R_initial <- theta_initial/lambda_initial
beta1_initial <- 1
beta2_initial <- 1
k_initial <- n
acceptk <- 0
acceptbeta <- 0

theta <- theta_initial
lambda <- lambda_initial
R <- R_initial
beta1 <- beta1_initial
beta2 <- beta2_initial
k <- k_initial

# initializing storage
theta_store <- rep(NA,T+B)
lambda_store <- rep(NA,T+B)
```

```
R_store <- rep(NA,T+B)
beta1_store <- rep(NA,T+B)
beta2_store <- rep(NA,T+B)
k_store <- rep(NA,T+B)
r_store <-rep(NA,T+B)
rk_store <-rep(NA,T+B)

klike <- function(y,k,theta,lambda) {
  # special cases of k
  if(k == length(y)) {
    return(theta^(sum(y[1:k])) * exp(-k*theta) )
  }
  else if(k == 1) {
    return( lambda^(sum(y[(k):n])) * exp(k*lambda) )
  }
  else {
    return(theta^(sum(y[1:k])) * exp(-k*theta) * lambda^(sum(y[(k+1):n])) * exp(k*lambda)
     ↪  )
  }
}

# logpost for MH for beta 1 and 2
post <- function(beta1,beta2,theta,lambda,a1,a2,c1,c2,d1,d2) {
  return( beta1^(c1 - a1 - 1)*beta2^(c2 - a2 - 1)*exp(- theta/beta1 - beta1/d1 -
   ↪  lambda/beta2 - beta2/d2))
}


# gibbs sampler
for(t in 1:(B+T)){
 # for( i in 1:1){
  # special cases for k
  if(k == n) {
    n1 <- k
    n2 <- 0
    y1. <- sum(y[1:k])
    y2. <- 0
  }
  else if (k == 1) {
    n1 <- 0
    n2 <- n
    y1. <- 0
    y2. <- sum(y[1:n])
  }
  else {
    n1 <- k
    n2 <- n-k
    y1. <- sum(y[1:k])
    y2. <- sum(y[(k+1):n])
  }


  # draw theta
```

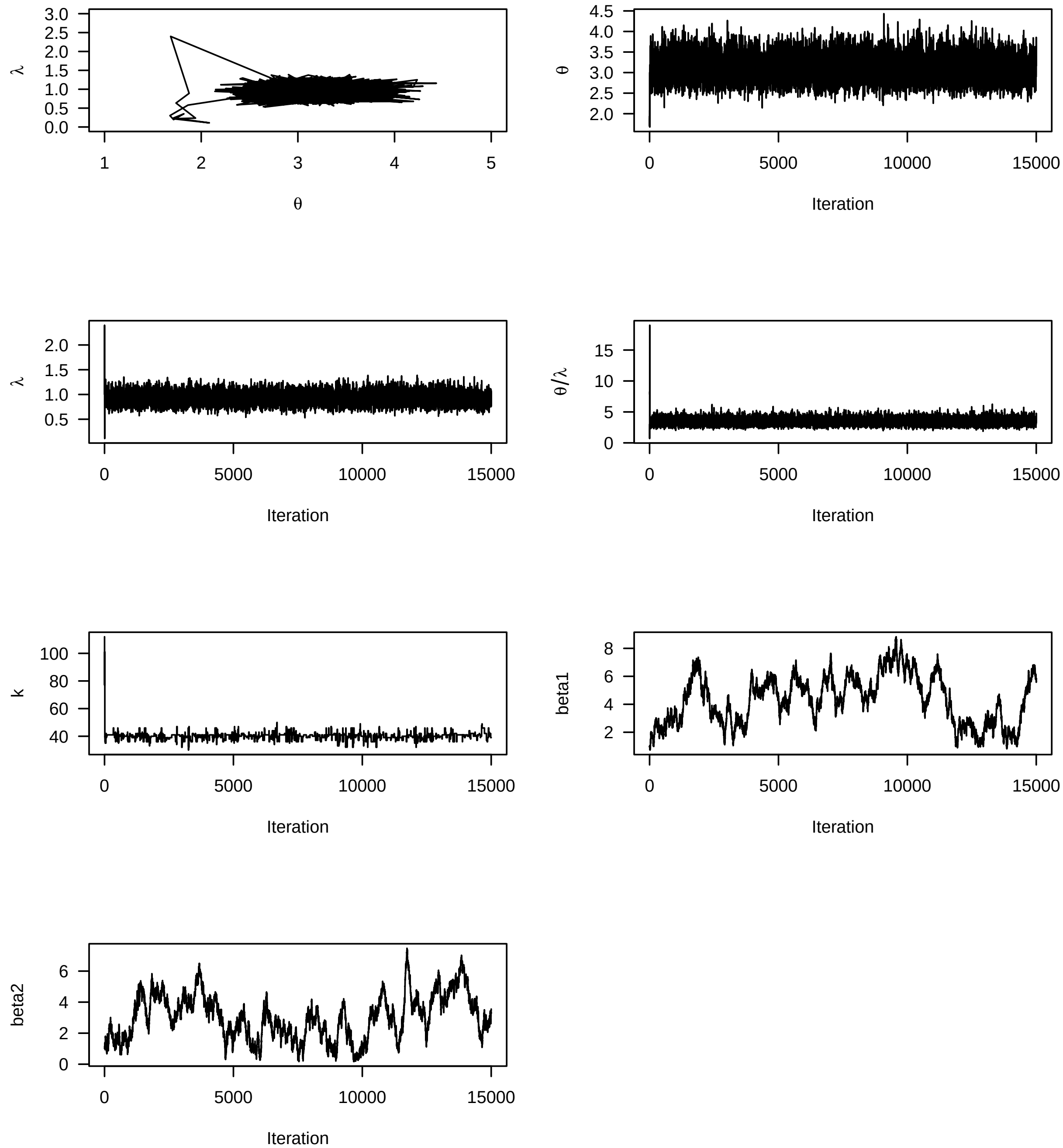
```r
  theta_shape <- a1 + y1.
  theta_rate <- (k*beta1 + 1)/beta1
  theta <- rgamma(1,shape = theta_shape,rate = theta_rate)

  # draw lambda
  lambda_shape <- a2 + y2.
  lambda_rate <- (n2*beta2 + 1)/beta2
  lambda <- rgamma(1,shape = lambda_shape,rate = lambda_rate)


  # metropolis- hastings beta1 and beta2
  beta1_star <- rnorm(1,mean = beta1,sd = psi_beta1)
  beta2_star <- rnorm(1,mean = beta2,sd = psi_beta2)
  r <- post(beta1_star,beta2_star,theta,lambda,a1,a2,c1,c2,d1,d2) /
↪  post(beta1,beta2,theta,lambda,a1,a2,c1,c2,d1,d2)

  u1 <- runif(1, min = 0,max = 1)

  if(r >= u1) {
    beta1 <- beta1_star
    beta2 <- beta2_star
    acceptbeta <- acceptbeta + 1
  }



  # draw k by mH
  # proposal dist is discrete uniform
  k_star <- sample(1:n,1,replace = T)
  rk <- klike(y,k_star,theta,lambda) / klike(y,k,theta,lambda)
  rk_store[t] <- rk
  u2 <- runif(1)

  if(rk >= u2) {
    k <- k_star
    acceptk <- acceptk+1
  }
 # }

  # store results
  theta_store[t] <- theta
  lambda_store[t] <- lambda
  R_store[t] <- theta/lambda
  beta1_store[t] <- beta1
  beta2_store[t] <- beta2
  k_store[t] <- k
}
```

```r
# calc posterior summary statistics theta
post.theta.meadian <- median(theta_store[(B+1):(B+T)])
post.theta.mean <- mean(theta_store[(B+1):(B+T)])
post.theta.sd <- sd(theta_store[(B+1):(B+T)])
```

```
post.theta.min <- min(theta_store[(B+1):(B+T)])
post.theta.max <- max(theta_store[(B+1):(B+T)])

post.theta.lower.cl <- quantile(theta_store[(B+1):(B+T)],probs = 0.025)
post.theta.upper.cl <- quantile(theta_store[(B+1):(B+T)],probs = 0.975)


# calc posterior summary statistics lambda
post.lambda.meadian <- median(lambda_store[(B+1):(B+T)])
post.lambda.mean <- mean(lambda_store[(B+1):(B+T)])
post.lambda.sd <- sd(lambda_store[(B+1):(B+T)])

post.lambda.lower.cl <- quantile(lambda_store[(B+1):(B+T)],probs = 0.025)
post.lambda.upper.cl <- quantile(lambda_store[(B+1):(B+T)],probs = 0.975)

post.lambda.min <- min(lambda_store[(B+1):(B+T)])
post.lambda.max <- max(lambda_store[(B+1):(B+T)])


# calc posterior summary statistics R
post.R.meadian <- median(R_store[(B+1):(B+T)])
post.R.mean <- mean(R_store[(B+1):(B+T)])
post.R.sd <- sd(R_store[(B+1):(B+T)])

post.R.lower.cl <- quantile(R_store[(B+1):(B+T)],probs = 0.025)
post.R.upper.cl <- quantile(R_store[(B+1):(B+T)],probs = 0.975)

post.R.min <- min(R_store[(B+1):(B+T)])
post.R.max <- max(R_store[(B+1):(B+T)])

