Methods Homework 1

Tim Vigers

September 9, 2018

Exercise 1

a.

Calculate the probability that 2.5% of Patagonians have the disease, assuming a sample size of 120 and population prevalence of 1%. Use both the exact binomial probability and the Poisson approximation of it. Compare the two.

```
# Calculate the probability using the binomial PMF. With a sample size 120, 2.5% is equal to three case
choose(120,3) * (0.01)^3 * (1 - 0.01)^(120-3)
```

```
## [1] 0.08665163
```

```
# Double check with dbinom().
dbinom(x = 3,size = 120,prob = 0.01)
```

```
## [1] 0.08665163
```

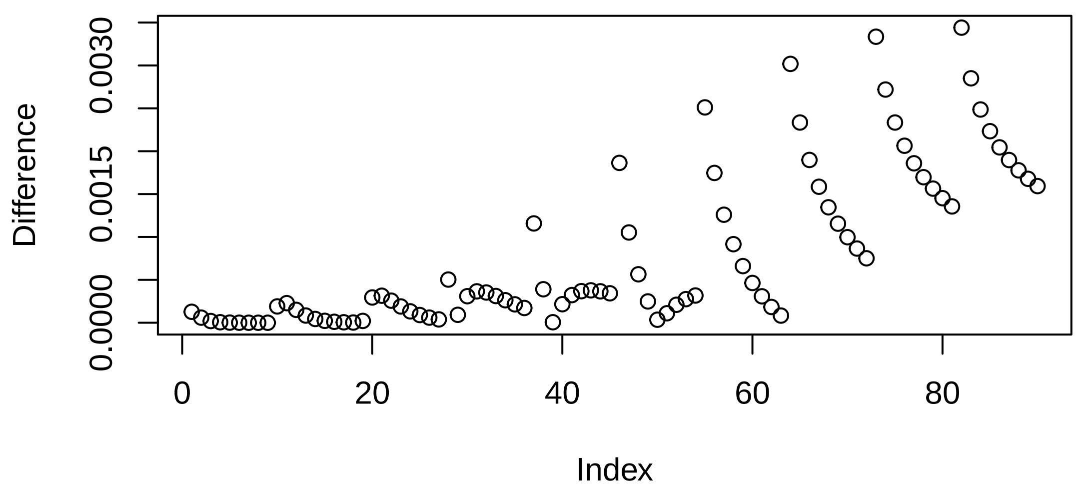
```
# For the Poisson disribution, lambda = np. Check the probability using dpois().
dpois(x = 3, lambda = (120 * 0.01))
```

```
## [1] 0.08674393
```

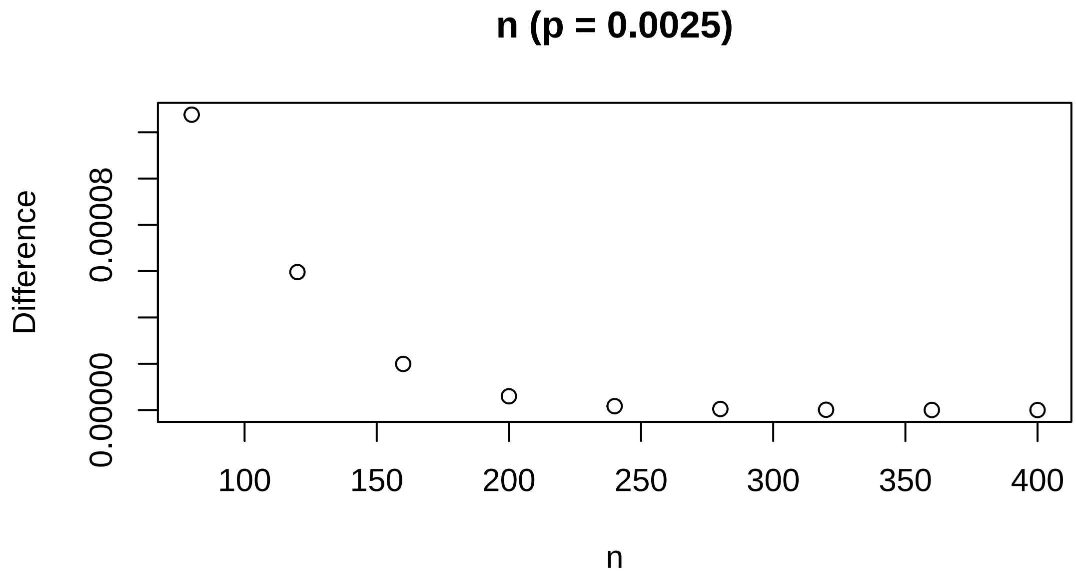
It looks like the Poisson approximation works well for this case (it also fits Rosner's rule where n >= 100 and p <=0.01).

