Homework #4

Dustin Leatherman

October 12, 2020

1

Assume the likelihood $Y|\lambda \sim Poisson(\lambda)$ and prior $p(\lambda) = 1$ for all $\lambda > 0$. Show that this is an improper prior, i.e. that this density function is not a valid density function. Compute the posterior (for example, $\lambda|Y \sim Beta(3, Y + 1)$), and argue that the posterior distribution is proper for any value of Y.

$$\int_0^\infty 1 = \infty$$

Since 1 is uniform over an infinite domain.

$$f(\lambda|Y) = \frac{f(Y|\lambda) \cdot f(\lambda)}{\int f(Y|\lambda) \cdot f(\lambda) d\lambda} = \frac{\frac{\lambda^Y e^{-\lambda}}{Y!} \cdot 1}{\int 1 \cdot \lambda^Y e^{-\lambda} Y!} = \frac{\lambda^Y e^{-\lambda}}{Y!} \sim Poisson(\lambda)$$

Regardless of the value of Y, the posterior will still be a Poisson since λ is the hyper-parameter controlling the posterior distribution.

2

Say that $Y \sim Exponential(\lambda)$ so that the likelihood is $P(Y|\lambda) = \lambda exp(-\lambda Y)$. Assuming the prior is $\lambda \sim Gamma(a, b)$, find the posterior of λ.

$$\begin{aligned} f(\lambda|Y_1, ..., Y_n) =& \frac{f(Y_1, ..., Y_n|\lambda) f(\lambda)}{f(Y_1, ..., Y_n)} \\ =& \frac{f(Y_1, |\lambda) \cdot f(Y_2, |\lambda) \cdot ... \cdot f(Y_n, |\lambda) \cdot f(\lambda)}{f(Y_1, ..., Y_n)} \\ \propto& \lambda^n e^{-\lambda \sum_{i=1}^n y_i} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda} \\ \propto& \lambda^{a+n-1} e^{-\lambda(b + \sum_{i=1}^n y_i)} \\ \lambda|Y_1, ..., Y_n \sim& Gamma(a + n, b + \sum_{i=1}^n y_i) \end{aligned} \tag{1}$$

3

The file hurricanes.csv on the Week Five course content on D@l has the year and Saffir-Sampson intensity category of all Atlantic hurricanes that made landfall between 1990 and 2019. The counts are downloaded from http://www.aoml.noaa.gov/hrd/hurdat/. Break the years into two intervals: 1990-2004 and 2005-2019. Compute the posterior probability that the average number of category k storms per year has increased from 1990-2004 to 2005-2019. Describe and justify your method and carry it out separately for each k=1,. . . ,5, and for all storms combined.

The variable of interest is average number of k category storms per year. The end goal is to compare means between two intervals: 1990-2004, and 2005-2019. Since there are two samples that are being compared, a two-sample t-test can be used. For each category K, an indicator variable can be used to capture the count of each category. Then Gibbs sampling can be applied to draw samples from posterior distributions for each period of interest. Finally, the average value for each category and total for 2005-2019 that exceeds 1990-2004 categories can be computed to give the posterior probabilities that a given category is more likely for one period over another.

```r
hurricanes <- read.csv("~/Downloads/hurricanes.csv")
hurricanes$interval <- ifelse(hurricanes$Year > 2004, 2, 1)

data.pre <- hurricanes %>% filter(interval == 1)
data.post <- hurricanes %>% filter(interval == 2)

# Set the priors

mu0 <- 0
s20 <-1000
a <- 0.1
b <- 0.1

# Gibbs sampling: Iterations and Elements to save results

n.iters <- 30000


# Initial Values
n <- 20
m <- 20

gibbs <- function(sample1, sample2, n.iters, mu0, s20, n, m, s2, a, b) {
  res <- matrix(0,n.iters,4)
  colnames(res) <- c("mu.1","mu.2","sigma2","Delta")
  mu.1 <- mu.2 <- s2 <- A <- B <- 0

  mu.1 <- mean(sample1)
  mu.2 <- mean(sample2)
  s2 <- (var(sample1)+var(sample2))/2
  res[1,] <- c(mu.1, mu.2, s2 ,mu.1-mu.2)
  # The loop to perform Gibbs sampling
  for(i in 2:n.iters){
    # Random draw from full conditional distribution of mu.pre
    A <- sum(sample1)/s2+mu0/s20
    B <- n/s2+1/s20
    mu.1 <- rnorm(1,A/B,1/sqrt(B))

    # Random draw from full conditional distribution of mu.post
    A <- sum(sample2)/s2+mu0/s20
    B <- m/s2+1/s20
    mu.2 <- rnorm(1,A/B,1/sqrt(B))

    # Random draw from full conditional distribution of sigma2
    A <- n/2+m/2+a
    B <- sum((sample1-mu.1)^2)/2+sum((sample2-mu.2)^2)/2+b
```

```r
    s2 <- 1/rgamma(1, A, B)

    res[i,] <- c(mu.1,mu.2,s2,mu.1-mu.2)
  }
  return(as.data.frame(res))
}
res.total <- gibbs(data.pre$Category, data.post$Category, n.iters, mu0, s20, n, m, s2, a, b)
for(i in sort(unique(hurricanes$Category))) {
  mycol <- str_c("k",i)

  somecol <- gibbs(
        ifelse(data.pre$Category == i, 1, 0),
        ifelse(data.post$Category == i, 1, 0),
        n.iters,
        mu0,
        s20,
        n,
        m,
        s2,
        a,
        b)
  assign(mycol, somecol)
}
mean(res.total$mu.2 > res.total$mu.1)
```

```
## [1] 0.7216
```

```r
mean(k1$mu.2 > k1$mu.1)
```

```
## [1] 0.6966333
```

```r
mean(k2$mu.2 > k2$mu.1)
```

```
## [1] 0.4221333
```

```r
mean(k3$mu.2 > k3$mu.1)
```

```
## [1] 0.2162333
```

```r
mean(k4$mu.2 > k4$mu.1)
```

```
## [1] 0.8652667
```

```r
mean(k5$mu.2 > k5$mu.1)
```

```
## [1] 0.8097667
```

The probability that the average Saffir-Sampson rating for 2005-2019 exceeds the Saffir-Sampson rating for 1990-2004 for all storms and each category is listed below.

Category	P(mu.2 > mu.1)
1	0.6966333
2	0.4221333
3	0.2162333
4	0.8652667
5	0.8097667
Total	0.7216

It appears that 2005-2019 is more likely to experience Category 1, 4, 5, and Total storms than 1990-2004.