# calc posterior summary statistics k
post.k.meadian <- median(k_store[(B+1):(B+T)])
post.k.mean <- mean(k_store[(B+1):(B+T)])
post.k.sd <- sd(k_store[(B+1):(B+T)])

post.k.lower.cl <- quantile(k_store[(B+1):(B+T)],probs = 0.025)
post.k.upper.cl <- quantile(k_store[(B+1):(B+T)],probs = 0.975)

post.k.min <- min(k_store[(B+1):(B+T)])
post.k.max <- max(k_store[(B+1):(B+T)])

# calc posterior summary statistics beta1
post.beta1.meadian <- median(beta1_store[(B+1):(B+T)])
post.beta1.mean <- mean(beta1_store[(B+1):(B+T)])
post.beta1.sd <- sd(beta1_store[(B+1):(B+T)])

post.beta1.lower.cl <- quantile(beta1_store[(B+1):(B+T)],probs = 0.025)
post.beta1.upper.cl <- quantile(beta1_store[(B+1):(B+T)],probs = 0.975)

post.beta1.min <- min(beta1_store[(B+1):(B+T)])
post.beta1.max <- max(beta1_store[(B+1):(B+T)])

# calc posterior summary statistics beta2
```

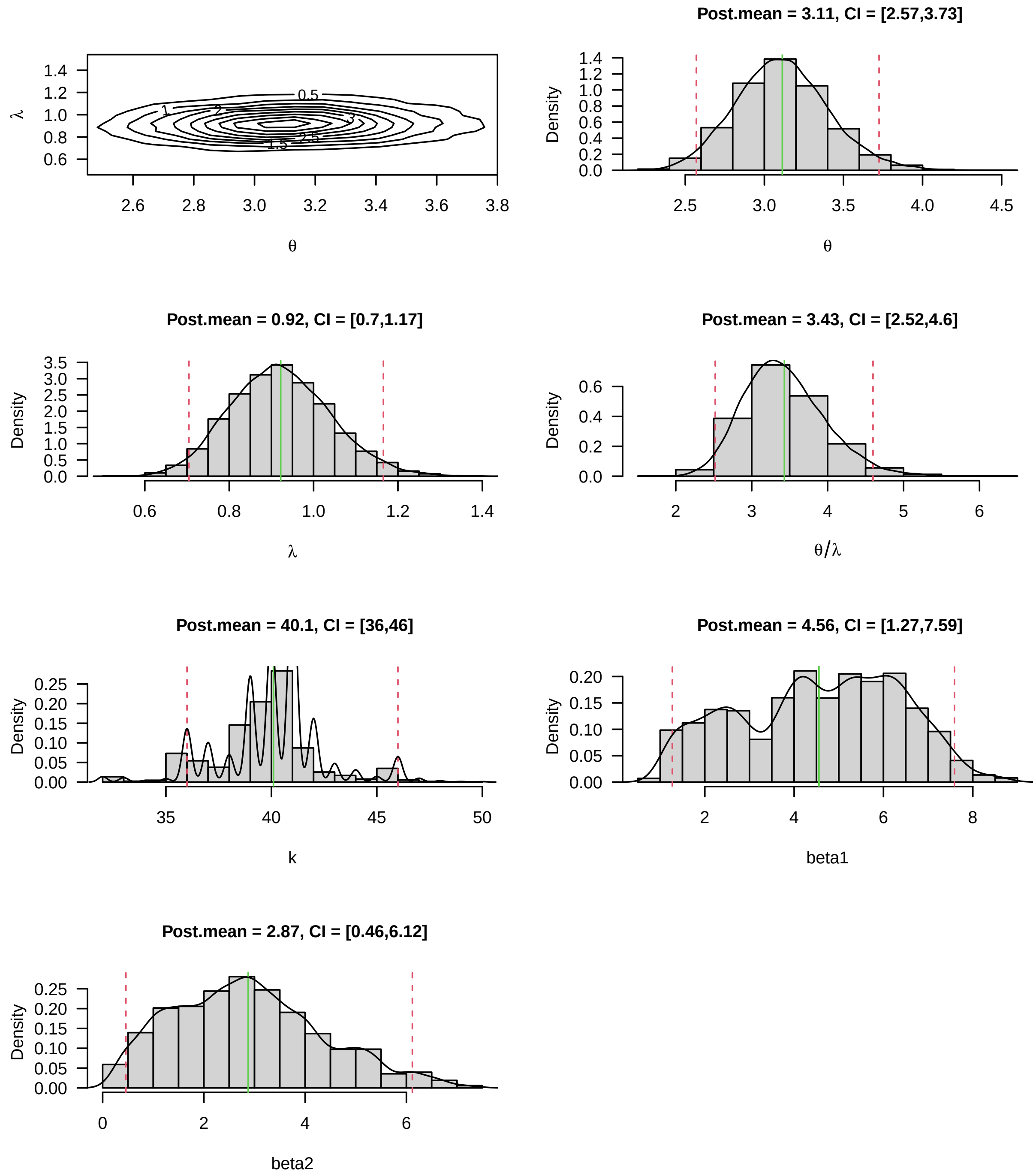
```r
post.beta2.meadian <- median(beta2_store[(B+1):(B+T)])
post.beta2.mean <- mean(beta2_store[(B+1):(B+T)])
post.beta2.sd <- sd(beta2_store[(B+1):(B+T)])

post.beta2.lower.cl <- quantile(beta2_store[(B+1):(B+T)],probs = 0.025)
post.beta2.upper.cl <- quantile(beta2_store[(B+1):(B+T)],probs = 0.975)

post.beta2.min <- min(beta2_store[(B+1):(B+T)])
post.beta2.max <- max(beta2_store[(B+1):(B+T)])
```

```r
# Monitor convergence
par(mfrow = c(4,2))
plot(cbind (c( theta_initial , theta_store ) , c( lambda_initial , lambda_store ) ) ,
↪  type = "l", xlim = c(1, 5) , ylim = c(0, 3),
      xlab = expression ( theta ) , ylab = expression ( lambda ) , las = 1)
plot(c( theta_initial , theta_store ) , type = "l",
      xlab = " Iteration ", ylab = expression ( theta ), las = 1)
plot(c( lambda_initial , lambda_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( lambda ) , las = 1)

plot(c( R_initial , R_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( R = theta/lambda ) , las = 1)
plot(c( k_initial , k_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( k ) , las = 1)
plot(c( beta1_initial , beta1_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( beta1 ) , las = 1)
plot(c( beta2_initial , beta2_store ) , type = "l",
      xlab = " Iteration ", ylab = expression( beta2 ) , las = 1)
```




```r
par(mfrow = c(4,2))
# Visualize the joint and marginal posterior
require(MASS)
# contour
z <- kde2d ( theta_store[(B+1):(B+T)] , lambda_store[(B+1):(B+T)])
contour(z , xlim = c(2.5, 3.75) , ylim = c(0.5, 1.5) , las = 1,
        xlab = expression ( theta ) , ylab =expression ( lambda ) )
# theta hist
```

```r
hist(theta_store[(B+1):(B+T)], freq = FALSE , xlab = expression( theta ), las = 1,
     main = paste0(" Post.mean = ", round(post.theta.mean, 2), ", CI = [",
      ↪ round(post.theta.lower.cl ,2), ",",round(post.theta.upper.cl,2), "]") , cex.main
      ↪ = 1)
abline ( v = post.theta.mean , col = 3)
abline ( v = c(post.theta.lower.cl,post.theta.upper.cl) , col = 2, lty = 2)
lines(density(theta_store[(B+1):(B+T)]))


# lambda hist
hist(lambda_store[(B+1):(B+T)], freq = FALSE, xlab = expression( lambda ), las = 1,
     main = paste0(" Post.mean = ", round(post.lambda.mean, 2), ", CI = [",
      ↪ round(post.lambda.lower.cl ,2), ",",round(post.lambda.upper.cl,2), "]") ,
      ↪ cex.main = 1)
abline ( v = post.lambda.mean , col = 3)
abline ( v = c(post.lambda.lower.cl,post.lambda.upper.cl) , col = 2, lty = 2)
lines(density(lambda_store[(B+1):(B+T)]))

# R hist
hist(R_store[(B+1):(B+T)], freq = FALSE, xlab = expression( R = theta/ lambda ), las = 1,
 ↪
     main = paste0(" Post.mean = ", round(post.R.mean, 2), ", CI = [",
      ↪ round(post.R.lower.cl ,2), ",",round(post.R.upper.cl,2), "]") , cex.main = 1)
abline ( v = post.R.mean , col = 3)
abline ( v = c(post.R.lower.cl,post.R.upper.cl) , col = 2, lty = 2)
lines(density(R_store[(B+1):(B+T)]))

