

STA 601/360 Homework 4

Yifei Wang

25 September, 2019

1. Hoff problem 4.3

(a). Make a histogram of $t^{(s)}$ and compare to the observed value of this statistic. Based on this statistic, assess the fit of the Poisson model for these data.

We have

$$\theta_A \sim \text{gamma}(120, 10)$$

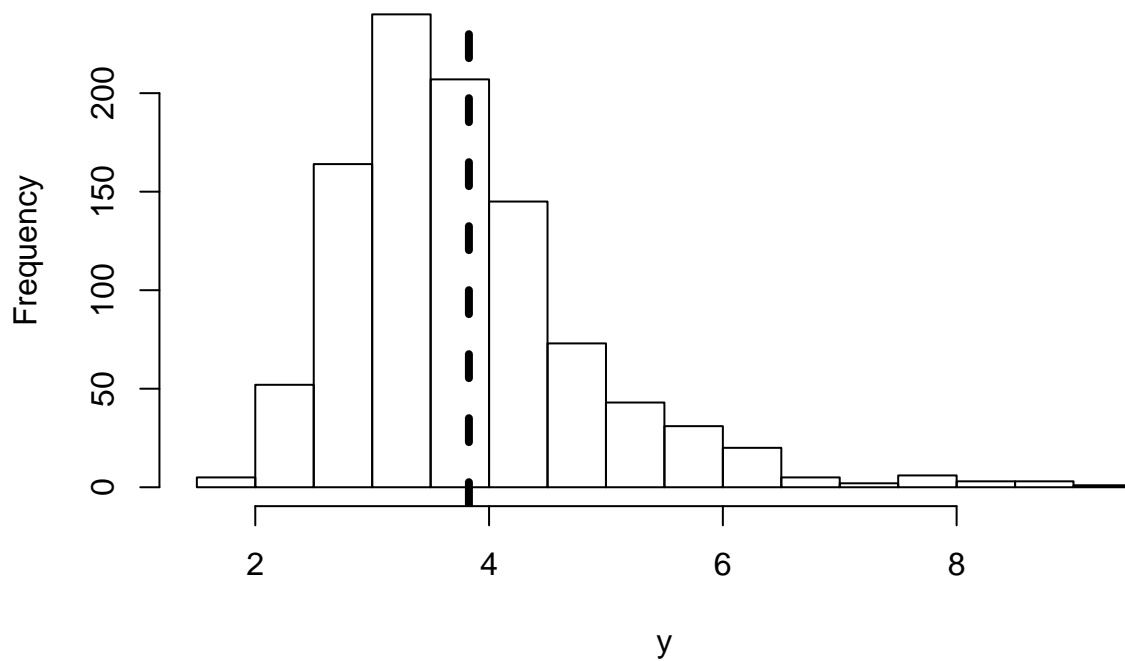
$$p(\theta_A | y_A) \sim \text{gamma}(120 + 117, 10 + 10)$$

```
y1 = c(12,9,12,14,13,13,15,8,15,6)
sy1 = sum(y1)
n1 = length(y1)
a1 = 120
b1 = 10
t1 = NULL

for (s in 1:1000) {
  theta1_s = rgamma(1, a1+sy1, b1+n1)
  y1_s = rpois(10, theta1_s)
  t1 = c(t1, mean(y1_s) / sd(y1_s))
}

hist(t1, main = "MC Distribution of Predicted Y", xlab = "y")
abline(v=mean(y1)/sd(y1), lty = 2, lwd = 4)
```

MC Distribution of Predicted Y



We can see from the histogram that this statistic of the data (thick dashed line) align with the distribution of it of the generated predicted samples. Thus, the Poisson model fit pretty good for group A in terms of this statistic.

(b). Repeat the above goodness of fit evaluation for the data in population B.