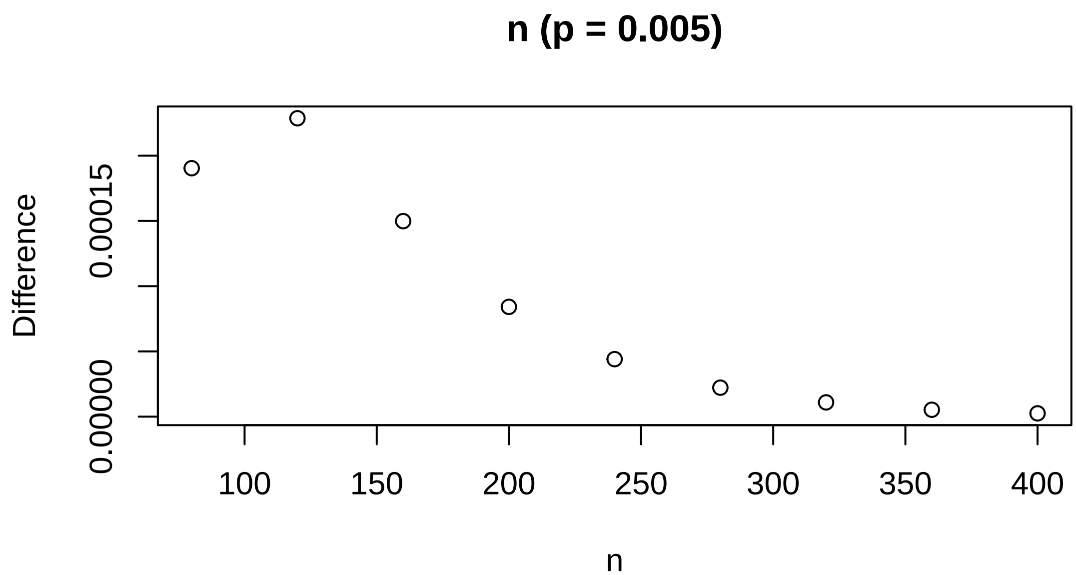
b.