# k hist
hist(k_store[(B+1):(B+T)], freq = FALSE, xlab = "k", las = 1,
     main = paste0(" Post.mean = ", round(post.k.mean, 2), ", CI = [",
      ↪ round(post.k.lower.cl ,2), ",",round(post.k.upper.cl,2), "]") , cex.main = 1)
abline ( v = post.k.mean , col = 3)
abline ( v = c(post.k.lower.cl,post.k.upper.cl) , col = 2, lty = 2)
lines(density(k_store[(B+1):(B+T)]))

# beta1 hist
hist(beta1_store[(B+1):(B+T)], freq = FALSE, xlab = "beta1", las = 1,
     main = paste0(" Post.mean = ", round(post.beta1.mean, 2), ", CI = [",
      ↪ round(post.beta1.lower.cl ,2), ",",round(post.beta1.upper.cl,2), "]") , cex.main
      ↪ = 1)
abline ( v = post.beta1.mean , col = 3)
abline ( v = c(post.beta1.lower.cl,post.beta1.upper.cl) , col = 2, lty = 2)
lines(density(beta1_store[(B+1):(B+T)]))

# beta2 hist
hist(beta2_store[(B+1):(B+T)], freq = FALSE, xlab = "beta2", las = 1,
     main = paste0(" Post.mean = ", round(post.beta2.mean, 2), ", CI = [",
      ↪ round(post.beta2.lower.cl ,2), ",",round(post.beta2.upper.cl,2), "]") , cex.main
      ↪ = 1)
abline ( v = post.beta2.mean , col = 3)
abline ( v = c(post.beta2.lower.cl,post.beta2.upper.cl) , col = 2, lty = 2)
lines(density(beta2_store[(B+1):(B+T)]))
```




```
# summary table
post.theta.summary <- c(round(post.theta.min,5),round(post.theta.meadian,5),
                        round(post.theta.mean,5),round(post.theta.sd,5),
                        paste("(",round(post.theta.lower.cl,5),",",
                              round(post.theta.upper.cl,5),")"),
                        round(post.theta.max,5))
post.lambda.summary <- c(round(post.lambda.min,5),round(post.lambda.meadian,5),
```

```
                          round(post.lambda.mean,5), round(post.lambda.sd,5),
                          paste("(",round(post.lambda.lower.cl,5),",",
                                round(post.lambda.upper.cl,5),")"),
                          round(post.lambda.max,5))
post.R.summary <- c(round(post.R.min,5),round(post.R.meadian,5),round(post.R.mean,5),
↪  round(post.R.sd,5),
                     paste("(",round(post.R.lower.cl,5),",",round(post.R.upper.cl,5),")"),
                     round(post.R.max,5))
post.k.summary <-
↪  c(round(post.k.min,5),round(post.k.meadian,5),round(post.k.mean,5),round(post.k.sd,5),
                     paste("(",round(post.k.lower.cl,5),",",round(post.k.upper.cl,5),")"),
                     round(post.k.max,5))
post.beta1.summary <-
↪  c(round(post.beta1.min,5),round(post.beta1.meadian,5),round(post.beta1.mean,5),
                         round(post.beta1.sd,5),

                     ↪  paste("(",round(post.beta1.lower.cl,5),",",round(post.beta1.upper.cl,5),")"),
                     round(post.beta1.max,5))
post.beta2.summary <-
↪  c(round(post.beta2.min,5),round(post.beta2.meadian,5),round(post.beta2.mean,5),
                         round(post.beta2.sd,5),

                     ↪  paste("(",round(post.beta2.lower.cl,5),",",round(post.beta2.upper.cl,5),")"),
                     round(post.beta2.max,5))

table.q22 <- as.data.frame(rbind(post.theta.summary,post.lambda.summary,
                                 post.R.summary,post.k.summary,
                                 post.beta1.summary,post.beta2.summary))
row.names(table.q22) <-
↪  c("theta","lambda","R","k","beta1","beta2")#paste("$",c("\\theta","\\lambda","R"),"$",sep
↪  = "")
  #c("$\\theta$","$\\lambda$","$R = \\theta/\\lambda$")