We have,

$$\theta_B \sim \text{gamma}(12, 1)$$

$$p(\theta_B \mid y_B) \sim \text{gamma}(12 + 113, 1 + 13)$$

```

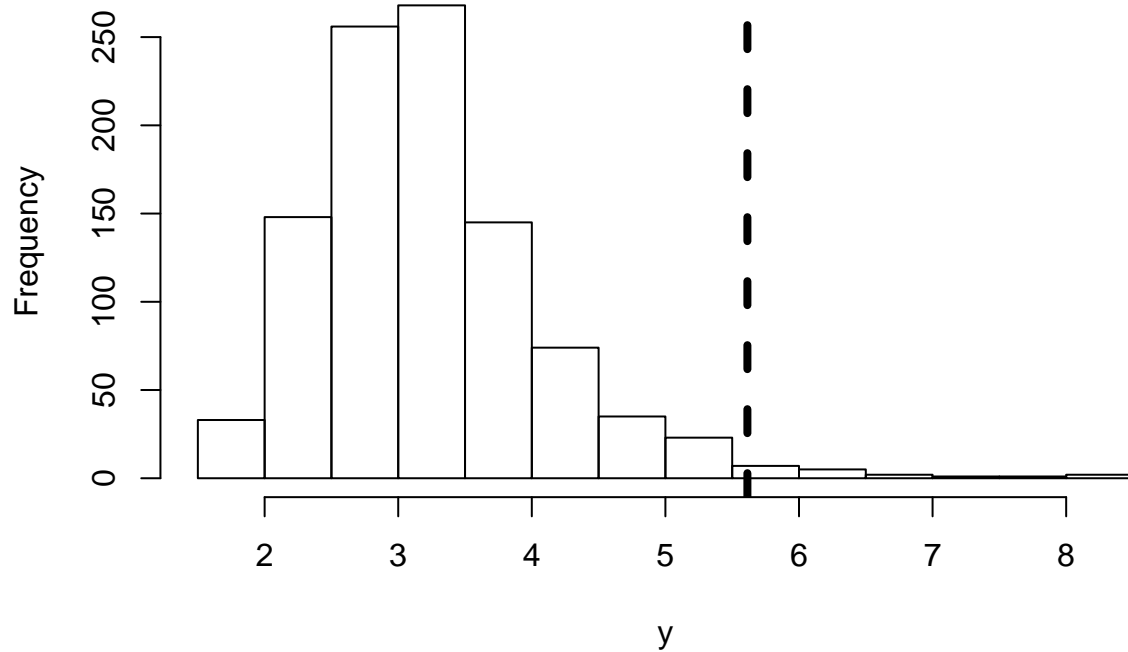
y2 = c(11,11,10,9,9,8,7,10,6,8,8,9,7)
sy2 = sum(y2)
n2 = length(y2)
a2 = 12
b2 = 1
t2 = NULL

for (s in 1:1000) {
  theta2_s = rgamma(1, a2+sy2, b2+n2)
  y2_s = rpois(10, theta2_s)
  t2 = c(t2, mean(y2_s) / sd(y2_s))
}

hist(t2, main = "MC Distribution of Predicted Y", xlab = "y")
abline(v=mean(y2)/sd(y2), lty = 2, lwd = 4)

```

MC Distribution of Predicted Y



We can see from the histogram that this statistic of the data (thick dashed line) lies almost outside of the distribution of it of the generated predicted samples. Thus, the Poisson model fit no so good for group B in terms of this statistic.

2. Hoff problem 4.7

(a). Sample at least 5,000 y values from the posterior predictive distribution.

```
Y = NULL
for (s in 1:5000) {
  prec = rgamma(1, 10, 2.5)
  sigma = sqrt(1 / prec)
  theta = rnorm(1, 4.1, sigma / sqrt(20))
  y = 0.31 * rnorm(1, theta, sigma) +
    0.46 * rnorm(1, 2*theta, 2*sigma) +
    0.23 * rnorm(1, 3*theta, 3*sigma)
  Y = c(Y, y)
}
```

(b). Form a 75% quantile-based confidence interval for a new value of Y .

```
(CI_quant = quantile(Y, probs = c(0.125, 0.875)))
```

```
##      12.5%      87.5%
## 7.122651 8.602160
```

(c). Form a 75% HPD region for a new Y :

```

# i).
dist_Y = density(Y)
Yx = dist_Y$x
Yy = dist_Y$y / sum(dist_Y$y)

# ii).
Yy_sort = sort(Yy, decreasing = TRUE)

# iii).
i = 1
while (sum(Yy_sort[1:i]) < 0.75) {
  i = i + 1
}

CI_HPD = c()
for (j in 1:length(Yx)) {
  if (any(Yy[j] == Yy_sort[1:i])) {
    CI_HPD = c(CI_HPD, Yx[j])
  }
}
c(min(CI_HPD), max(CI_HPD))

```

```
## [1] 7.128415 8.621894
```

The 75% HPD region looks almost the same as the 75% quantile-based confidence interval.

(d). Can you think of a physical justification for the mixture sampling distribution of Y ?

Yes. Suppose the whole population of the squash plants are sampled from three different kinds of squash plant, say A, B and C, from three different farms. Some farms might have better farmers and sunlight so they have larger squash C. Some might have less skillful farmers so their squash A are smaller. On average, squash C is 3 times as heavy as squash A, and squash B is 2 times of squash A (because of the sunlight and farmers). However, the larger the squash, the harder of it to grow. Therefore, the standard deviation of squash C is 3 times of the squash A, and squash B is 2 times of squash A. For any squash, the probability of growing to be larger than the average is equal to the probability of growing smaller than the average. That is to say, the weight distribution of any squash is symmetric to the average. That's the reason we choose Gaussian to model the weight. In addition, the coefficients of each Gaussian components could be seen as the empirical proportions of the whole population for each kind of squash. This implicates the probability of getting squash A, B and C from a random draw, respectively.

