# Homework 6

Due by end of day Monday, 4/21

PUBH 8442: Bayes Decision Theory and Data Analysis

Include any code used to generate answers at the end of your assignment, and submit electronically via Canvas.

1. Mining Safety (*Modified from Exercise* 3.9 *in Carlin* & *Louis*)

   Consider the count of coal mining disasters in England from 1851 to 1962, available at http://ericfrazerlock.com/coal_data.txt. We will use a hierarchical Poisson model for the number of disasters $Y_i$ during year $i$, with a "changepoint" at year 1890:

$$Y_i \sim \begin{cases} \text{Poisson}(\theta) & \text{for } i = 1851, \ldots, 1890 \\ \text{Poisson}(\lambda) & \text{for } i = 1891, \ldots, 1962, \end{cases}$$

$$\theta \sim \text{Gamma}(1/2, b_1),\ \lambda \sim \text{Gamma}(1/2, b_2),\ \theta \text{ and } \lambda \text{ independent;}$$

$$b_1 \sim \text{Gamma}(1, 1),\ b_2 \sim \text{Gamma}(1, 1),\ b_1 \text{ and } b_2 \text{ independent.}$$

   (a) Let $\mathbf{Y} = (Y_{1851}, \ldots, Y_{1962})$. Derive the full conditional distributions for $\theta$, $\lambda$, $b_1$, $b_2$: $p(\theta \mid \lambda, b_1, b_2, \mathbf{Y})$, $p(\lambda \mid \theta, b_1, b_2, \mathbf{Y})$, $p(b_1 \mid \theta, \lambda, b_2, \mathbf{Y})$, and $p(b_2 \mid \theta, \lambda, b_1, \mathbf{Y})$. (Hint: they will all be Gamma distributions.)

   (b) Based on part (a), describe a Gibbs sampling algorithm to simulate from the joint posterior distribution $p(\theta, \lambda, b_1, b_2 \mid \mathbf{Y})$.

   (c) Implement your Gibbs sampler in R, and use histograms of the output to approximate the marginal posterior densities for $\theta$, $\lambda$, and $R = \theta/\lambda$.

   (d) Compute a 95% credible interval for $R = \theta/\lambda$, the ratio in mining disaster rate pre-1890 vs post-1890.

   NOTE: Here the second parameter of the Gamma distribution defines the *rate*, as in class. In the Carlin & Louis book, the second parameter defines the *scale*, which is $1/rate$.

a.) Let $k$ denote the number of years until the "changepoint" (here, $k = 40$). We first derive

$$\begin{aligned} P(\theta|b_1, \mathbf{y}) &\sim Gamma(1/2 + \sum_{i=1851}^{1890} y_i, k + b_1) \\ P(b_1|\theta, \mathbf{y}) &\sim Gamma(3/2, 1 + \theta) \\ P(\lambda|b_2, \mathbf{y}) &\sim Gamma(1/2 + \sum_{i=1891}^{1962} y_i, n - k + b_2) \\ P(b_2|\lambda, \mathbf{y}) &\sim Gamma(3/2, 1 + \lambda). \end{aligned}$$

b.) (a) Initialize the first element of $b_1$ and $b_2$ to $b_1^{(0)}$ and $b_1^{(1)}$.

   (b) For draws t=1, ..., N,

      i. Draw $\theta^{(t)} \sim Gamma(1/2 + \sum_{i=1851}^{1890} y_i, k + b_1^{(t-1)})$

ii. Draw $b_1^{(t)} \sim Gamma(3/2, 1 + \theta^{(t)})$

iii. Draw $\lambda^{(t)} \sim Gamma(1/2 + \sum_{i=1891}^{1962} y_i, n - k + b_2^{(t-1)})$

iv. Draw $b_2^{(t)} \sim Gamma(3/2, 1 + \lambda^{(t)})$

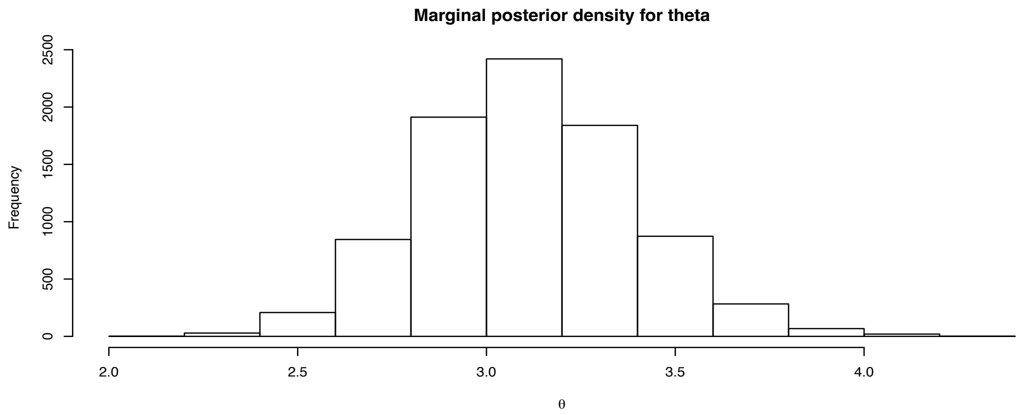
c.) Histograms for $\theta$, $\lambda$, and $R = \theta/\lambda$ are attached on the back page.

