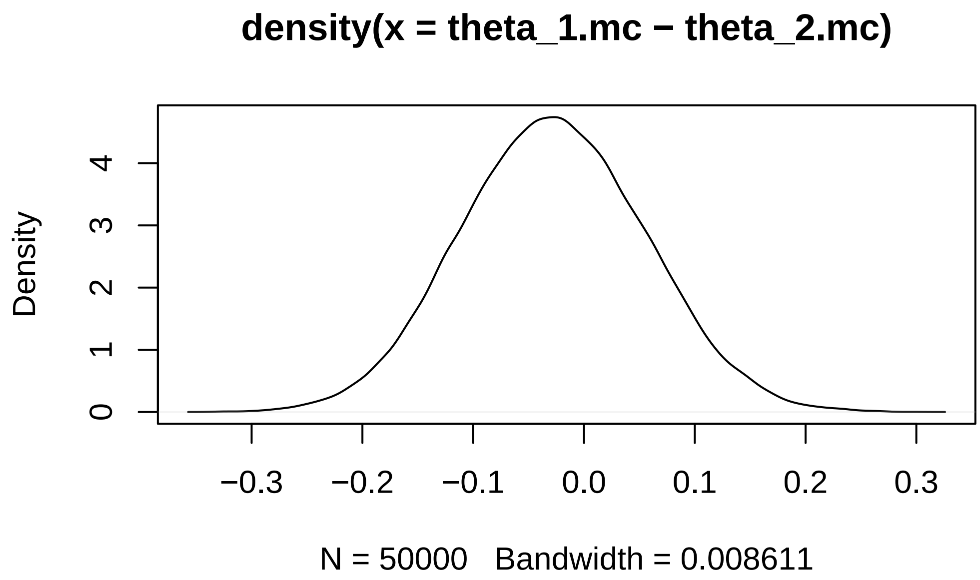
JC STAT 638 HW 4

Jack Cunningham

4.1)

Given that the prior for θ_2 is uniform, its posterior would be beta$(1+30, 1+50-30)$. Given the prior for θ_1 is uniform, its posterior would be beta(1+57,1+100-57).

```
set.seed(2)
a <- 1; b <- 1
sy1 <- 57; n1 <- 100
sy2 <- 30; n2 <- 50
theta_1.mc <- rbeta(50000, a + sy1, b + n1 - sy1)
theta_2.mc <- rbeta(50000, a + sy2, b + n2 - sy2)
plot(density(theta_1.mc - theta_2.mc))
```



$Pr(\theta_1 < \theta_2|\text{data,prior})$ is the observed proportion of times where $\theta_1 < \theta_2$ in our Monte Carlo sample.

```
t_1_less_t_2 = mean(theta_1.mc<theta_2.mc)
```

In this sample we have 0.63104.

4.2)

a)

```
set.seed(101)
y_a = c(12,9,12,14,13,13,15,8,15,6)
y_b = c(11,11,10,9,9,8,7,10,6,8,8,9,7)
sy1 <- sum(y_a); n_a <- 10
sy2 <- sum(y_b); n_b <- 13
a_a <- 120; b_a <- 10
a_b <- 12;  b_b <- 1

theta_a.mc <- rgamma(100000, a_a + sy1, b_a + n_a)
```

```
theta_b.mc <- rgamma(100000, a_b + sy2, b_b + n_b)
```

We have a posterior distribution of θ_a, θ_b as gamma(120+117, 10+10) and gamma(12+113,1+13) respectively.

So we use Monte Carlo sampling with $S = 100000$, and $Pr(\theta_B < \theta_A | y_a, y_b)$ is the percentage of times $\theta_b < \theta_a$ in our sample.

```
t_b_less_t_a = mean(theta_b.mc<theta_a.mc)
```

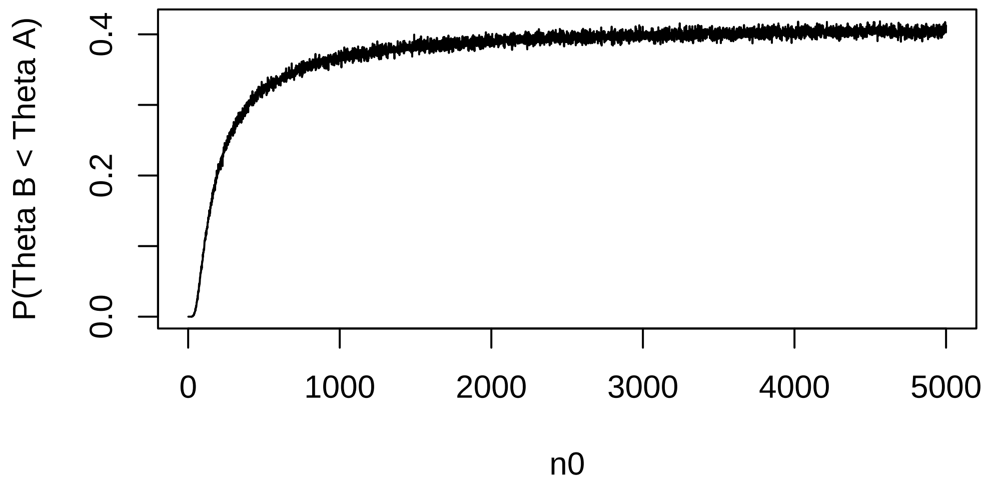
We get 0.99563 as the probability θ_b is less than θ_a given the observed data.