3. Hoff problem 4.8

(a). Using a Poisson sampling model, a $\text{gamma}(2, 1)$ prior for each θ and the given data, obtain 5,000 samples of \tilde{Y}_A and \tilde{Y}_B from the posterior predictive distribution of the two samples. Plot the Monte Carlo approximations to these two posterior predictive distributions.

We know that given a Poisson sampling model ($\sum y$ and n) and a gamma prior with parameters α and β , the posterior distribution is $\text{gamma}(\alpha + \sum y, \beta + n)$.

```

men_A = c(1, 0, 0, 1, 2, 2, 1, 5, 2, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 2, 1, 3,
          2, 0, 0, 3, 0, 0, 0, 2, 1, 0, 2, 1, 0, 0, 1, 3, 0, 1, 1, 0, 2, 0, 0, 2, 2, 1,
          3, 0, 0, 0, 1, 1)
men_B = c(2, 2, 1, 1, 2, 2, 1, 2, 1, 0, 2, 1, 1, 2, 0, 2, 2, 0, 2, 1, 0, 0, 3, 6, 1, 6,
          4, 0, 3, 2, 0, 1, 0, 0, 0, 3, 0, 0, 0, 0, 0, 1, 0, 4, 2, 1, 0, 0, 1, 0, 3, 2,

```

```

5, 0, 1, 1, 2, 1, 2, 1, 2, 0, 0, 0, 2, 1, 0, 2, 0, 2, 4, 1, 1, 1, 2, 0, 1, 1,
1, 1, 0, 2, 3, 2, 0, 2, 1, 3, 1, 3, 2, 2, 3, 2, 0, 0, 0, 1, 0, 0, 0, 1, 2, 0,
3, 3, 0, 1, 2, 2, 2, 0, 6, 0, 0, 0, 2, 0, 1, 1, 1, 3, 3, 2, 1, 1, 0, 1, 0, 0,
2, 0, 2, 0, 1, 0, 2, 0, 0, 2, 2, 4, 1, 2, 3, 2, 0, 0, 0, 1, 0, 0, 1, 5, 2, 1,
3, 2, 0, 2, 1, 1, 3, 0, 5, 0, 0, 2, 4, 3, 4, 0, 0, 0, 0, 0, 2, 2, 0, 0, 2,
0, 0, 1, 1, 0, 2, 1, 3, 3, 2, 2, 0, 0, 2, 3, 2, 4, 3, 3, 4, 0, 3, 0, 1, 0, 1,
2, 3, 4, 1, 2, 6, 2, 1, 2, 2)

```

```

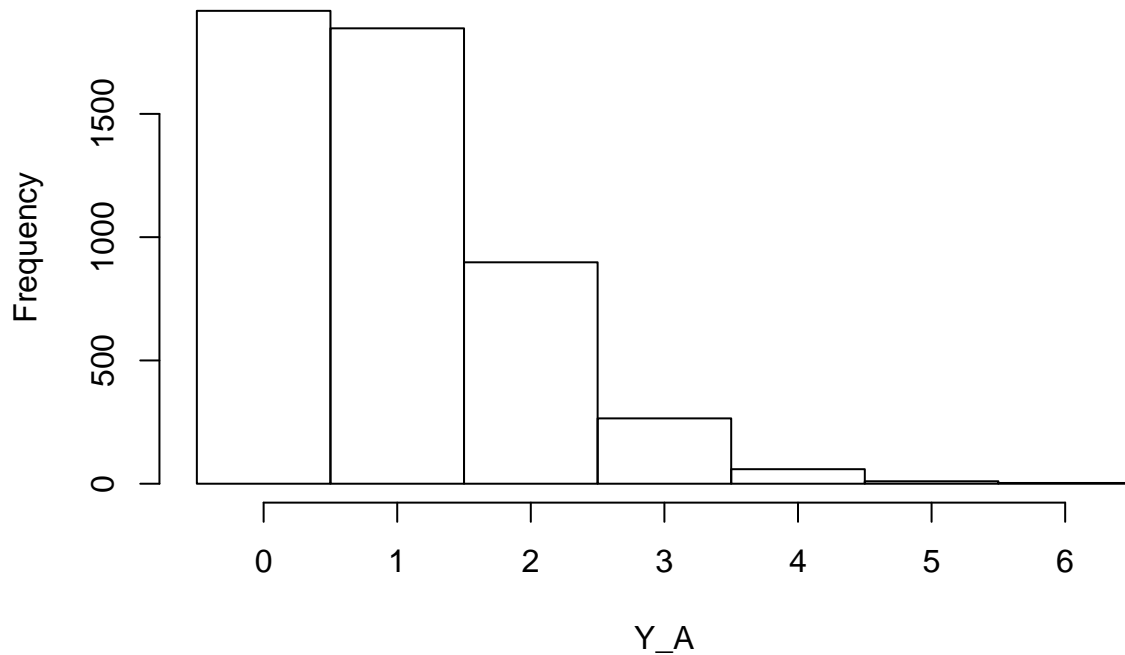
a = 2
b = 1
sum_A = sum(men_A)
sum_B = sum(men_B)
n_A = length(men_A)
n_B = length(men_B)

Y_A = NULL
Y_B = NULL
for (s in 1:5000) {
  theta_A = rgamma(1, a+sum_A, b+n_A)
  theta_B = rgamma(1, a+sum_B, b+n_B)
  Y_A = c(Y_A, rpois(1, theta_A))
  Y_B = c(Y_B, rpois(1, theta_B))
}