library(kableExtra)
knitr::kable(table.q22,booktabs = T,col.names = c("Posterior Min","Posterior
↪  Median","Posterior Mean","Posterior sd", "95% Credible interval","Posterior
↪  Max"),caption = "Posterior Summary Statistics of $\\theta,\\lambda,R = \\theta /
↪  \\lambda$, k, $\\beta_1$, $\\beta_2$", align = "cccccc") %>%
  kable_styling(latex_options = c("hold_position","scale_down"))
```

Table 4: Posterior Summary Statistics of $\theta, \lambda, R = \theta/\lambda$, k, β_1, β_2

	Posterior Min	Posterior Median	Posterior Mean	Posterior sd	95% Credible interval	Posterior Max
theta	2.20264	3.10523	3.11349	0.29188	(2.56977 , 3.72504)	4.43249
lambda	0.52861	0.91713	0.92208	0.11809	(0.70458 , 1.16551)	1.38715
R	1.87189	3.38182	3.43041	0.53547	(2.51964 , 4.59751)	6.24889
k	32	40	40.0984	2.44723	(36 , 46)	50
beta1	0.81342	4.69465	4.55587	1.80343	(1.27229 , 7.58913)	8.83001
beta2	0.1677	2.78378	2.8749	1.47073	(0.45902 , 6.11796)	7.47165

Problem 3: Model comparison on regression models

For the following data we consider two competing models: H_1: a linear vs. H_2: a quadratic regression.

x_i	-1.9	-0.39	0.79	-0.20	0.42	-0.35	0.67	0.63	-0.024	1.2
y_i	-1.7	-0.23	0.50	-0.66	1.97	0.10	0.60	1.13	-0.943	2.6

Model H_1:

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_i + \epsilon_i, \quad i = 1, \ldots, n \\ \beta_1 &\sim N(0,1) \\ \beta_2 &\sim N(1,1) \\ \epsilon_i &\sim N(0,1) \end{aligned}$$

with β_1 and β_2 a priori independent.

Model H_2:

$$\begin{aligned} y_i &= \gamma_1 + \gamma_2 x_i + \gamma_3 x_i^2 + \epsilon_i, \quad i = 1, \ldots, n \\ \gamma_1, \gamma_3 &\sim N(0,1) \\ \gamma_2 &\sim N(1,1) \\ \epsilon_i &\sim N(0,1) \end{aligned}$$

with γ_1, γ_2, and γ_3 a prior independent.

Question 3.1 (7 points)

Find the marginal distributions $\pi(y_1, \ldots, y_n|H_1) = \int f(y_1, \ldots, y_n|\beta_1, \beta_2)\pi(\beta_1)\pi(\beta_2)d\beta_1 d\beta_2$

$$\pi(y_1, \ldots, y_n|H_1) = \int f(y_1, \ldots, y_n|\beta_1, \beta_2)\pi(\beta_1)\pi(\beta_2)d\beta_1 d\beta_2$$

Integrate out β_1 :

$$\begin{aligned} p(\boldsymbol{y}|\beta_2) &= \int f(y_1, \ldots, y_n|\beta_1, \beta_2)\pi(\beta_1)d\beta_1 \\ &= \int N(y_1, \ldots, y_n; \beta_1 1_n + \boldsymbol{x}\beta_2, I_n)N(\beta_1; 0, 1)d\beta_1 \\ &= \int (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \beta_1 \mathbf{1}_n - \boldsymbol{x}\beta_2)^\top (\boldsymbol{y} - \beta_1 \mathbf{1}_n - \boldsymbol{x}\beta_2)\right)(2\pi)^{-\frac{1}{2}} \exp\left(-\frac{\beta_1^2}{2}\right) d\beta_1 \\ &= (2\pi)^{-\frac{n+1}{2}} \int \exp\left(-\frac{1}{2}(\boldsymbol{y} - \beta_1 \mathbf{1}_n - \boldsymbol{x}\beta_2)^\top (\boldsymbol{y} - \beta_1 \mathbf{1}_n - \boldsymbol{x}\beta_2)\right) \exp\left(-\frac{\beta_1^2}{2}\right) d\beta_1 \end{aligned}$$

$$
\begin{aligned}
&=(2\pi)^{-\frac{n+1}{2}}\int \exp\Big(-\frac{1}{2}(\boldsymbol{y}\boldsymbol{y}^\top - n\beta_1^2 - \beta_2^2\boldsymbol{x}^\top\boldsymbol{x}\Big)\exp\Big(\beta_1\mathbf{1}_n^\top\boldsymbol{y} - \beta_1\beta_2\mathbf{1}_n^\top\boldsymbol{x} - \beta_2\boldsymbol{x}^\top\boldsymbol{y}\Big)\exp\Big(-\frac{\beta_1^2}{2}\Big)d\beta_1\\
&=(2\pi)^{-\frac{n+1}{2}}\exp\Big(-\frac{1}{2}(\boldsymbol{y}\boldsymbol{y}^\top)\Big)\exp\Big(-\frac{1}{2}\beta_2^2\boldsymbol{x}^\top\boldsymbol{x} + \beta_2\boldsymbol{x}^\top\boldsymbol{y}\Big)\int\exp\Big(-\frac{n+1}{2}\beta_1^2 + \Big(\mathbf{1}_n^\top\boldsymbol{y} - \beta_2\mathbf{1}_n^\top\boldsymbol{x}\Big)\beta_1\Big)d\beta_1\\
&=(2\pi)^{-\frac{n+1}{2}}\exp\Big(-\frac{1}{2}(\boldsymbol{y}\boldsymbol{y}^\top)\Big)\exp\Big(-\frac{1}{2}\beta_2^2\boldsymbol{x}^\top\boldsymbol{x} + \beta_2\boldsymbol{x}^\top\boldsymbol{y}\Big)(2\pi)^{\frac{1}{2}}\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\\
&\quad\exp\Big(\frac{1}{2}\frac{1}{n+1}(\boldsymbol{y}-\beta_2\boldsymbol{x})^\top\mathbf{1}_n\mathbf{1}_n^\top(\boldsymbol{y}-\beta_2\boldsymbol{x})\Big)\int N\Big(\beta_1;\frac{1}{n+1}\Big(\mathbf{1}_n^\top\boldsymbol{y} - \beta_2\mathbf{1}_n^\top\boldsymbol{x}\Big),\frac{1}{n+1}\Big)d\beta_1\\
&=(2\pi)^{-\frac{n}{2}}\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}(\boldsymbol{y}\boldsymbol{y}^\top)\Big)\exp\Big(-\frac{1}{2}\beta_2^2\boldsymbol{x}^\top\boldsymbol{x} + \beta_2\boldsymbol{x}^\top\boldsymbol{y}\Big)\\
&\quad\exp\Big(\frac{1}{2}\frac{1}{n+1}\boldsymbol{y}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{y} + \frac{1}{2}\frac{1}{n+1}\beta_2^2\boldsymbol{x}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{x} - \frac{1}{n+1}\beta_2\boldsymbol{x}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{y}\Big)\\
&=(2\pi)^{-\frac{n}{2}}\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}\boldsymbol{y}(\boldsymbol{I}_n - \frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{y}^\top\Big)\\
&\quad\exp\Big(-\frac{1}{2}\beta_2^2\boldsymbol{x}^\top(\boldsymbol{I}_n - \frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{x} + \beta_2\boldsymbol{x}^\top(\boldsymbol{I}_n - \frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{y}\Big)\\
&=(2\pi)^{-\frac{n}{2}}\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\beta_2^2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x} + \beta_2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\\
&\quad\text{where } \boldsymbol{H}_n = \boldsymbol{I}_n - \frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top
\end{aligned}
$$

Integrate out β_2 :

$$
\begin{aligned}
p(\boldsymbol{y}|H_1) &= \int p(\boldsymbol{y}|\beta_2)p(\beta_2)d\beta_2\\
&= \int p(\boldsymbol{y}|\beta_2)N(\beta_2;1,1)d\beta_2\\
&= \int (2\pi)^{-\frac{n}{2}}\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\beta_2^2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x} + \beta_2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\\
&\quad(2\pi)^{-\frac{1}{2}}\exp\Big(-\frac{(\beta_2-1)^2}{2}\Big)d\beta_2\\
&=(2\pi)^{-\frac{n+1}{2}}\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\int\exp\Big(-\frac{1}{2}\beta_2^2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x} + \beta_2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\\
&\quad\exp\Big(-\frac{\beta_2^2}{2} + \beta_2 - \frac{1}{2}\Big)d\beta_2\\
&=(2\pi)^{-\frac{n+1}{2}}\exp\Big(-\frac{1}{2}\Big)\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\int\exp\Big(-\frac{1}{2}\beta_2^2(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}) + \beta_2(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y})\Big)d\beta_2\\
&=(2\pi)^{-\frac{n+1}{2}}\exp\Big(-\frac{1}{2}\Big)\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\\
&\quad(2\pi)^{\frac{1}{2}}\Big(\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)^{\frac{1}{2}}\exp\Big(\frac{1}{2}\frac{(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y})^2}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)\int N\Big(\beta_2;\frac{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}},\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)d\beta_2\\
&=(2\pi)^{-\frac{n}{2}}\exp\Big(-\frac{1}{2}\Big)\Big(\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)^{\frac{1}{2}}\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\exp\Big(\frac{1}{2}\frac{(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y})^\top(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y})}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)\\
&=(2\pi)^{-\frac{n}{2}}\exp\Big(-\frac{1}{2}\Big)\Big(\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)^{\frac{1}{2}}\Big(\frac{1}{n+1}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)
\end{aligned}
$$

$$
\begin{aligned}
&\exp\Big(\frac{1}{2}\frac{\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}+\frac{\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}+\frac{1}{2}\Big)\\
&\propto\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top(\boldsymbol{H}_n-\frac{\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}})\boldsymbol{y}+\frac{\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)\\
\boldsymbol{y}|H_1 &\sim MN\Big(\Big(\boldsymbol{H}_n-\frac{\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)^{-1}\frac{\boldsymbol{H}_n^\top\boldsymbol{x}}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}},\Big(\boldsymbol{H}_n-\frac{\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)^{-1}\Big)
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{y}|H_1 &\sim MN\Big(\boldsymbol{C}^{-1}\boldsymbol{u},\boldsymbol{C}^{-1}\Big)\\
&\text{where } \boldsymbol{C}=\boldsymbol{H}_n-\frac{\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\\
&\text{where } \boldsymbol{u}=\frac{\boldsymbol{H}_n^\top\boldsymbol{x}}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}
\end{aligned}
$$

Question 3.2 (7 points)

Find the marginal distributions $\pi(y_1,\dots,y_n|H_2)=\int f(y_1,\dots,y_n|\gamma_1,\gamma_2,\gamma_3)\pi(\gamma_1)\pi(\gamma_2)\pi(\gamma_3)d\gamma_1 d\gamma_2 d\gamma_3$

$$
\begin{aligned}
\pi(y_1,\dots,y_n|H_2) &= \int f(y_1,\dots,y_n|\gamma_1,\gamma_2,\gamma_3)\pi(\gamma_1)\pi(\gamma_2)\pi(\gamma_3)d\gamma_1 d\gamma_2 d\gamma_3\\
&\text{Integrate out } \gamma_1:\\
p(\boldsymbol{y}|\gamma_2,\gamma_3) &= \int f(y_1,\dots,y_n|\gamma_1,\gamma_2,\gamma_3)\pi(\gamma_1)d\gamma_1\\
&= \int N(\boldsymbol{y};\gamma_1\mathbf{1}_n+\gamma_2\boldsymbol{x}+\gamma_3\boldsymbol{x}\circ\boldsymbol{x},\boldsymbol{I}_n)N(\gamma_1;0,1)d\gamma_1\\
&\propto \int \exp\Big(-\frac{1}{2}(\boldsymbol{y}-\gamma_1\mathbf{1}_n-\gamma_2\boldsymbol{x}-\gamma_3\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{y}-\gamma_1\mathbf{1}_n-\gamma_2\boldsymbol{x}-\gamma_3\boldsymbol{x}\circ\boldsymbol{x})\Big)\exp\Big(-\frac{\gamma_1^2}{2}\Big)d\gamma_1\\
&= \int \exp\Big(-\frac{1}{2}\Big(\boldsymbol{y}^\top\boldsymbol{y}+n\gamma_1^2+\gamma_2^2\boldsymbol{x}^\top\boldsymbol{x}+\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{x}\circ\boldsymbol{x})\Big)\Big)\\
&\quad \exp\Big(\gamma_1\mathbf{1}_n^\top\boldsymbol{y}-\gamma_1\gamma_2\mathbf{1}_n^\top\boldsymbol{x}-\gamma_1\gamma_3\mathbf{1}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_2\boldsymbol{x}^\top\boldsymbol{y}-\gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{y})\Big)\exp\Big(-\frac{\gamma_1^2}{2}\Big)d\gamma_1\\
&= \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{y}\Big)\\
&\quad \exp\Big(-\frac{1}{2}\Big(\gamma_2^2\boldsymbol{x}^\top\boldsymbol{x}\Big)+\gamma_2\boldsymbol{x}^\top\boldsymbol{y}-\gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{x}\circ\boldsymbol{x})\Big)\\
&\quad \int \exp\Big(-\frac{n+1}{2}\gamma_1^2+\mathbf{1}_n^\top(\boldsymbol{y}-\gamma_2\boldsymbol{x}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x}))\gamma_1\Big)d\gamma_1\\
&= \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{y}\Big)\\
&\quad \exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{y}\Big)\\
&\quad \exp\Big(-\frac{1}{2}\Big(\gamma_2^2\boldsymbol{x}^\top\boldsymbol{x}\Big)+\gamma_2\boldsymbol{x}^\top\boldsymbol{y}-\gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{x}\circ\boldsymbol{x})\Big)\\
&\quad \exp\Big(\frac{1}{2}\frac{1}{n+1}(\boldsymbol{y}-\gamma_2\boldsymbol{x}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x}))^\top\mathbf{1}_n\mathbf{1}_n^\top(\boldsymbol{y}-\gamma_2\boldsymbol{x}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x}))\Big)
\end{aligned}
$$

$$
\begin{aligned}
&\int N\bigg(\gamma_1; \frac{1}{n+1}(\boldsymbol{y}-\gamma_2\boldsymbol{x}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})), \frac{1}{n+1}\bigg)d\gamma_1\\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{y}\bigg)\\
&\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{y}\bigg)\\
&\exp\bigg(-\frac{1}{2}\bigg(\gamma_2^2\boldsymbol{x}^\top\boldsymbol{x}\bigg)+\gamma_2\boldsymbol{x}^\top\boldsymbol{y}-\gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{x}\circ\boldsymbol{x})\bigg)\\
&\exp\bigg(\frac{1}{2}\frac{1}{n+1}(\boldsymbol{y}-\gamma_2\boldsymbol{x}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x}))^\top\mathbf{1}_n\mathbf{1}_n^\top(\boldsymbol{y}-\gamma_2\boldsymbol{x}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x}))\bigg)\\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{y}\bigg)\\
&\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{y}\bigg)\\
&\exp\bigg(-\frac{1}{2}\bigg(\gamma_2^2\boldsymbol{x}^\top\boldsymbol{x}\bigg)+\gamma_2\boldsymbol{x}^\top\boldsymbol{y}-\gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{x}\circ\boldsymbol{x})\bigg)\\
&\exp\bigg(\frac{1}{2}\frac{1}{n+1}\boldsymbol{y}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{y}+\frac{1}{2}\frac{1}{n+1}\gamma_2^2\boldsymbol{x}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{x}+\frac{1}{2}\frac{1}{n+1}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\mathbf{1}_n\mathbf{1}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})\bigg)\\
