Homework #5

Dustin Leatherman

October 18, 2020

Refer to the dataset ConcussionsByTeamAndYear.csv found under the Week Six course content of D2L. Let Y_i be the number of concussions from team $i = 1, ..., 32$. The model is $Y_i|\lambda_i \sim Poisson(\lambda_i)$ and the prior is $\lambda_i|\theta \sim Gamma(1, \theta)$ where $\theta \sim Gamma(0.1, 0.1)$.

1

Derive the full conditional distribution of λ_1.

$$\begin{aligned}
f(\lambda_1|Y,\theta) \propto & f(Y,|\lambda_1,\theta)f(\lambda_1|\theta) \\
=& f(Y|\lambda_1,\theta)f(\theta)f(\lambda_1|\theta) \\
=& \frac{e^{-\lambda_1}\lambda_1^Y}{Y!} \cdot \theta^{0.1-1}e^{-0.1\theta} \cdot \theta e^{-\theta\lambda_1} \\
\propto & \theta^{0.1}\lambda^Y e^{-\lambda_1 - 0.1\theta - \theta\lambda_1}
\end{aligned} \tag{1}$$

When holding everything but λ_1, then

$$f(\lambda_1|Y,\theta) \propto \lambda^{Y+1}e^{-\lambda_1(\theta+1)} \sim Gamma(Y+1, \theta+1)$$

2

Derive the full conditional distribution of θ.

Following the result in (a) and holding θ fixed,

$$\begin{aligned}
P(\theta|Y,\lambda) \propto & p(Y|\lambda,\theta)p(\lambda|\theta)p(\theta) \\
=& \prod_{i=1}^{n}\left[\frac{\lambda_i^{y_i}e^{-\lambda_i}}{y_i!}\right]\prod_{i=1}^{n}\left[\theta e^{-\theta\lambda_i}\right]\frac{0.1^{0.1}}{\Gamma(0.1)}\theta^{0.1-1}e^{-0.1\theta} \\
\propto & \theta^n e^{-\theta\sum_{i=1}^{n}\lambda_i}\theta^{0.1-1}e^{-0.1\theta} \\
=& \theta^{n+0.1-1}e^{-(0.1+\sum_{i=1}^{n}\lambda_i)\theta}
\end{aligned} \tag{2}$$

$$\theta|Y,\lambda \sim Gamma(n+0.1, 0.1+\sum_{i=1}^{n}\lambda_i)$$

3

Write Gibbs sampling code to draw samples from the joint distribution of $(\lambda_1, ..., \lambda_{32}, \theta)$.

```r
num.teams <- 32
num.params <- num.teams + 1 # for theta
n.iters <- 30000

res.2012 <- res.2013 <- matrix(0,n.iters, num.params)
lambda.names <- sapply(1:num.teams, function(x) paste0("lambda_", x))
colnames(res.2012) <- colnames(res.2013) <- c(lambda.names,"theta")

# initial values for first row
lambdas.init <- concussions %>%
  mutate(mean = (X2012 + X2013) / 2)

theta <- 0.1
lambdas <- lambdas.init
res.2012[1,] <- res.2013[1,] <- c(lambdas.init$mean, theta)

# The loop to perform Gibbs sampling
for(i in 2:n.iters){
  # Random draw from full conditional distribution of lambdas
  lambdas.2012 <- rgamma(num.teams, lambdas.init$X2012 + 1, theta + 1)
  lambdas.2013 <- rgamma(num.teams, lambdas.init$X2013 + 1, theta + 1)

  # Random draw from full conditional distribution of theta
  theta.2012 <- rgamma(1, n.iters + 0.1, sum(lambdas.2012) + 0.1)
  theta.2013 <- rgamma(1, n.iters + 0.1, sum(lambdas.2013) + 0.1)

  res.2012[i,] <- c(lambdas.2012, theta.2012)
  res.2013[i,] <- c(lambdas.2013, theta.2013)
}
res.2012 <- as.data.frame(res.2012)
res.2013 <- as.data.frame(res.2013)
```