```

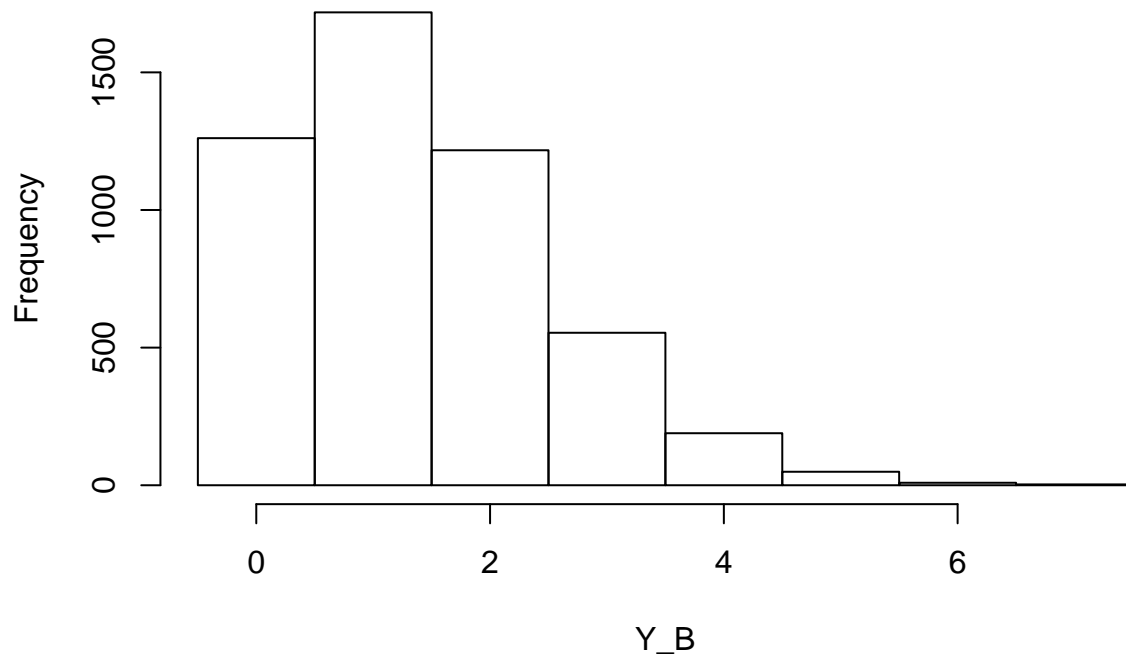
```
hist(Y_A, breaks = seq(-0.5, max(Y_A)+0.5, 1), main = "MC Generated Posterior Predictive Y_A")
```

MC Generated Posterior Predictive Y_A



```
hist(Y_B, breaks = seq(-0.5, max(Y_B)+0.5, 1), main = "MC Generated Posterior Predictive Y_B")
```

MC Generated Posterior Predictive Y_B



(b).

Find 95% quantile-based posterior confidence intervals for $\theta_B - \theta_A$ and $\tilde{Y}_A - \tilde{Y}_B$. Describe in words the differences between the two populations using these quantities and the plots in (a), along with any other results that may be of interest to you.

```
theta_A = NULL
theta_B = NULL
for (s in 1:5000) {
  theta_A = c(theta_A, rgamma(1, a+sum_A, b+n_A))
  theta_B = c(theta_B, rgamma(1, a+sum_B, b+n_B))
}
```

```
quantile(theta_B - theta_A, probs = c(0.025, 0.975))
```