&\exp\bigg(-\frac{1}{n+1}\gamma_2\boldsymbol{x}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{y}+\frac{1}{n+1}\gamma_2\gamma_3\boldsymbol{x}^\top\mathbf{1}_n\mathbf{1}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})-\frac{1}{n+1}\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{y}\bigg)\\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top(\boldsymbol{I}_n-\frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{y}\bigg)\\
&\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{I}_n-\frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{I}_n-\frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{y}\bigg)\\
&\exp\bigg(-\frac{1}{2}\bigg(\gamma_2^2\boldsymbol{x}^\top(\boldsymbol{I}_n-\frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{x}\bigg)+\gamma_2\boldsymbol{x}^\top(\boldsymbol{I}_n-\frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{y}-\gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{I}_n-\frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top)(\boldsymbol{x}\circ\boldsymbol{x})\bigg)\\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\\
&\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\\
&\exp\bigg(-\frac{1}{2}\gamma_2^2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}+\gamma_2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_2\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})\bigg)\\
&\text{where } \boldsymbol{H}_n=\boldsymbol{I}_n-\frac{1}{n+1}\mathbf{1}_n\mathbf{1}_n^\top
\end{aligned}
$$

Integrate out γ_2 :

$$
\begin{aligned}
p(\boldsymbol{y}|\gamma_3)=&\int p(y|\gamma_2,\gamma_3)p(\gamma_2)d\gamma_2\\
=&\int p(y|\gamma_2,\gamma_3)N(\gamma_2;1,1)d\gamma_2\\
\propto&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\\
&\int\exp\bigg(-\frac{1}{2}\gamma_2^2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}+\gamma_2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_2\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})\bigg)\exp\bigg(-\frac{\gamma_2^2}{2}+\gamma_2\bigg)d\gamma_2\\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)
\end{aligned}
$$

$$
\begin{aligned}
&\int \exp\bigg(-\frac{1}{2}\gamma_2^2(1+\boldsymbol{x}^\top \boldsymbol{H}_n\boldsymbol{x}) + \gamma_2(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y} - \gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))\bigg) d\gamma_2 \\
\propto&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\bigg) \\
&\exp\bigg(\frac{1}{2}\frac{1}{(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x})}(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))^\top(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))\bigg) \\
&\int N\bigg(\gamma_2; \frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})), \frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\bigg)d\gamma_2 \\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\bigg) \\
&\exp\bigg(\frac{1}{2}\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))^\top(1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))\bigg) \\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\bigg) \\
&\exp\bigg(\frac{1}{2}\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\bigg(1+\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}+\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})\bigg)\bigg) \\
&\exp\bigg(\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\bigg(\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}-\gamma_3\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n^\top\boldsymbol{x}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}\bigg)\bigg) \\
\propto&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\bigg(\boldsymbol{H}_n-\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\bigg)\boldsymbol{y}+\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}\bigg) \\
&\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\bigg(\boldsymbol{H}_n-\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\bigg)(\boldsymbol{x}\circ\boldsymbol{x})+ \\
&\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\bigg(\boldsymbol{H}_n-\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\bigg)\boldsymbol{y}-\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\gamma_3\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})- \\
&\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n^\top\boldsymbol{x}\bigg) \\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{G}_n\boldsymbol{y}+\boldsymbol{y}^\top\boldsymbol{s}\bigg)\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\boldsymbol{y}-\gamma_3\boldsymbol{z}^\top\boldsymbol{y}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\bigg) \\
&\text{where } \boldsymbol{G}_n=\boldsymbol{H}_n-\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n \\
&\text{where } \boldsymbol{z}=\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) \\
&\text{where } \boldsymbol{s}=\frac{1}{1+\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}
\end{aligned}
$$

Integrate out γ_3 :

$$
\begin{aligned}
p(\boldsymbol{y}|H_2)=&\int p(\boldsymbol{y}|\gamma_3)p(\gamma_3)d\gamma_3 \\
=&\int p(\boldsymbol{y}|\gamma_3)N(\gamma_3;0,1)d\gamma_3 \\
\propto&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{G}_n\boldsymbol{y}+\boldsymbol{y}^\top\boldsymbol{s}\bigg) \\
&\int\exp\bigg(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\boldsymbol{y}-\gamma_3\boldsymbol{z}^\top\boldsymbol{y}-\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\bigg)\exp\bigg(-\frac{\gamma_3^2}{2}\bigg)d\gamma_3 \\
=&\exp\bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{G}_n\boldsymbol{y}+\boldsymbol{y}^\top\boldsymbol{s}\bigg)
\end{aligned}
$$

$$
\begin{aligned}
&\int \exp\Bigg(-\frac{1}{2}\gamma_3^2(1+(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})) + \gamma_3\Bigg((\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n\boldsymbol{y} - \boldsymbol{z}^\top\boldsymbol{y} - (\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg)\Bigg) d\gamma_3 \\
=&\exp\Bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{G}_n\boldsymbol{y} + \boldsymbol{y}^\top\boldsymbol{s}\Bigg) \\
&\exp\Bigg(\frac{1}{2}\frac{1}{1+(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}\Bigg((\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n\boldsymbol{y} - \boldsymbol{z}^\top\boldsymbol{y} - (\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg)^\top \Bigg((\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n\boldsymbol{y} - \boldsymbol{z}^\top\boldsymbol{y} - (\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg)\Bigg) \\
&\int N\Bigg(\gamma_3; \frac{1}{1+(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}\Bigg((\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n\boldsymbol{y} - \boldsymbol{z}^\top\boldsymbol{y} - (\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg), \frac{1}{1+(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}\Bigg) d\gamma_3 \\
=&\exp\Bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{G}_n\boldsymbol{y} + \boldsymbol{y}^\top\boldsymbol{s}\Bigg) \\
&\exp\Bigg(\frac{1}{\boldsymbol{2d}}\Bigg((\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n\boldsymbol{y} - \boldsymbol{z}^\top\boldsymbol{y} - (\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg)^\top \Bigg((\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n\boldsymbol{y} - \boldsymbol{z}^\top\boldsymbol{y} - (\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg)\Bigg) \\
=&\exp\Bigg(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{G}_n\boldsymbol{y} + \boldsymbol{y}^\top\boldsymbol{s}\Bigg) \\
&\exp\Bigg(\frac{1}{\boldsymbol{2d}}\Bigg(\boldsymbol{y}^\top\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\boldsymbol{y} + \boldsymbol{y}^\top\boldsymbol{z}\boldsymbol{z}^\top\boldsymbol{y} + \boldsymbol{s}^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg)\Bigg) \\
&\exp\Bigg(\frac{1}{\boldsymbol{d}}\Bigg(-\boldsymbol{y}^\top\boldsymbol{z}(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\boldsymbol{y} + \boldsymbol{y}^\top\boldsymbol{z}(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s} - \boldsymbol{y}^\top\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg)\Bigg) \\
\propto&\exp\Bigg(-\frac{1}{2}\boldsymbol{y}^\top\Bigg(\boldsymbol{G}_n - \frac{1}{d}\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n - \frac{1}{d}\boldsymbol{z}\boldsymbol{z}^\top + \frac{2}{d}\boldsymbol{z}(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\Bigg)\boldsymbol{y} \\
&+\boldsymbol{y}^\top\Bigg(\boldsymbol{s} + \frac{1}{\boldsymbol{d}}\boldsymbol{z}(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s} - \frac{1}{\boldsymbol{d}}\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}\Bigg)\Bigg)
\end{aligned}
$$

$$
\boldsymbol{y}|H_2 \sim MN\Bigg(\boldsymbol{A}^{-1}\boldsymbol{b}, \boldsymbol{A}^{-1}\Bigg)
$$

$$
\begin{aligned}
&\text{where } \boldsymbol{d} = 1 + (\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x}) \\
&\text{where } \boldsymbol{A} = \boldsymbol{G}_n - \frac{1}{d}\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n - \frac{1}{d}\boldsymbol{z}\boldsymbol{z}^\top + \frac{2}{d}\boldsymbol{z}(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n \\
&\text{where } \boldsymbol{b} = \boldsymbol{s} + \frac{1}{\boldsymbol{d}}\boldsymbol{z}(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s} - \frac{1}{\boldsymbol{d}}\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{s}
\end{aligned}
$$

Question 3.3 (8 points)

Write down the Bayes factor $B = f(y_1, \ldots, y_n|H_2)/f(y_1, \ldots, y_n|H_1)$ for comparing the two models and evaluate it for the given data set.

$$
\begin{aligned}
\boldsymbol{y}|H_1 \sim & MN\Big(\boldsymbol{C}^{-1}\boldsymbol{u}, \boldsymbol{C}^{-1}\Big) \\
& \text{where } \boldsymbol{H}_n = \boldsymbol{I}_n - \frac{1}{n+1}\boldsymbol{1}_n\boldsymbol{1}_n^\top \\
& \text{where } \boldsymbol{C} = \boldsymbol{H}_n - \frac{\boldsymbol{H}_n^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{H}_n}{1 + \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}} \\
& \text{where } \boldsymbol{u} = \frac{\boldsymbol{H}_n^\top \boldsymbol{x}}{1 + \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{y}|H_2 \sim & MN\Big(\boldsymbol{A}^{-1}\boldsymbol{b}, \boldsymbol{A}^{-1}\Big) \\
& \text{where } \boldsymbol{H}_n = \boldsymbol{I}_n - \frac{1}{n+1}\boldsymbol{1}_n\boldsymbol{1}_n^\top \\
& \text{where } \boldsymbol{G}_n = \boldsymbol{H}_n - \frac{1}{1 + \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}}\boldsymbol{H}_n^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{H}_n \\
& \text{where } \boldsymbol{z} = \frac{1}{1 + \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}}\boldsymbol{H}_n^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{H}_n(\boldsymbol{x} \circ \boldsymbol{x}) \\
& \text{where } \boldsymbol{s} = \frac{1}{1 + \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}}\boldsymbol{H}_n^\top \boldsymbol{x} \\
& \text{where } \boldsymbol{d} = 1 + (\boldsymbol{x} \circ \boldsymbol{x})^\top \boldsymbol{G}_n(\boldsymbol{x} \circ \boldsymbol{x}) \\
& \text{where } \boldsymbol{A} = \boldsymbol{G}_n - \frac{1}{d}\boldsymbol{G}_n^\top(\boldsymbol{x} \circ \boldsymbol{x})(\boldsymbol{x} \circ \boldsymbol{x})^\top \boldsymbol{G}_n - \frac{1}{d}\boldsymbol{z}\boldsymbol{z}^\top + \frac{2}{d}\boldsymbol{z}(\boldsymbol{x} \circ \boldsymbol{x})^\top \boldsymbol{G}_n \\
& \text{where } \boldsymbol{b} = \boldsymbol{s} + \frac{1}{\boldsymbol{d}}\boldsymbol{z}(\boldsymbol{x} \circ \boldsymbol{x})^\top \boldsymbol{s} - \frac{1}{\boldsymbol{d}}\boldsymbol{G}_n^\top(\boldsymbol{x} \circ \boldsymbol{x})(\boldsymbol{x} \circ \boldsymbol{x})^\top \boldsymbol{s}
\end{aligned}
$$

$$
\begin{aligned}
B =& \frac{(2\pi)^{-\frac{n}{2}}|\boldsymbol{A}|^{1/2}\exp\Big(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}^{-1}\boldsymbol{b})^\top \boldsymbol{A}(\boldsymbol{y} - \boldsymbol{A}^{-1}\boldsymbol{b})\Big)}{(2\pi)^{-\frac{n}{2}}|\boldsymbol{C}|^{1/2}\exp\Big(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{C}^{-1}\boldsymbol{u})^\top \boldsymbol{C}(\boldsymbol{y} - \boldsymbol{C}^{-1}\boldsymbol{u})\Big)} \\
=& \frac{|\boldsymbol{A}|^{1/2}\exp\Big(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}^{-1}\boldsymbol{b})^\top \boldsymbol{A}(\boldsymbol{y} - \boldsymbol{A}^{-1}\boldsymbol{b})\Big)}{|\boldsymbol{C}|^{1/2}\exp\Big(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{C}^{-1}\boldsymbol{u})^\top \boldsymbol{C}(\boldsymbol{y} - \boldsymbol{C}^{-1}\boldsymbol{u})\Big)}
\end{aligned}
$$

