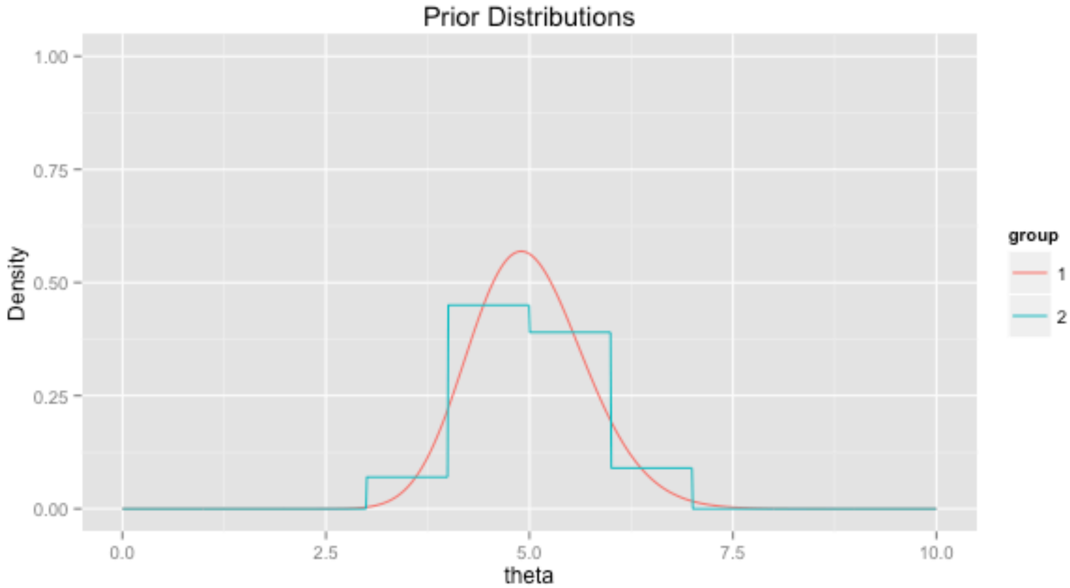
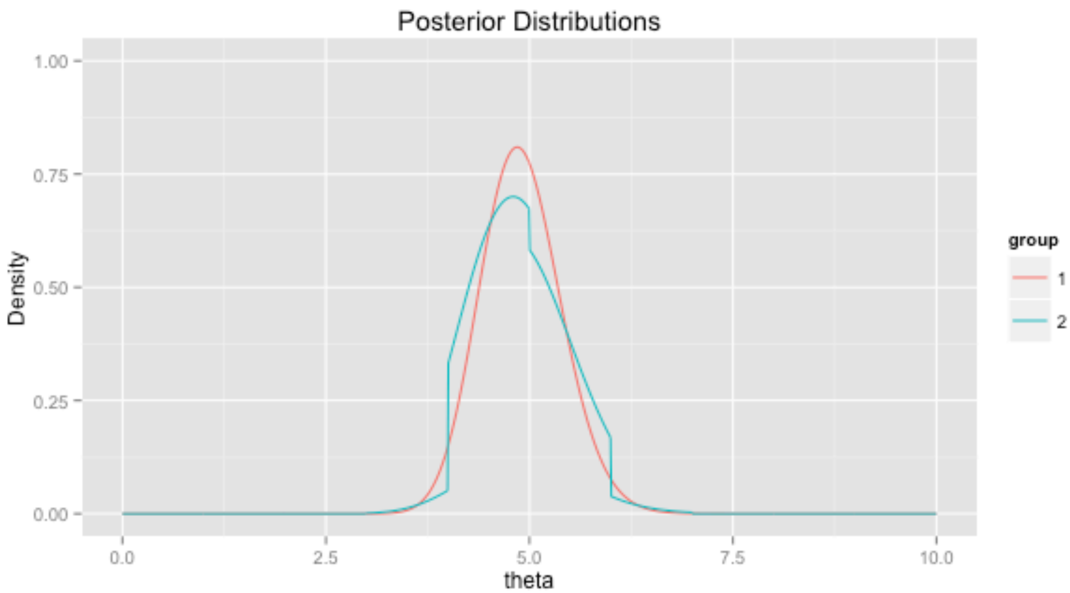
Kelvin Niu

Lab 2

<u>Plots</u>





<u>Equations for distributions</u>

Prior 1: θ ~ Gamma(shape = 50, scale = 0.1)

Prior 2: $p(\theta) = \begin{cases} 0.07, if\ \theta \in (3,4] \\ 0.45, if\ \theta \in (4,5] \\ 0.39, if\ \theta \in (5,6] \\ 0.09, if\ \theta \in (6,7] \end{cases}$