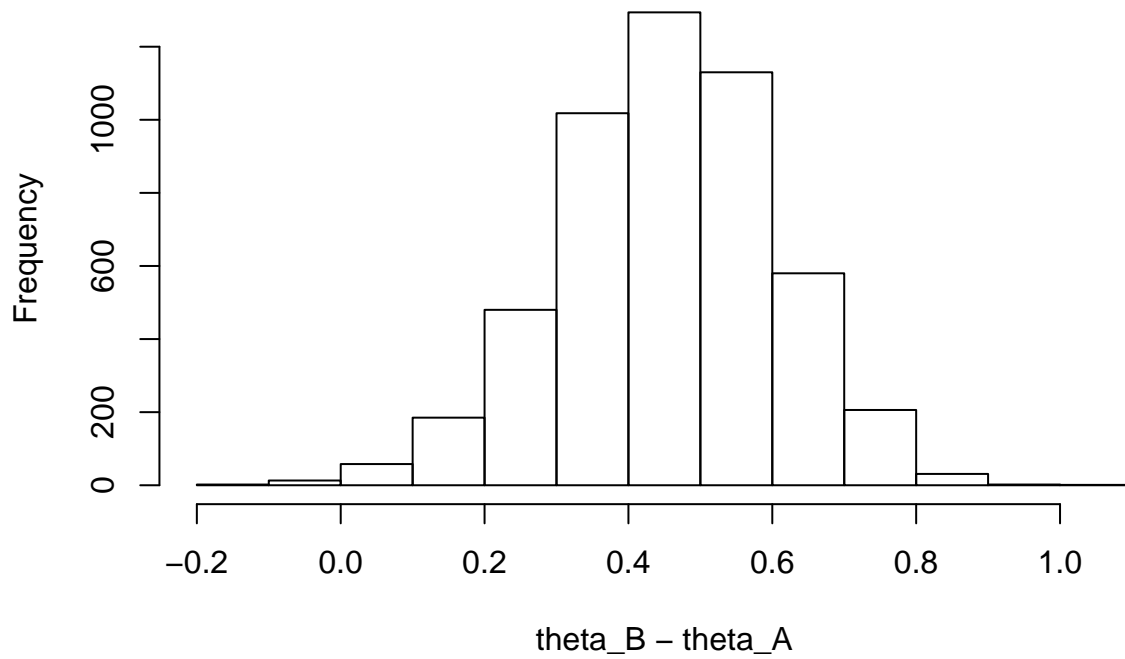
```
##      2.5%      97.5%
## 0.1436837 0.7429073
```

```
quantile(Y_B - Y_A, probs = c(0.025, 0.975))
```

```
## 2.5% 97.5%
## -2    4
```