```r
# load the data
x <- c(-1.9 , -0.39 , 0.79 , -0.20 , 0.42 , -0.35 , 0.67 , 0.63 , -0.024 , 1.2)
y <- c(-1.7 ,-0.23 , 0.50 , -0.66 , 1.97 , 0.10 , 0.60 , 1.13 , -0.943 , 2.6)

# data summary and calc needed for BF
n <- length(x)
x2 <- x^2
one <- rep(1,n)
```

```r
I <- diag(n)

# H1 stats
Hn1 <- I - (1/(n+1)) *one %*% t(one)
C <- Hn1 - (t(Hn1) %*% x %*% t(x) %*% Hn1)/ (1 + t(x) %*% Hn1 %*% x)[1]
u <- (t(Hn1) %*% x) / (1 + t(x) %*% Hn1 %*% x)[1]

# H2 stats
Hn2 <- I - (1/(n+1)) *one %*% t(one)
G <- Hn2 - (t(Hn2) %*% x %*% t(x) %*% Hn2) / (1 + t(x)  %*% Hn2 %*% x)[1]
z <- (t(Hn2) %*% x %*% t(x) %*% Hn2 %*% x2) / (1 + t(x)  %*% Hn2 %*% x)[1]
s <- (t(Hn2) %*% x) / (1 + t(x)  %*% Hn2 %*% x)[1]
d <- (1 + t(x2) %*% G %*% x2)[1]
A <- G - (t(G) %*% x2 %*% t(x2) %*% G)/d - (z %*% t(z))/d + (2/d)* (z %*% t(x2) %*% G)
b <- s + (z %*% t(x2) %*% s)/d - (t(G) %*% x2 %*% t(x2) %*% s ) / d

# bayesfactor
B <- (det(A)^(1/2) * exp( -(1/2) * t(y - solve(A) %*% b) %*% A %*% (y - solve(A) %*%
↪  b))[1]) /
  (det(C)^(1/2) * exp( -(1/2) * t(y - solve(C) %*% u) %*% C %*% (y - solve(C) %*% u))[1])
print(paste("B = ",round(B,5)))
```

