HW3

Andrew Liang

9/21/2020

4.2

a)

```r
#sampling a
ya <- c(12,9,12,14,13,13,15,8,15,6)
na <- length(ya)
sya <- sum(ya)

a1 <- 120; a2 <- 10 #prior parameters

yb <- c(11,11,10,9,9,8,7,10,6,8,8,9,7)
nb <- length(yb)
syb <- sum(yb)

b1 <- 12; b2 <- 1

#Monte Carlo sampling
set.seed(1651)

thetaa.mc <- rgamma(10000, a1 + sya, a2 + na) #using 10000 samplings for each theta
thetab.mc <- rgamma(10000, b1 + syb, b2 + nb)

mean(thetab.mc < thetaa.mc)
```

```
## [1] 0.9953
```

b)

```r
n0 <- seq(1:100)

set.seed(1651)
postprob <- c() # for inputting probability value for each iteration

#calculating posterior prob. theta b < theta a for values of n0
for(i in 1:length(n0)){
  thetab.mc <- rgamma(10000, b1*i + syb, i + nb)
  postprob[i] <- mean(thetab.mc < thetaa.mc)
}

print(postprob)
```

```
##    [1] 0.9956 0.9927 0.9889 0.9866 0.9829 0.9758 0.9685 0.9674 0.9620 0.9557
##   [11] 0.9474 0.9413 0.9355 0.9245 0.9171 0.9134 0.9069 0.8990 0.8912 0.8853
##   [21] 0.8759 0.8681 0.8614 0.8551 0.8545 0.8474 0.8326 0.8331 0.8249 0.8220
##   [31] 0.8162 0.8063 0.8017 0.8009 0.7862 0.7895 0.7796 0.7783 0.7714 0.7669
##   [41] 0.7608 0.7559 0.7515 0.7438 0.7416 0.7407 0.7340 0.7308 0.7298 0.7230
##   [51] 0.7203 0.7126 0.7162 0.7119 0.7074 0.7004 0.6976 0.6972 0.6891 0.6833
##   [61] 0.6862 0.6824 0.6741 0.6742 0.6708 0.6713 0.6677 0.6618 0.6591 0.6565
##   [71] 0.6520 0.6500 0.6522 0.6419 0.6470 0.6417 0.6451 0.6379 0.6364 0.6370
##   [81] 0.6342 0.6268 0.6267 0.6277 0.6256 0.6240 0.6243 0.6224 0.6224 0.6210
##   [91] 0.6152 0.6103 0.6047 0.6093 0.6059 0.6054 0.6082 0.6049 0.6021 0.6025
```

```
plot(n0, postprob, type = "l", main = "Posterior Prob. Theta B < Theta A", xlab = "n0", ylab = "Probabil
```


Posterior Prob. Theta B < Theta A
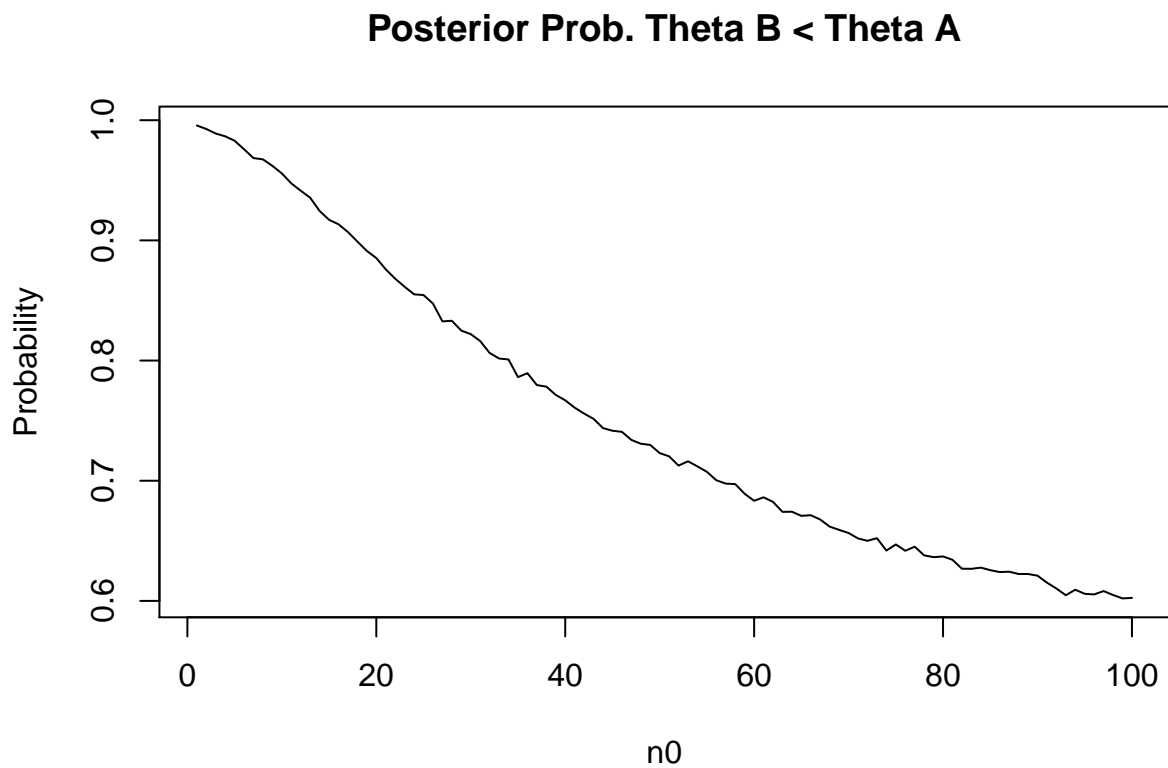
From the output above, we can see all the values of $Pr(\theta_B < \theta_A | y_A, y_B)$ for each $n_0 = 1, 2, \cdots, 100$ and the plot associated with it. We can see that although there is an obvious effect on the posterior probability for an increasing n_0, it is arguable that conclusions about event $\theta_B < \theta_A$ is not very sensitive from the prior distribution of θ_B. Even if $n_0 = 100$, $Pr(\theta_B < \theta_A | y_A, y_B)$ is still greater than 0.6.