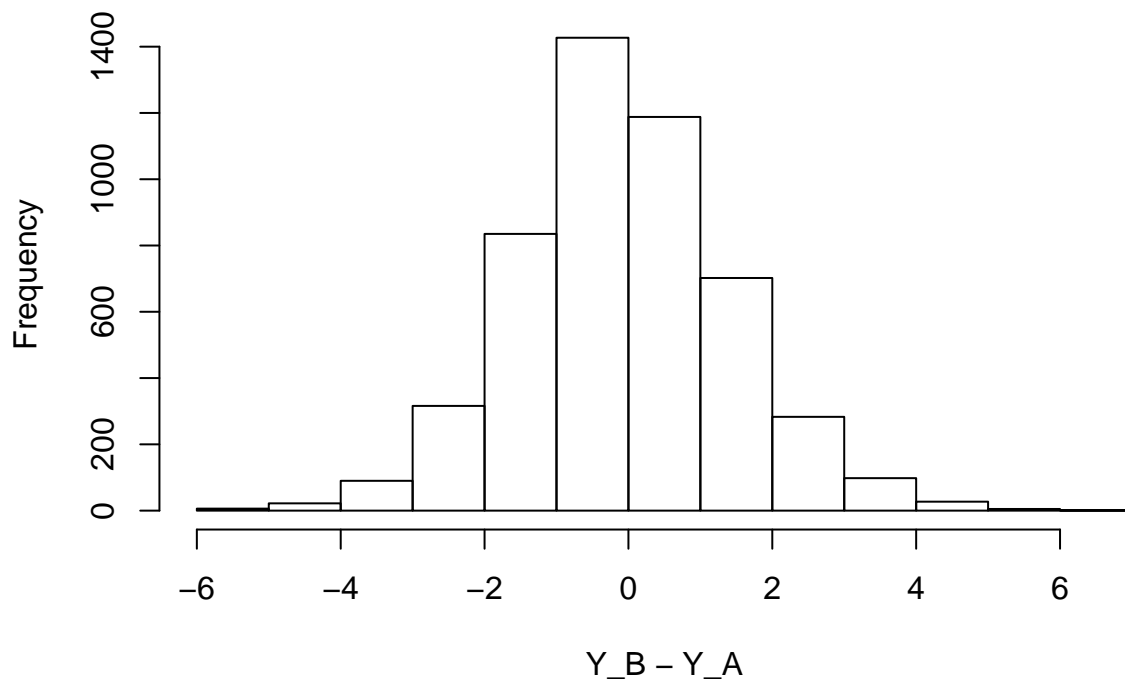
```
hist(theta_B - theta_A)
```

Histogram of $\theta_B - \theta_A$



```
hist(Y_B - Y_A)
```

Histogram of $Y_B - Y_A$



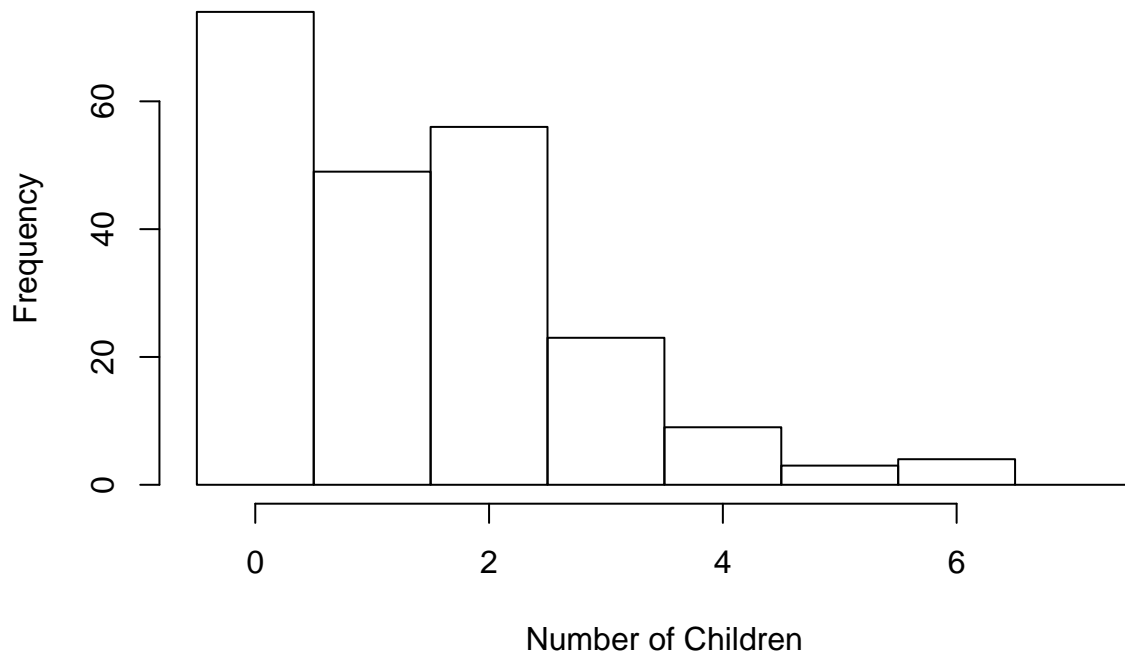
According to the 95% quantile-based confidence interval, and the distribution of these two generated samples, we find that θ_A is extremely unlikely to be greater than θ_B . The confidence interval do not cover zero and only a small portion of it in the histogram is smaller than zero. On the other hand, the values of $Y_B - Y_A$

are almost evenly distributed on either side of zero in the histogram. The confidence interval contains zero. All shows that the distribution of Y_B (θ_B) is more right-shifted than Y_A (θ_A), but the predictive samples has a greater variance than the posterior parameters.

(c). Obtain the empirical distribution of the data in group B. Compare this to the Poisson distribution with mean $\hat{\theta} = 1.4$. Do you think the Poisson model is a good fit? Why or why not?