```
## [1] "B =  0.42017"
```

The calculated Bayes Factor is 0.42017 this indicates that model 1 H_1 may be a better model for the data.

Question 3.4 (10 points)

We now replace the prior distribution by improper constant priors: $\pi(\beta_1) = \pi(\beta_2) = c_1$ in model H_1 and $\pi(\gamma_1) = \pi(\gamma_2) = \pi(\gamma_3) = c_2$ in model H_2. We can still obtain the marginal distributions and define a Bayes factor. Show that the value of the Bayes factor depends on both c_1 and c_2.

$$\pi(y_1, \dots, y_n | H_1) = \int f(y_1, \dots, y_n | \beta_1, \beta_2)\pi(\beta_1)\pi(\beta_2) d\beta_1 d\beta_2$$

Integrate out β_1 :

$$\begin{aligned}
p(\boldsymbol{y}|\beta_2) =& \int f(y_1, \dots, y_n | \beta_1, \beta_2)\pi(\beta_1) d\beta_1 \\
=& \int N(y_1, \dots, y_n; \beta_1 1_n + \boldsymbol{x}\beta_2, I_n) c_1 d\beta_1 \\
=& c_1 \int (2\pi)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2}(\boldsymbol{y} - \beta_1 \mathbf{1}_n - \boldsymbol{x}\beta_2)^\top (\boldsymbol{y} - \beta_1 \mathbf{1}_n - \boldsymbol{x}\beta_2) \Big) d\beta_1 \\
=& c_1 (2\pi)^{-\frac{n}{2}} \int \exp\Big(-\frac{1}{2}(\boldsymbol{y} - \beta_1 \mathbf{1}_n - \boldsymbol{x}\beta_2)^\top (\boldsymbol{y} - \beta_1 \mathbf{1}_n - \boldsymbol{x}\beta_2) \Big) d\beta_1 \\
=& c_1 (2\pi)^{-\frac{n}{2}} \int \exp\Big(-\frac{1}{2}(\boldsymbol{y}\boldsymbol{y}^\top - n\beta_1^2 - \beta_2^2 \boldsymbol{x}^\top \boldsymbol{x} \Big) \exp\Big(\beta_1 \mathbf{1}_n^\top \boldsymbol{y} - \beta_1\beta_2 \mathbf{1}_n^\top \boldsymbol{x} - \beta_2 \boldsymbol{x}^\top \boldsymbol{y} \Big) d\beta_1 \\
=& c_1 (2\pi)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2}(\boldsymbol{y}\boldsymbol{y}^\top) \Big) \exp\Big(-\frac{1}{2}\beta_2^2 \boldsymbol{x}^\top \boldsymbol{x} + \beta_2 \boldsymbol{x}^\top \boldsymbol{y} \Big) \int \exp\Big(-\frac{n}{2}\beta_1^2 + \Big(\mathbf{1}_n^\top \boldsymbol{y} - \beta_2 \mathbf{1}_n^\top \boldsymbol{x} \Big) \beta_1 \Big) d\beta_1 \\
\propto& c_1 \exp\Big(-\frac{1}{2}(\boldsymbol{y}\boldsymbol{y}^\top) \Big) \exp\Big(-\frac{1}{2}\beta_2^2 \boldsymbol{x}^\top \boldsymbol{x} + \beta_2 \boldsymbol{x}^\top \boldsymbol{y} \Big) \\
& \exp\Big(\frac{1}{2n}(\boldsymbol{y} - \beta_2 \boldsymbol{x})^\top \mathbf{1}_n \mathbf{1}_n^\top (\boldsymbol{y} - \beta_2 \boldsymbol{x}) \Big) \int N\Big(\beta_1; \frac{1}{n}\Big(\mathbf{1}_n^\top \boldsymbol{y} - \beta_2 \mathbf{1}_n^\top \boldsymbol{x} \Big), \frac{1}{n} \Big) d\beta_1 \\
=& c_1 \exp\Big(-\frac{1}{2}(\boldsymbol{y}\boldsymbol{y}^\top) \Big) \exp\Big(-\frac{1}{2}\beta_2^2 \boldsymbol{x}^\top \boldsymbol{x} + \beta_2 \boldsymbol{x}^\top \boldsymbol{y} \Big) \\
& \exp\Big(\frac{1}{2n}\boldsymbol{y}^\top \mathbf{1}_n \mathbf{1}_n^\top \boldsymbol{y} + \frac{1}{2n}\beta_2^2 \boldsymbol{x}^\top \mathbf{1}_n \mathbf{1}_n^\top \boldsymbol{x} - \frac{1}{n}\beta_2 \boldsymbol{x}^\top \mathbf{1}_n \mathbf{1}_n^\top \boldsymbol{y} \Big) \\
=& c_1 \exp\Big(-\frac{1}{2}\boldsymbol{y}(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n^\top)\boldsymbol{y}^\top \Big) \\
& \exp\Big(-\frac{1}{2}\beta_2^2 \boldsymbol{x}^\top (\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n^\top)\boldsymbol{x} + \beta_2 \boldsymbol{x}^\top (\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n^\top)\boldsymbol{y} \Big) \\
=& c_1 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{H}_n \boldsymbol{y} \Big) \exp\Big(-\frac{1}{2}\beta_2^2 \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x} + \beta_2 \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{y} \Big) \\
& \text{where } \boldsymbol{H}_n = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n^\top
\end{aligned}$$

Integrate out β_2 :

$$\begin{aligned}
p(\boldsymbol{y}|H_1) =& \int p(\boldsymbol{y}|\beta_2) p(\beta_2) d\beta_2 \\
=& \int p(\boldsymbol{y}|\beta_2) c_1 d\beta_2 \\
\propto& \int c_1^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{H}_n \boldsymbol{y} \Big) \exp\Big(-\frac{1}{2}\beta_2^2 \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x} + \beta_2 \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{y} \Big) d\beta_2 \\
=& c_1^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{H}_n \boldsymbol{y} \Big) \int \exp\Big(-\frac{1}{2}\beta_2^2 \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x} + \beta_2 \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{y} \Big) d\beta_2
\end{aligned}$$

$$
\begin{aligned}
&= c_1^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{H}_n \boldsymbol{y}\Big) \int \exp\Big(-\frac{1}{2}\beta_2^2 \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x} + \beta_2 \boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{y}\Big) d\beta_2 \\
&= c_1^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{H}_n \boldsymbol{y}\Big) \exp\Big(\frac{1}{2}\frac{\boldsymbol{y}^\top \boldsymbol{H}_n^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{y}}{\boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}}\Big) \int N\Big(\beta_2; \frac{\boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{y}}{\boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}}, \frac{1}{\boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}}\Big) d\beta_2 \\
&\propto c_1^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \Big(\boldsymbol{H}_n - \frac{1}{\boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}} \boldsymbol{H}_n^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{H}_n\Big)\boldsymbol{y}\Big) \\
&= c_1^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{C} \boldsymbol{y}\Big)
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{y}|H_1 &\sim c_1^2 MN\Big(\boldsymbol{0}_n, \boldsymbol{C}^{-1}\Big) \\
&\quad \text{where } \boldsymbol{C} = \boldsymbol{H}_n - \frac{1}{\boldsymbol{x}^\top \boldsymbol{H}_n \boldsymbol{x}} \boldsymbol{H}_n^\top \boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{H}_n
\end{aligned}
$$

$$
\pi(y_1, \ldots, y_n|H_2) = \int f(y_1, \ldots, y_n|\gamma_1, \gamma_2, \gamma_3)\pi(\gamma_1)\pi(\gamma_2)\pi(\gamma_3) d\gamma_1 d\gamma_2 d\gamma_3
$$

Integrate out γ_1 :

$$
\begin{aligned}
p(\boldsymbol{y}|\gamma_2, \gamma_3) &= \int f(y_1, \ldots, y_n|\gamma_1, \gamma_2, \gamma_3)\pi(\gamma_1) d\gamma_1 \\
&= \int N(\boldsymbol{y}; \gamma_1 \boldsymbol{1}_n + \gamma_2 \boldsymbol{x} + \gamma_3 \boldsymbol{x} \circ \boldsymbol{x}, \boldsymbol{I}_n) c_2 d\gamma_1 \\
&\propto \int c_2 \exp\Big(-\frac{1}{2}(\boldsymbol{y} - \gamma_1 \boldsymbol{1}_n - \gamma_2 \boldsymbol{x} - \gamma_3 \boldsymbol{x} \circ \boldsymbol{x})^\top (\boldsymbol{y} - \gamma_1 \boldsymbol{1}_n - \gamma_2 \boldsymbol{x} - \gamma_3 \boldsymbol{x} \circ \boldsymbol{x})\Big) d\gamma_1 \\
&= c_2 \int \exp\Big(-\frac{1}{2}\Big(\boldsymbol{y}^\top \boldsymbol{y} + n\gamma_1^2 + \gamma_2^2 \boldsymbol{x}^\top \boldsymbol{x} + \gamma_3^2 (\boldsymbol{x} \circ \boldsymbol{x})^\top (\boldsymbol{x} \circ \boldsymbol{x})\Big)\Big) \\
&\quad \exp\Big(\gamma_1 \boldsymbol{1}_n^\top \boldsymbol{y} - \gamma_1 \gamma_2 \boldsymbol{1}_n^\top \boldsymbol{x} - \gamma_1 \gamma_3 \boldsymbol{1}_n^\top (\boldsymbol{x} \circ \boldsymbol{x}) + \gamma_2 \boldsymbol{x}^\top \boldsymbol{y} - \gamma_2 \gamma_3 \boldsymbol{x}^\top (\boldsymbol{x} \circ \boldsymbol{x}) + \gamma_3 (\boldsymbol{x} \circ \boldsymbol{x})^\top \boldsymbol{y}\Big) d\gamma_1 \\
&= c_2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{y}\Big) \exp\Big(-\frac{1}{2}\gamma_3^2 (\boldsymbol{x} \circ \boldsymbol{x})^\top (\boldsymbol{x} \circ \boldsymbol{x}) + \gamma_3 (\boldsymbol{x} \circ \boldsymbol{x})^\top \boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\Big(\gamma_2^2 \boldsymbol{x}^\top \boldsymbol{x}\Big) + \gamma_2 \boldsymbol{x}^\top \boldsymbol{y} - \gamma_2 \gamma_3 \boldsymbol{x}^\top (\boldsymbol{x} \circ \boldsymbol{x})\Big) \\
&\quad \int \exp\Big(-\frac{n}{2}\gamma_1^2 + \boldsymbol{1}_n^\top (\boldsymbol{y} - \gamma_2 \boldsymbol{x} - \gamma_3 (\boldsymbol{x} \circ \boldsymbol{x}))\gamma_1\Big) d\gamma_1 \\
&= c_2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\gamma_3^2 (\boldsymbol{x} \circ \boldsymbol{x})^\top (\boldsymbol{x} \circ \boldsymbol{x}) + \gamma_3 (\boldsymbol{x} \circ \boldsymbol{x})^\top \boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\Big(\gamma_2^2 \boldsymbol{x}^\top \boldsymbol{x}\Big) + \gamma_2 \boldsymbol{x}^\top \boldsymbol{y} - \gamma_2 \gamma_3 \boldsymbol{x}^\top (\boldsymbol{x} \circ \boldsymbol{x})\Big) \\
&\quad \exp\Big(\frac{1}{2}\frac{1}{n}(\boldsymbol{y} - \gamma_2 \boldsymbol{x} - \gamma_3 (\boldsymbol{x} \circ \boldsymbol{x}))^\top \boldsymbol{1}_n \boldsymbol{1}_n^\top (\boldsymbol{y} - \gamma_2 \boldsymbol{x} - \gamma_3 (\boldsymbol{x} \circ \boldsymbol{x}))\Big) \\
&\quad \int N\Big(\gamma_1; \frac{1}{n}(\boldsymbol{y} - \gamma_2 \boldsymbol{x} - \gamma_3 (\boldsymbol{x} \circ \boldsymbol{x})), \frac{1}{n}\Big) d\gamma_1