b)

```
set.seed(10)
n0 = seq(1,5000,1)
t_b_less_t_a_seq = rep(0,5000)
for(i in 1:length(t_b_less_t_a_seq)){
theta_a.mc <- rgamma(10000, a_a + sy1, b_a + n_a)
theta_b.mc <- rgamma(10000, 12*n0[i] + sy2,n0[i] + b_b)
t_b_less_t_a_seq[i] =  mean(theta_b.mc<theta_a.mc)
}
```

Plot:

```
plot(x = n0, y = t_b_less_t_a_seq, ylab = "P(Theta B < Theta A)", type = "l")
```



The conclusions clearly depend a lot on the selection of the prior, this makes it very important to understand how much weight we intend on placing onto our prior beliefs and checking to see how different prior beliefs would impact the posterior inference.

c)

a)

To obtain $Pr(\tilde{Y}_B < \tilde{Y}_A | y_A, y_B)$ we get the Monte Carlo sample of θ_B, θ_A and then we use those samples to perform a Monte Carlo sample using the poisson distribution given each θ sample.

```
set.seed(95)
y_a = c(12,9,12,14,13,13,15,8,15,6)
y_b = c(11,11,10,9,9,8,7,10,6,8,8,9,7)
sy1 <- sum(y_a); n_a <- 10
sy2 <- sum(y_b); n_b <- 13
a_a <- 120; b_a <- 10
a_b <- 12;  b_b <- 1

theta_a.mc <- rgamma(100000, a_a + sy1, b_a + n_a)
theta_b.mc <- rgamma(100000, a_b + sy2, b_b + n_b)
```

```
y_a.mc <- rpois(100000, theta_a.mc)
y_b.mc <- rpois(100000, theta_b.mc)
```

Now that we have Monte Carlo samples $\tilde{Y}_A$ and $\tilde{Y}_B$ we check the proportion of times we see $\tilde{Y}_{B(i)} < \tilde{Y}_{A(i)}$.

```
y_b_lt_y_a = mean(y_b.mc < y_a.mc)
```

We have $Pr(\tilde{Y}_B < \tilde{Y}_A | y_A, y_B) = 0.69997$.

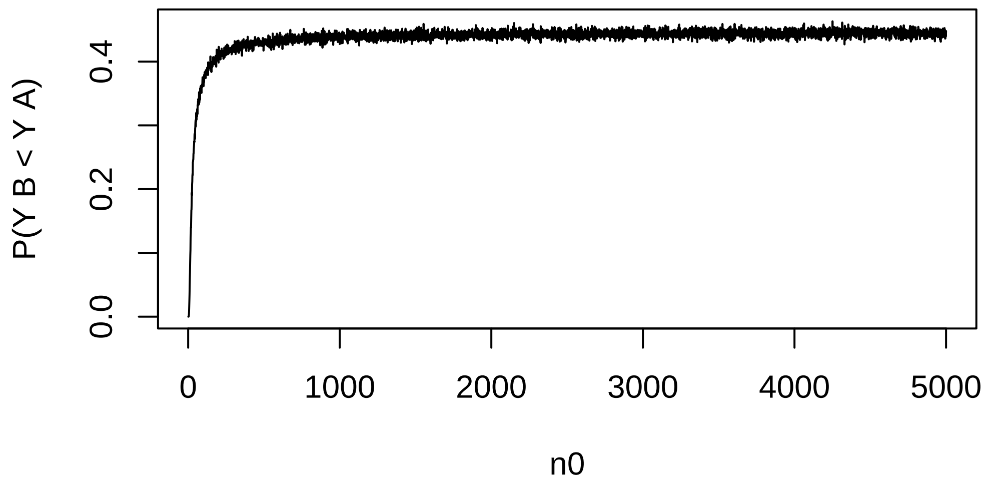
b)

Similarly to before, we perform the same procedure over varying choices of n_0.

```
set.seed(10)
n0 = seq(1,5000,1)
y_b_lt_y_a_seq = rep(0,5000)
for(i in 1:length(y_b_lt_y_a_seq)){
theta_a.mc <- rgamma(10000, a_a + sy1, b_a + n_a)
theta_b.mc <- rgamma(10000, 12*n0[i] + sy2,n0[i] + b_b)
y_a.mc <- rpois(10000, theta_a.mc)
y_b.mc <- rpois(10000, theta_b.mc)
y_b_lt_y_a_seq[i] = mean(y_b.mc < y_a.mc)
}
```

Below is a plot of the probabilities over varying n_0:

```
plot(x = n0, y = y_b_lt_y_a_seq, ylab = "P(Y B < Y A)", type = "l")
```



Similarly to before, the probability is very sensitive to the choice of n_0, as n_0 increases we are basically ignoring the observed data entirely and only relying on the prior. Its important that we choose a prior carefully.