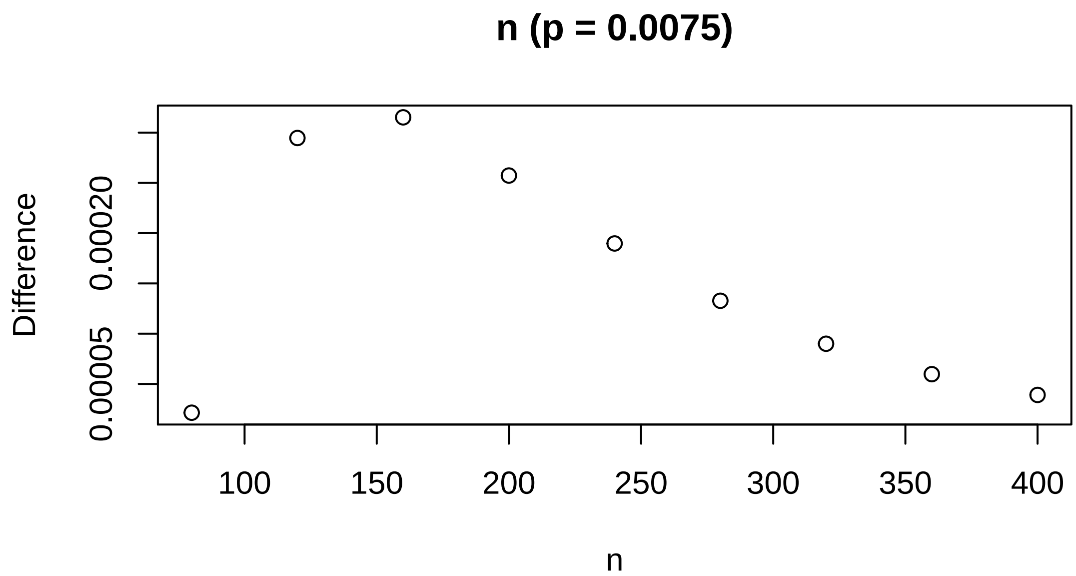
```
# Use the hint provided.
n=seq(80,400,by=40)
p=seq(0.0025,.025,by=.0025)
np<-expand.grid(n=n,p=p)
# Add a column to np for each of the exact binomial, the Poisson approximation of the binomial, and the
np$binom <- dbinom(x = (0.025 * np$n),size = np$n,prob = np$p)
np$poisson <- dpois(x = (0.025 * np$n), lambda = (np$n * np$p))
np$diff <- abs(np$binom - np$poisson)
# Plot everything together.
plot(np$diff,ylab = "Difference")
```