Posterior 1: $\theta \mid y_{1:10} \sim$ Gamma(shape = 98, scale = 0.05)

$$\text{Posterior 2: } p(\theta \mid y_{1:10}) = \begin{cases} \frac{0.07\theta^{\sum y_i} e^{-n\theta}}{c}, if\ \theta \in (3,4] \\ \frac{0.45\theta^{\sum y_i} e^{-n\theta}}{c}, if\ \theta \in (4,5] \\ \frac{0.39\theta^{\sum y_i} e^{-n\theta}}{c}, if\ \theta \in (5,6] \\ \frac{0.09\theta^{\sum y_i} e^{-n\theta}}{c}, if\ \theta \in (6,7] \end{cases}$$

where $c = \frac{\Gamma(\sum y_i+1)[0.07(pg(4)-pg(3))+0.45(pg(5)-pg(4))+0.39(pg(6)-pg(5))+0.09(pg(7)-pg(6))]}{n^{\sum y_i+1}}$

and $pg(x)$ is the probability density function for a Gamma(shape = 49, rate = 10).

95% central credible interval

The 95% central credible interval is calculated from the 2.5th and 97.5th percentile of the posterior distribution. This is easily calculated for posterior 1, since it is a gamma distribution (using the function qgamma): **[3.9781, 5.9166]**

However, since the quantile function is not explicitly known for posterior 2, we can estimate the 2.5th and 97.5th percentiles by testing different limits for integrating the posterior. We can compute the integral $\int_{-\infty}^{c} p(\theta \mid y_{1:10}) d\theta$ for slightly increasing values of c until the integral has a value of approximately 0.025 and 0.975. These values of c will make up the 95% central credible interval for posterior 2.

Code

```
library(ggplot2)
library(data.table)

y <- c(2, 1, 9, 4, 3, 3, 7, 7, 5, 7)
n <- length(y)
constant <- gamma(sum(y) + 1)*(0.07*(pgamma(4, sum(y) + 1, rate = n) -
pgamma(3, sum(y) + 1, rate = n)) + 0.45*(pgamma(5, sum(y) + 1, rate = n) -
pgamma(4, sum(y) + 1, rate = n)) + 0.39*(pgamma(6, sum(y) + 1, rate = n) -
pgamma(5, sum(y) + 1, rate = n)) + 0.09*(pgamma(7, sum(y) + 1, rate = n) -
pgamma(6, sum(y) + 1, rate = n)))/(n^(sum(y) + 1))

theta <- seq(0, 10, length.out = 1000)
prior_1 <- dgamma(theta, 50, scale = 0.1)
prior_2 <- (theta > 3 & theta <= 4)*0.07 + (theta > 4 & theta <= 5)*0.45 + (theta > 5 &
theta <= 6)*0.39 + (theta > 6 & theta <= 7)*0.09
post_1 <- dgamma(theta, 50 + sum(y), scale = 1/(1/0.1 + n))
post_2 <- (theta > 3 & theta <= 4)*0.07*(theta^sum(y))*exp(-n*theta)/constant +
(theta > 4 & theta <= 5)*0.45*(theta^sum(y))*exp(-n*theta)/constant + (theta > 5 &
theta <= 6)*0.39*(theta^sum(y))*exp(-n*theta)/constant + (theta > 6 & theta <=
7)*0.09*(theta^sum(y))*exp(-n*theta)/constant
```

```
#Get data into a data.table format for ggplot
data_prior_1 <- data.table(theta = theta, prior = prior_1, group = "1")
data_prior_2 <- data.table(theta = theta, prior = prior_2, group = "2")
data_prior <- rbindlist(list(data_prior_1, data_prior_2))
data_post_1 <- data.table(theta = theta, post = post_1, group = "1")
data_post_2 <- data.table(theta = theta, post = post_2, group = "2")
data_post <- rbindlist(list(data_post_1, data_post_2))

#Plot priors and posteriors
ggplot(data_prior, aes(x = theta, y = prior, group = group)) + geom_line(aes(color =
group)) + labs(title = "Prior Distributions") + scale_x_continuous(limits = c(0, 10)) +
scale_y_continuous(limits = c(0, 1)) + labs(y = "Density")
ggplot(data_post, aes(x = theta, y = post, group = group)) + geom_line(aes(color =
group)) + labs(title = "Posterior Distributions") + scale_x_continuous(limits = c(0,
10)) + scale_y_continuous(limits = c(0, 1)) + labs(y = "Density")

#Calculate 95% central credible interval
interval_1 <- qgamma(c(0.025, 0.975), 50 + sum(y), scale = 1/(1/0.1 + n))
```