c)

```
thetaa.mc <- rgamma(10000, a1 + sya, a2 + na)
thetab.mc <- rgamma(10000, b1 + syb, b2 + nb)

#posterior predictive sampling
set.seed(1651)
```

```
ya.mc <- rpois(10000, thetaa.mc)
yb.mc <- rpois(10000, thetab.mc)

mean(yb.mc < ya.mc)
```

```
## [1] 0.6996
```

```
set.seed(1651)
postpred <- c()

#calculating posterior prob. theta b < theta a for values of n0
for(i in 1:length(n0)){
  thetab.mc <- rgamma(10000, b1*i + syb, i + nb)
  newyb.mc <- rpois(10000, thetab.mc)
  postpred[i] <- mean(newyb.mc < ya.mc)
}

print(postpred)
```

```
##   [1] 0.7037 0.6839 0.6660 0.6510 0.6437 0.6351 0.6310 0.6106 0.6000 0.6003
##  [11] 0.5941 0.5941 0.5858 0.5765 0.5780 0.5710 0.5671 0.5630 0.5536 0.5534
##  [21] 0.5529 0.5527 0.5552 0.5441 0.5461 0.5346 0.5389 0.5345 0.5279 0.5319
##  [31] 0.5224 0.5205 0.5180 0.5264 0.5265 0.5216 0.5235 0.5213 0.5131 0.5034
##  [41] 0.5099 0.5168 0.5170 0.5151 0.5101 0.5074 0.5023 0.5032 0.4997 0.5026
##  [51] 0.4954 0.4960 0.4986 0.5001 0.5024 0.5034 0.4924 0.4948 0.4970 0.4938
##  [61] 0.4932 0.4939 0.4908 0.4972 0.4926 0.5052 0.4847 0.4926 0.4926 0.4901
##  [71] 0.4865 0.4958 0.4886 0.4921 0.4882 0.4901 0.4843 0.4823 0.4813 0.4827
##  [81] 0.4821 0.4846 0.4777 0.4836 0.4773 0.4874 0.4829 0.4773 0.4898 0.4846
##  [91] 0.4809 0.4757 0.4768 0.4817 0.4843 0.4828 0.4732 0.4806 0.4804 0.4846
```
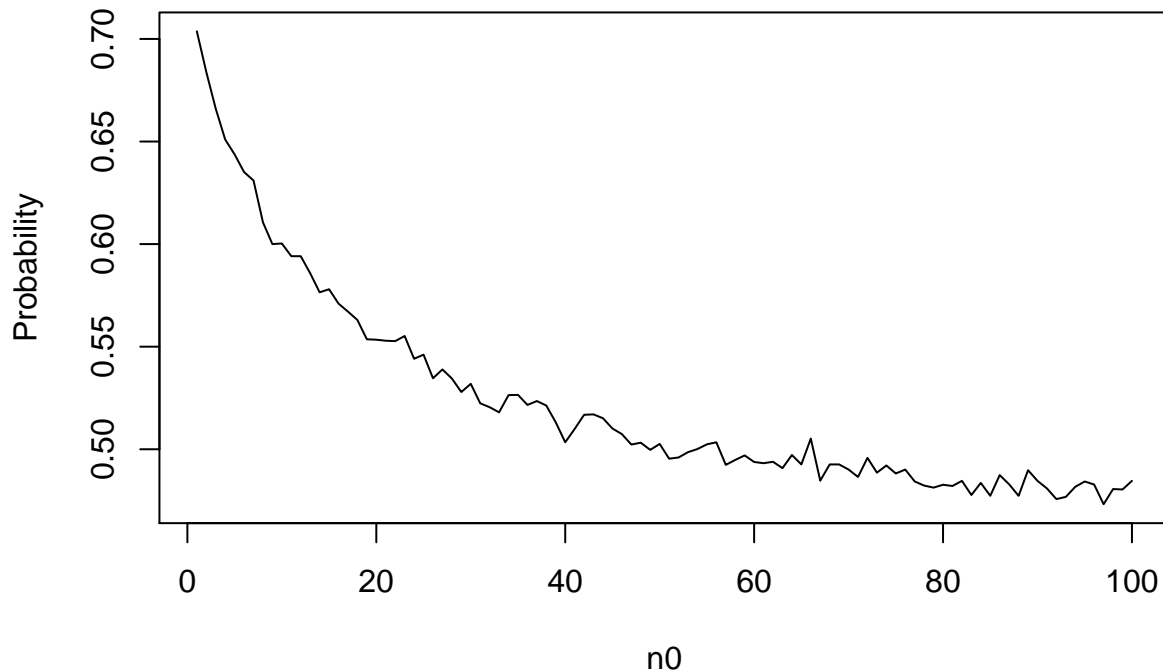
```
plot(n0, postpred, type = "l", main = "Posterior Prob. Y tilda B < Y tilda A", xlab = "n0", ylab = "Prol
```