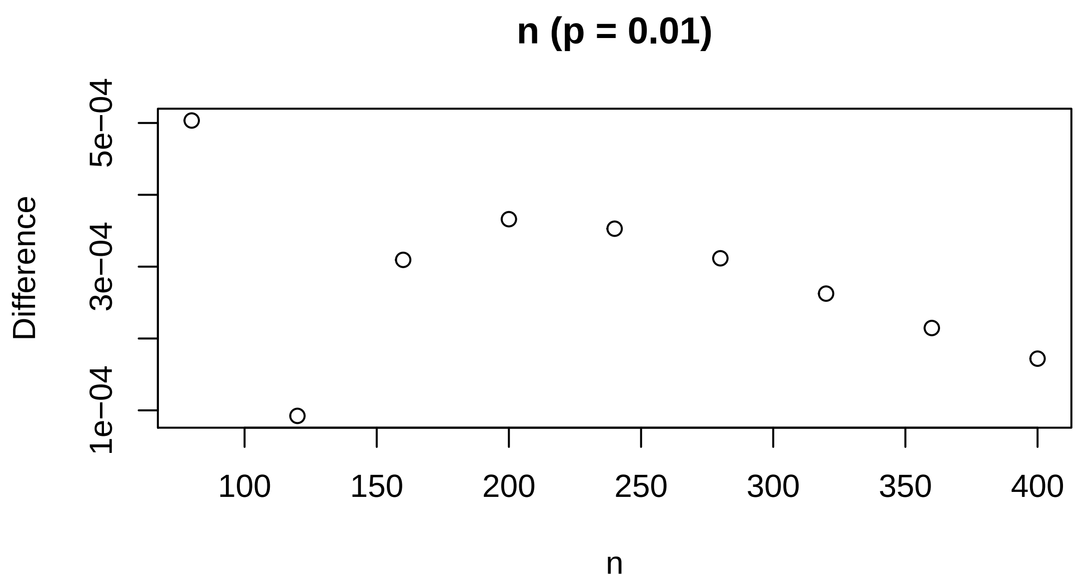


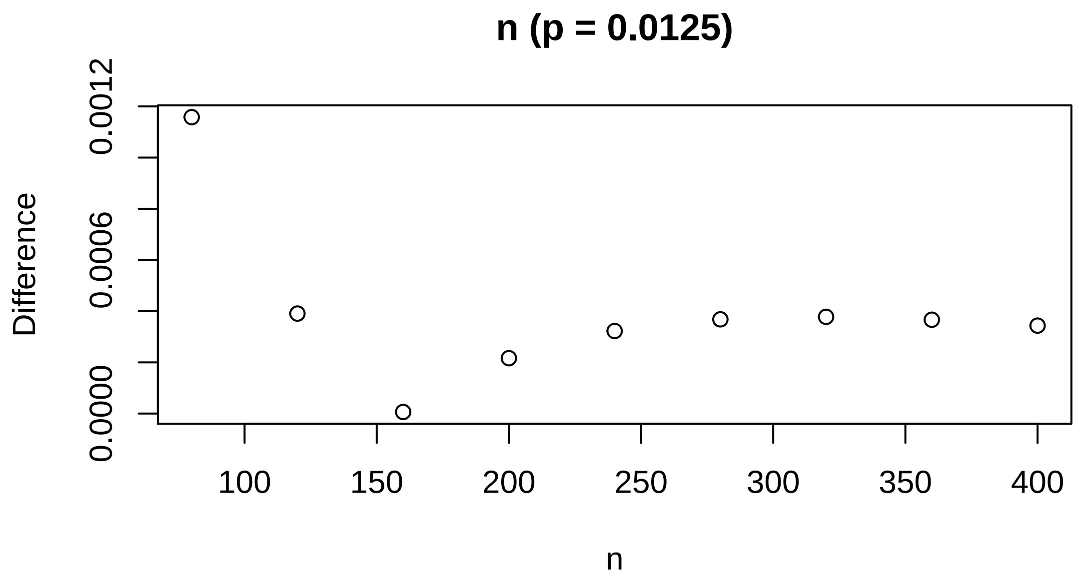
Plot each probability separately.