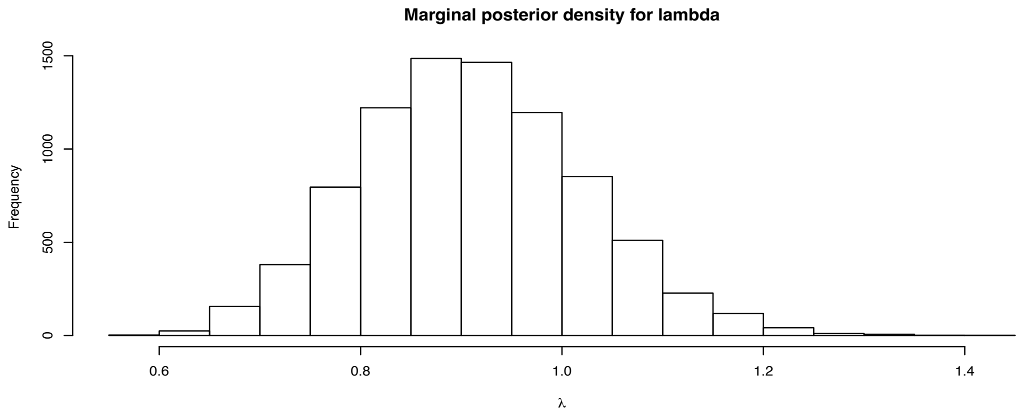
Following is the R program

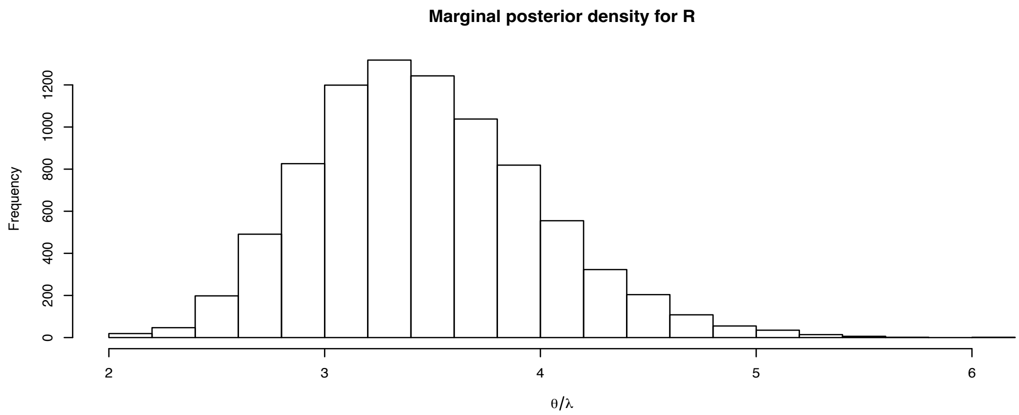
```
# Data
coal <- read.table(url("http://ericfrazerlock.com/coal_data.txt"))
k <- 40; n <- nrow(coal)
y1 <- sum(coal[1:k, 2])
y2 <- sum(coal[(k+1):n, 2])
n1 <- k; n2 <- n-k
# Parameters
b1.init <- b2.init <- 1
# Run Gibbs
MC.n <- 10000
burn.n <- 1500
theta <- rep(NA, MC.n)
lambda <- rep(NA, MC.n)
b1 <- b2 <- rep(NA, MC.n)
pre.b1 <- b1.init; pre.b2 <- b2.init

for (i in 1:MC.n)
{
  # Draw for theta
  theta[i] <- rgamma(1, shape = 0.5+y1, rate = n1+pre.b1)
  b1[i] <- rgamma(1, shape = 1.5, rate = 1+theta[i])
  pre.b1 = b1[i]
  # Draw for lambda
  lambda[i] <- rgamma(1, shape = 0.5+y2, rate = n2+pre.b2)
  b2[i] <- rgamma(1, shape = 1.5, rate = 1+lambda[i])
  pre.b2 = b2[i]
}
R.samp <- theta/lambda

windows(width = 8)
pdf(file = "Histogram.pdf", family = "serif", width = 8)
par(mfrow=c(3,1))
hist(theta[(burn.n+1):MC.n], main = "Marginal posterior density for theta",
     xlab = expression(theta))
hist(lambda[(burn.n+1):MC.n], main = "Marginal posterior density for lambda",
     xlab = expression(lambda))
hist(R.samp[(burn.n+1):MC.n], main = "Marginal posterior density for R",
     xlab = expression(theta/lambda))
```

d.) (2.529, 4.599) computed in R by

```
round(quantile(R.samp[(burn.n+1):MC.n], probs = c(.025, .975)), 3)
```


Marginal posterior density for theta
Frequency
0
500
1000
1500
2000
2500
2.0
2.5
3.0
3.5
4.0
θ


Marginal posterior density for lambda
Frequency
0
500
1000
1500
0.6
0.8
1.0
1.2
1.4
λ


Marginal posterior density for R
Frequency
0
200
400
600
800
1000
1200
2
3
4
5
6
θ/λ