Posterior Prob. Y tilda B < Y tilda A

From the output above, we can see all the values of $Pr(\tilde{Y}_B < \tilde{Y}_A | y_A, y_B)$ for $n_0 = 1, 2, \cdots, 100$, as well as a plot of the two. Comparing this to the result in part b), we can argue that the conclusions of $\tilde{Y}_B < \tilde{Y}_A$ are more sensitive. For even $n_0 = 1$, we can see that the probability starts initially at 0.7, and goes below 0.5 as n_0 tends to 100. We also note that the pattern of probabilities are much more sensitive/volatile as n_0 increases, relative to that of part b).
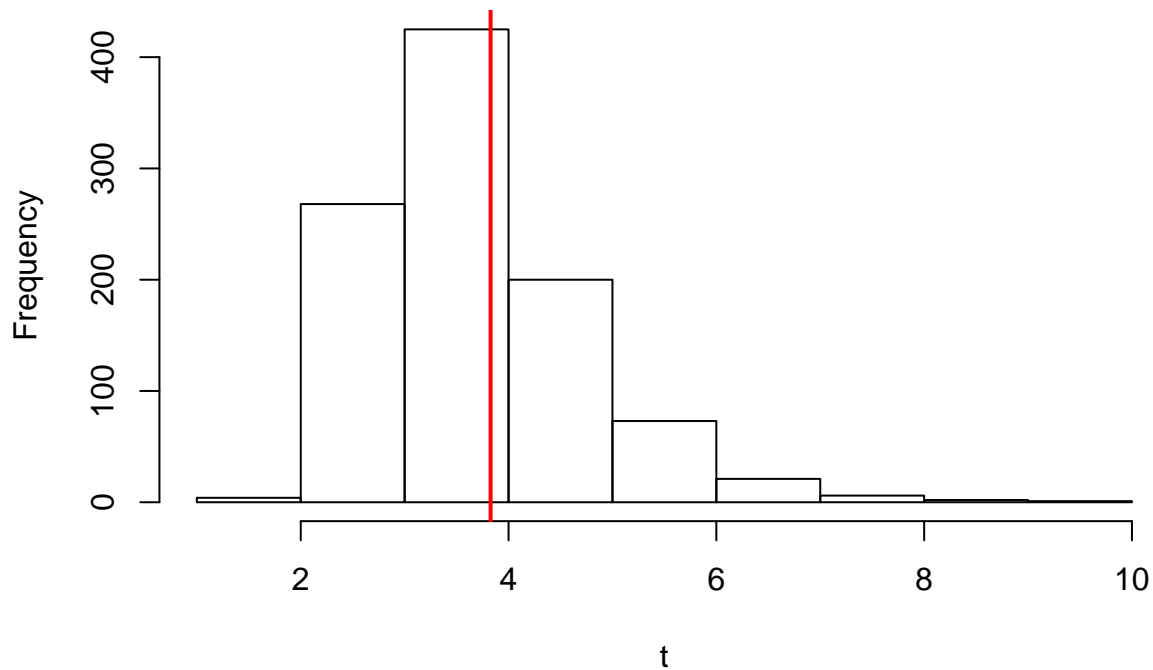
4.3