\end{aligned}
$$

$$
\begin{aligned}
&= c_2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\Big(\gamma_2^2\boldsymbol{x}^\top\boldsymbol{x}\Big) + \gamma_2\boldsymbol{x}^\top\boldsymbol{y} - \gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{x}\circ\boldsymbol{x})\Big) \\
&\quad \exp\Big(\frac{1}{2n}(\boldsymbol{y} - \gamma_2\boldsymbol{x} - \gamma_3(\boldsymbol{x}\circ\boldsymbol{x}))^\top\mathbf{1}_n\mathbf{1}_n^\top(\boldsymbol{y} - \gamma_2\boldsymbol{x} - \gamma_3(\boldsymbol{x}\circ\boldsymbol{x}))\Big) \\
&= c_2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\Big(\gamma_2^2\boldsymbol{x}^\top\boldsymbol{x}\Big) + \gamma_2\boldsymbol{x}^\top\boldsymbol{y} - \gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{x}\circ\boldsymbol{x})\Big) \\
&\quad \exp\Big(\frac{1}{2n}\boldsymbol{y}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{y} + \frac{1}{2n}\gamma_2^2\boldsymbol{x}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{x} + \frac{1}{2n}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\mathbf{1}_n\mathbf{1}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})\Big) \\
&\quad \exp\Big(-\frac{1}{n}\gamma_2\boldsymbol{x}^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{y} + \frac{1}{n}\gamma_2\gamma_3\boldsymbol{x}^\top\mathbf{1}_n\mathbf{1}_n^\top(\boldsymbol{x}\circ\boldsymbol{x}) - \frac{1}{n}\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\mathbf{1}_n\mathbf{1}_n^\top\boldsymbol{y}\Big) \\
&= c_2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\Big(\gamma_2^2\boldsymbol{x}^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{x}\Big) + \gamma_2\boldsymbol{x}^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\boldsymbol{y} - \gamma_2\gamma_3\boldsymbol{x}^\top(\boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)(\boldsymbol{x}\circ\boldsymbol{x})\Big) \\
&= c_2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\Big) \\
&\quad \exp\Big(-\frac{1}{2}\gamma_2^2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x} + \gamma_2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y} - \gamma_2\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})\Big) \\
&\quad \text{where } \boldsymbol{H}_n = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top
\end{aligned}
$$

Integrate out γ_2 :

$$
\begin{aligned}
p(\boldsymbol{y}|\gamma_3) &= \int p(y|\gamma_2,\gamma_3)p(\gamma_2)d\gamma_2 \\
&= \int p(y|\gamma_2,\gamma_3)c_2 d\gamma_2 \\
&\propto c_2^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\Big) \\
&\quad \int \exp\Big(-\frac{1}{2}\gamma_2^2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x} + \gamma_2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y} - \gamma_2\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})\Big)d\gamma_2 \\
&= c_2^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{H}_n\boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n\boldsymbol{y}\Big) \\
&\quad \int \exp\Big(-\frac{1}{2}\gamma_2^2\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x} + \gamma_2(\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y} - \gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))\Big)d\gamma_2
\end{aligned}
$$

$$
\begin{aligned}
&= c_2^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{H}_n \boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{H}_n\boldsymbol{y}\Big)\\
&\quad \exp\Big(\frac{1}{2}\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}(\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))^\top(\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))\Big)\\
&\quad \int N\Big(\gamma_2;\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}(\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})),\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big)d\gamma_2\\
&= c_2^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{H}_n \boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{H}_n\boldsymbol{y}\Big)\\
&\quad \exp\Big(\frac{1}{2}\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}(\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))^\top(\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}-\gamma_3\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}))\Big)\\
&= c_2^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{H}_n \boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{H}_n\boldsymbol{y}\Big)\\
&\quad \exp\Big(\frac{1}{2}\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big(\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{y}+\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})\Big)\Big)\\
&\quad \exp\Big(\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\Big(-\gamma_3\boldsymbol{y}^\top\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n(\boldsymbol{x}\circ\boldsymbol{x})\Big)\Big)\\
&\propto c_2^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\Big(\boldsymbol{H}_n-\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\Big)\boldsymbol{y}\Big)\\
&\quad \exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\Big(\boldsymbol{H}_n-\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\Big)(\boldsymbol{x}\circ\boldsymbol{x})\\
&+\gamma_3\boldsymbol{y}^\top\Big(\boldsymbol{H}_n-\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\Big)^\top(\boldsymbol{x}\circ\boldsymbol{x})\Big)\\
&= c_2^2 \exp\Big(-\frac{1}{2}\boldsymbol{y}^\top \boldsymbol{G}_n \boldsymbol{y}\Big)\exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x}) + \gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top \boldsymbol{G}_n\boldsymbol{y}\Big)\\
&\quad \text{where } \boldsymbol{G}_n = \boldsymbol{H}_n-\frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n
\end{aligned}
$$

Integrate out γ_3 :

$$
\begin{aligned}
p(\boldsymbol{y}|H_2) &= \int p(\boldsymbol{y}|\gamma_3)p(\gamma_3)d\gamma_3\\
&= \int p(\boldsymbol{y}|\gamma_3)c_2 d\gamma_3\\
&\propto c_2^3\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{G}_n\boldsymbol{y}\Big)\int\exp\Big(-\frac{1}{2}\gamma_3^2(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})+\gamma_3(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\boldsymbol{y}\Big)d\gamma_3\\
&\propto c_2^3\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{G}_n\boldsymbol{y}\Big)\exp\Big(\frac{1}{2}\frac{1}{(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}\boldsymbol{y}^\top\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\boldsymbol{y}\Big)\\
&\quad \int N\Big(\gamma_3;\frac{1}{(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\boldsymbol{y},\frac{1}{(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}\Big)d\gamma_3\\
&= c_2^3\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\Big(\boldsymbol{G}_n-\frac{1}{(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\Big)\boldsymbol{y}\Big)\\
&= c_2^3\exp\Big(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{y}\Big)\\
\boldsymbol{y}|H_2 &\sim c_2^3 MN\Big(\boldsymbol{0},\boldsymbol{A}^{-1}\Big)\\
&\quad \text{where } \boldsymbol{A} = \boldsymbol{G}_n-\frac{1}{(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n
\end{aligned}
$$

Model 1 with improper constant prior c_1 :

$$\begin{aligned}\boldsymbol{y}|H_1 \sim & c_1^2 MN\left(\boldsymbol{0}_n, \boldsymbol{C}^{-1}\right)\\ & \text{where } \boldsymbol{H}_n = \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n^\top\\ & \text{where } \boldsymbol{C} = \boldsymbol{H}_n - \frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\end{aligned}$$

Model 2 with improper constant prior c_2 :

$$\begin{aligned}\boldsymbol{y}|H_2 \sim & c_2^3 MN\left(\boldsymbol{0}, \boldsymbol{A}^{-1}\right)\\ & \text{where } \boldsymbol{H}_n = \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}_n\boldsymbol{1}_n^\top\\ & \text{where } \boldsymbol{G}_n = \boldsymbol{H}_n - \frac{1}{\boldsymbol{x}^\top\boldsymbol{H}_n\boldsymbol{x}}\boldsymbol{H}_n^\top\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{H}_n\\ & \text{where } \boldsymbol{A} = \boldsymbol{G}_n - \frac{1}{(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n(\boldsymbol{x}\circ\boldsymbol{x})}\boldsymbol{G}_n^\top(\boldsymbol{x}\circ\boldsymbol{x})(\boldsymbol{x}\circ\boldsymbol{x})^\top\boldsymbol{G}_n\end{aligned}$$

Bayes Factor:

$$\begin{aligned}B = & \frac{c_2^3(2\pi)^{-\frac{n}{2}}|\boldsymbol{A}|^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{y}\right)}{c_1^2(2\pi)^{-\frac{n}{2}}|\boldsymbol{C}|^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{C}\boldsymbol{y}\right)}\\ = & \frac{c_2^3}{c_1^2}\cdot\frac{|\boldsymbol{A}|^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{y}\right)}{|\boldsymbol{C}|^{1/2}\exp\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{C}\boldsymbol{y}\right)}\end{aligned}$$

From the Bayes Factor we can see that it depends on the values of c_1 and c_2.