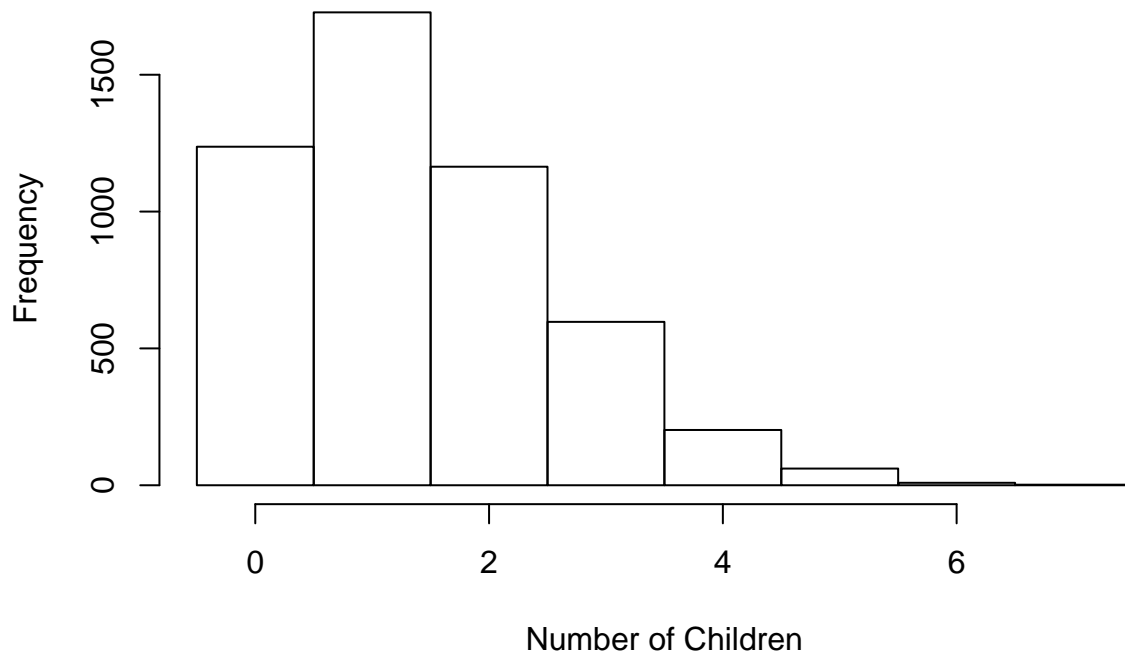
```
hist(men_B, breaks = seq(-0.5, max(Y_B)+0.5, 1), xlab = "Number of Children",
     main = "Empirical Distribution of Data in Group B")
```

Empirical Distribution of Data in Group B



```
sam = rpois(5000, 1.4)
hist(sam, breaks = seq(-0.5, max(sam)+0.5, 1), xlab = "Number of Children",
     main = "Distribution of Poisson with Mean 1.4")
```


Distribution of Poisson with Mean 1.4



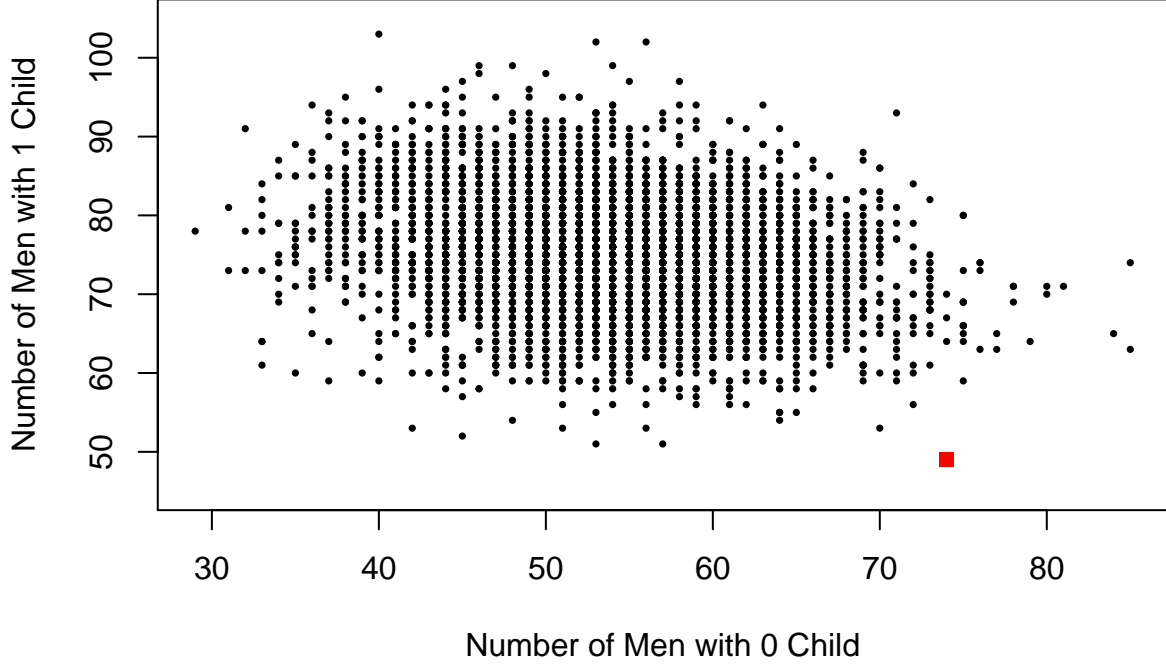
I think it's not a good fit. In the Poisson model with mean 1.4, the number of 0s is almost equal to the number of 2s, and the number of 1s is nearly 1.5 times of the number of 0s (or 2s). In the empirical distribution, however, the number of 1s is nearly 1.5 times of the number of 1s and 2s, and the number of 1s is even smaller than the number of 2s. The Poisson model could not reflect the fact that the number of 1s is smaller than the number of 0s and 2s, and the number of 0s is 1.5 times to the number of 1s.