a)

```r
ta.mc <- c() #sample statistic of interest
set.seed(1651)

for(s in 1:1000){
  thetaa.mc <- rgamma(10, a1 + sya, a2 + na)
  ya.mc <- rpois(10,thetaa.mc)
  ta.mc[s] <- mean(ya.mc)/sd(ya.mc)
}
ta.obs <- mean(ya)/sd(ya)

hist(ta.mc, main = "Histogram of t-statistic for sample A", xlab = "t")
abline(v = ta.obs, col = "red", lwd = 2)
```

Histogram of t-statistic for sample A


```r
#calculating absolute mean difference for all t.mc iterations
for(i in 1:1000){
  diff <- abs(ta.mc[i] - ta.obs)
}
mean(diff)
```

```
## [1] 0.222347
```

From the histogram, we can see that the mode of the statistic through the Monte Carlo samples is pretty close to the observed value of the statistic, with an mean absolute difference of about 0.22. Thus, the Poisson is a pretty good fit for the test statistic based on sample A.
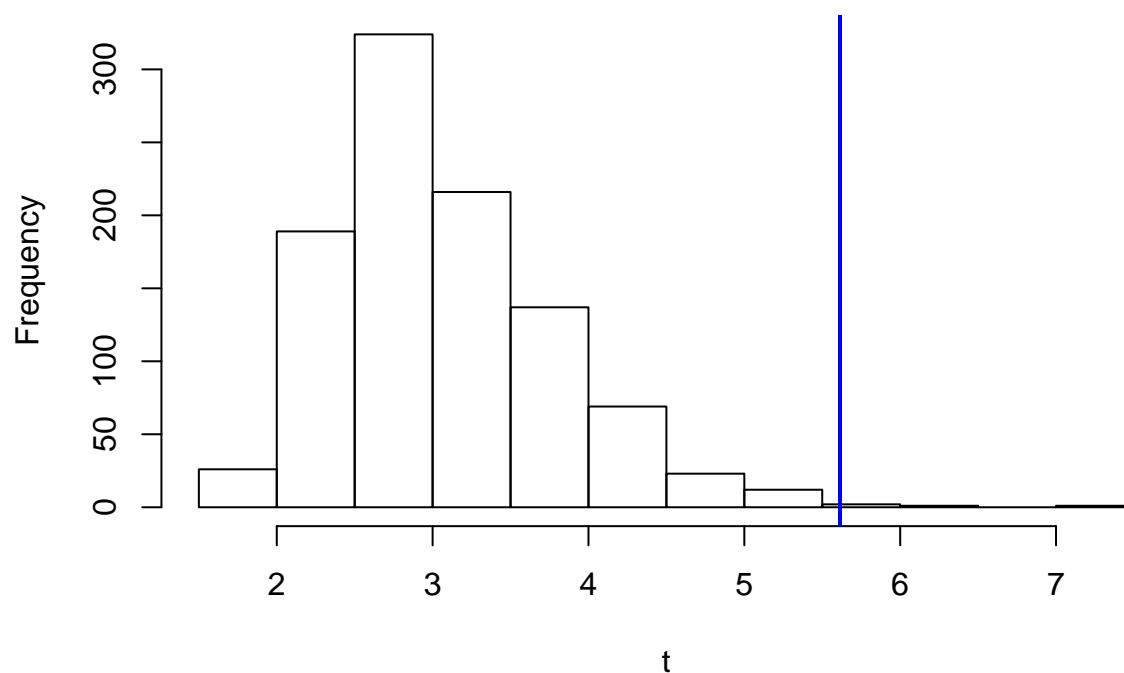
b)

```r
tb.mc <- c()
set.seed(1651)

for(s in 1:1000){
  thetab.mc <- rgamma(13, b1 + syb, b2 + nb)
  yb.mc <- rpois(13,thetab.mc)
  tb.mc[s] <- mean(yb.mc)/sd(yb.mc)
}
tb.obs <- mean(yb)/sd(yb)

hist(tb.mc, main = "Histogram of t-statistic for sample B", xlab = "t")
abline(v = tb.obs, col = "blue", lwd = 2)
```

Histogram of t−statistic for sample B


```r
for(i in 1:1000){
  diff <- abs(tb.mc[i] - tb.obs)
}
mean(diff)
```

```
## [1] 2.707255
```

From the histogram, we can see that the mode of the statistic through the Monte Carlo samples is much further way from the observed value of the statistic, with an mean absolute difference of about 2.71. Thus, the Poisson is not a good fit for the test statistic based on sample B.