4

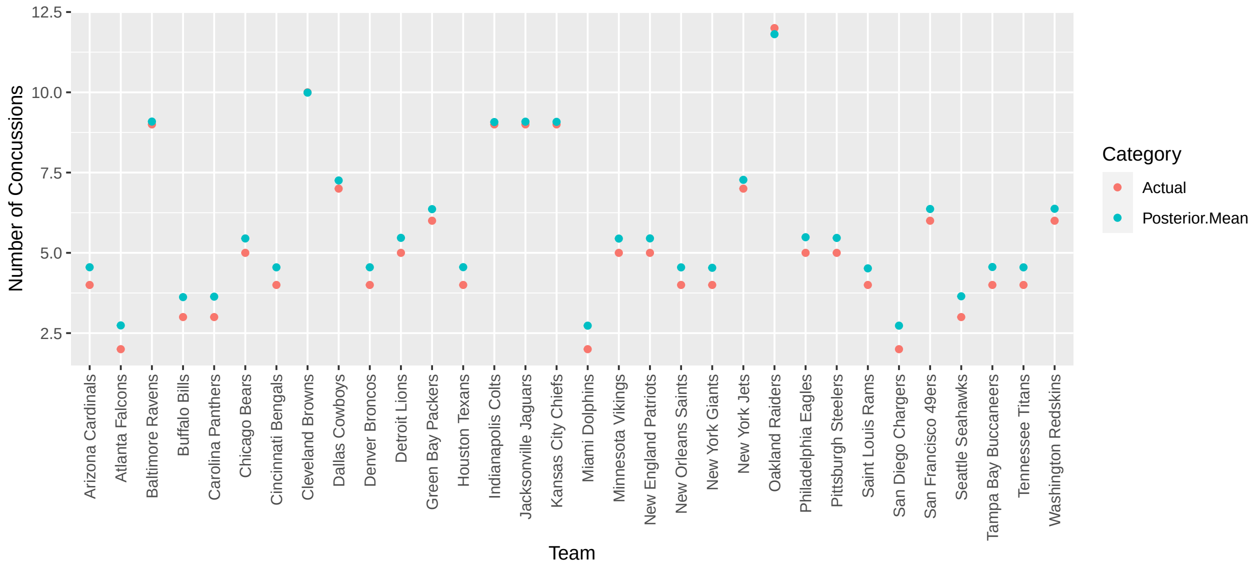
Show trace plots of the samples for λ_1 and θ. (These plots of the value of the parameter on the y-axis and the iteration number on the x-axis).

5

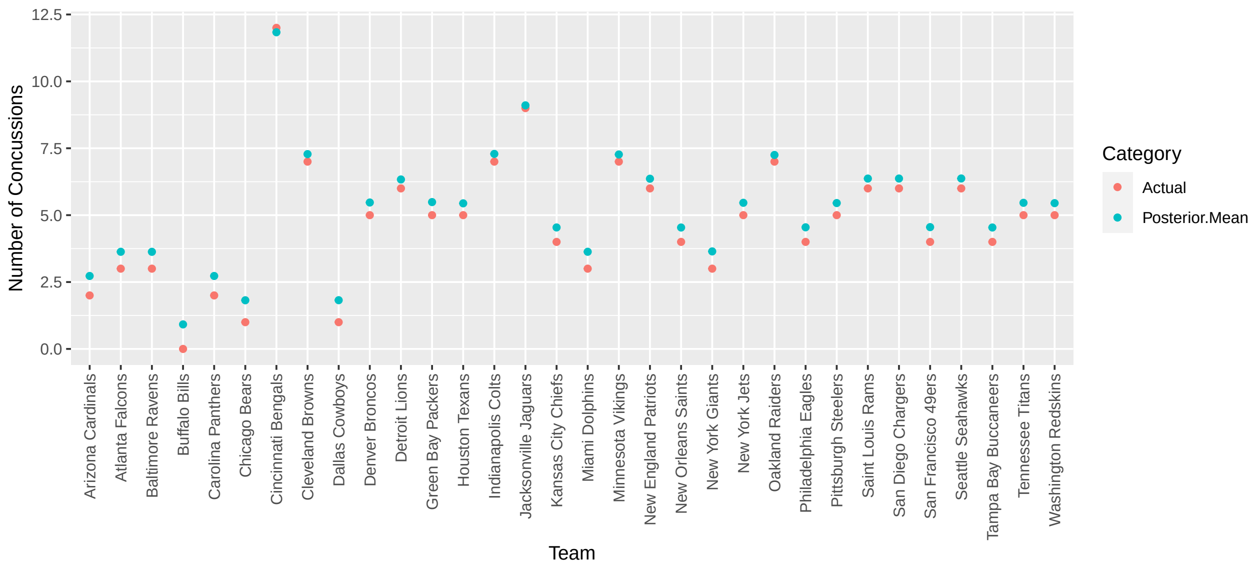
Plot the estimated posterior mean of λ_i versus Y_i and comment on whether the code is returning reasonable estimates

```r
lambdas.mean.2012 <- colMeans(res.2012) %>% head(-1)
concussions %>%
  bind_cols(Posterior.Mean = lambdas.mean.2012) %>%
  select(X, Actual = X2012, Posterior.Mean) %>%
  pivot_longer(-X, names_to = "Category", values_to = "value") %>%
  mutate(X = str_remove(X, "\xa0")) %>%
  ggplot(aes(x = X, color = Category, y = value)) +
  geom_point() +
```

```r
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  labs(x = "Team", y = "Number of Concussions")
```




```r
lambdas.mean.2013 <- colMeans(res.2013) %>% head(-1)
concussions %>%
  bind_cols(Posterior.Mean = lambdas.mean.2013) %>%
  select(X, Actual = X2013, Posterior.Mean) %>%
  pivot_longer(-X, names_to = "Category", values_to = "value") %>%
  mutate(X = str_remove(X, "\xa0")) %>%
  ggplot(aes(x = X, color = Category, y = value)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
  labs(x = "Team", y = "Number of Concussions")
```



The posterior mean is pretty close to the actual estimates for each year indicating that this is not a bad model.