(d). For each of the 5,000 θ_B -values you sampled, sample $n_B = 218$ Poisson random variables and count the number of 0s and the number of 1s in each of the 5,000 simulated datasets. You should now have two sequences of length 5,000 each, one sequence counting the number of people having zero children for each of the 5,000 posterior predictive datasets, the other counting the number of people with one child. Plot the two sequences against one another (one on the x-axis, one on the y-axis). Add to the plot a point marking how many people in the observed dataset had zero children and one child. Using this plot, describe the adequacy of the Poisson model.

```
count_0 = NULL
count_1 = NULL
for (i in theta_B) {
  sams = rpois(n_B, i)
  count_0 = c(count_0, sum(sams == 0))
  count_1 = c(count_1, sum(sams == 1))
}

plot(count_0, count_1, pch = 16, cex = 0.5, ylim = c(45, max(count_1)+2),
     xlab = "Number of Men with 0 Child", ylab = "Number of Men with 1 Child",
     main = "Generated Predictive Samples and Empirical Data in Group B")
points(sum(men_B == 0), sum(men_B == 1), pch = 15, cex = 1, col = "red")
```

Generated Predictive Samples and Empirical Data in Group B



Still, this Poisson model do not fit the data in terms of this statistics. The empirical data point lies outside of the generated samples area. It means that the majority of the generated predictive samples do not have the same numbers of both 0s and 1s as in the empirical data. This also coincide with what we have mentioned in part (c).

4. Let $Y_1, \dots, Y_n \mid \alpha, \beta$ be i.i.d. $\text{Gamma}(\alpha, \beta)$ with α know.

(a). Find the conjugate family of priors for β .

We have,

$$\begin{aligned}
 p(Y_1, \dots, Y_n \mid \alpha, \beta) &= \prod_{i=1}^n p(Y_i \mid \alpha, \beta) \\
 &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n \prod_{i=1}^n y_i^{\alpha-1} e^{-\beta y_i} \\
 &= \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n \left(\prod_{i=1}^n y_i \right)^{\alpha-1} e^{-\beta \sum_{i=1}^n y_i} \\
 p(\beta \mid Y_1, \dots, Y_n, \alpha) &\propto p(\beta \mid \alpha) p(Y_1, \dots, Y_n \mid \alpha, \beta) \\
 &\propto p(\beta \mid \alpha) \beta^{n\alpha} e^{-\beta \sum_{i=1}^n y_i}
 \end{aligned}$$

We can see that the prior of β in this model should at least contain $\beta^{c_1} e^{-c_2 \beta}$. A gamma distribution could satisfy this condition.

(b). Find the posterior.

Suppose we have $p(\beta \mid \alpha) = \text{Gamma}(\alpha_0, \beta_0)$. Then we have,

$$\begin{aligned} p(\beta \mid Y_1, \dots, Y_n, \alpha) &\propto p(\beta \mid \alpha) \beta^{n\alpha} e^{-\beta \sum_{i=1}^n y_i} \\ &\propto \beta^{\alpha_0-1} e^{-\beta_0 \beta} \beta^{n\alpha} e^{-\beta \sum_{i=1}^n y_i} \\ &\propto \beta^{\alpha_0+n\alpha-1} e^{-(\beta_0 + \sum_{i=1}^n y_i) \beta} \\ &\propto \text{Gamma}(\alpha_0 + n\alpha, \beta_0 + \sum_{i=1}^n y_i) \end{aligned}$$

By using the kernel trick we can find that the posterior will be $\text{Gamma}(\alpha_0 + n\alpha, \beta_0 + \sum_{i=1}^n x_i)$.

(c) Give an interpretation of the prior parameters as things like “prior mean”, “prior variance”, “prior sample size”, etc.

$$\begin{aligned} \mathbb{E}[\beta \mid Y_1, \dots, Y_n] &= \frac{\alpha_0 + n\alpha}{\beta_0 + \sum_{i=1}^n y_i} \\ &= \frac{\alpha}{\beta_0 + \sum_{i=1}^n y_i} \left(\frac{\alpha_0}{\alpha} + n \right) \end{aligned}$$

Let us say for our data Y_1, \dots, Y_n we have n trials with the sum of outcomes $\sum_{i=1}^n x_i$. Then we could interpretate the prior parameters as we have ran $\frac{\alpha_0}{\alpha}$ trials before and we received the sum of outcome β_0 .