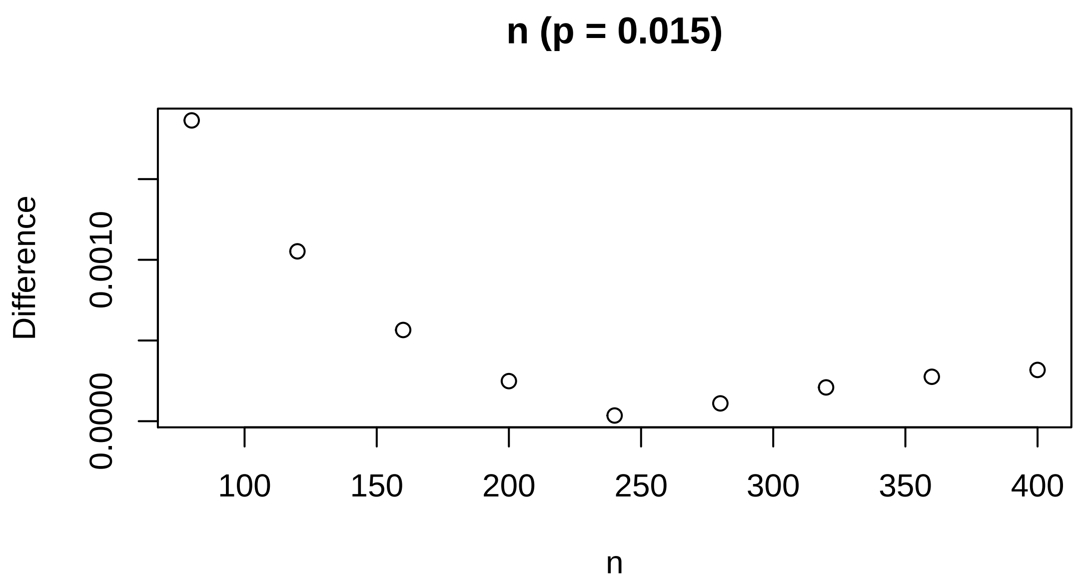



n (p = 0.0075)
Difference
0.00005
0.00020
100
150
200
250
300
350
400
n

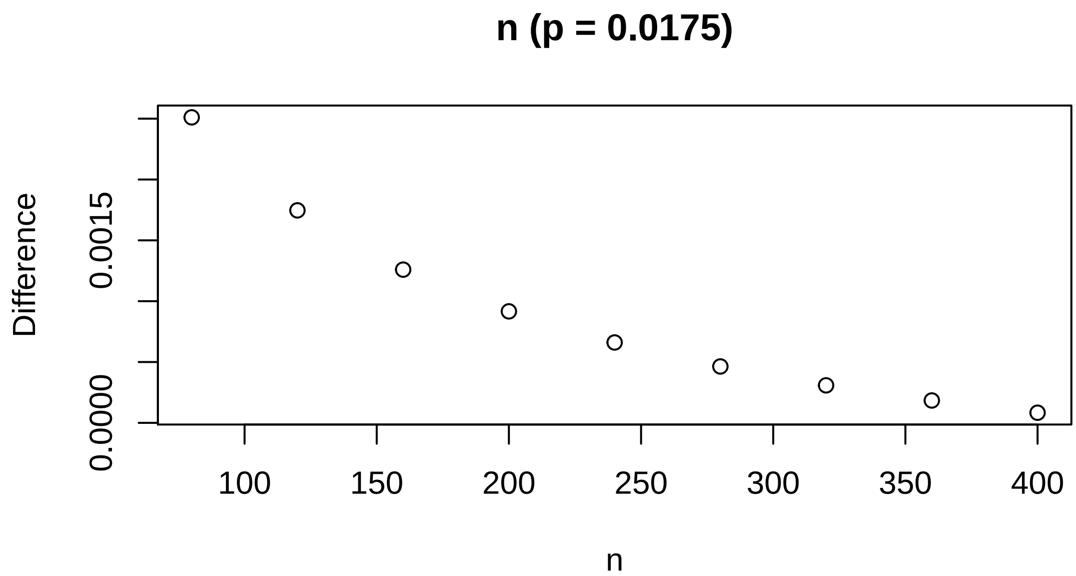

n (p = 0.01)
Difference
1e−04
3e−04
5e−04
100
150
200
250
300
350
400
n

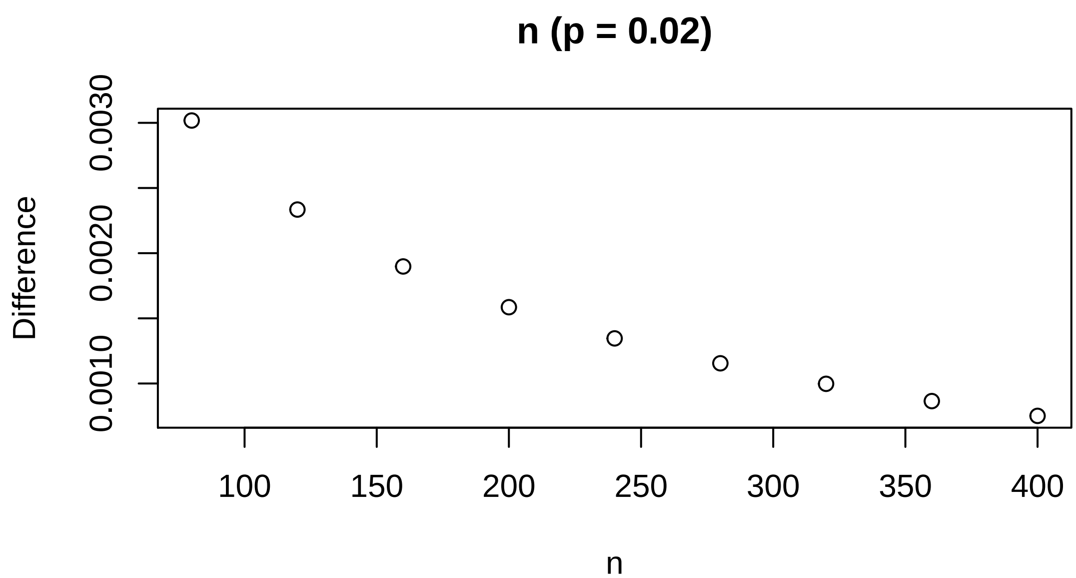


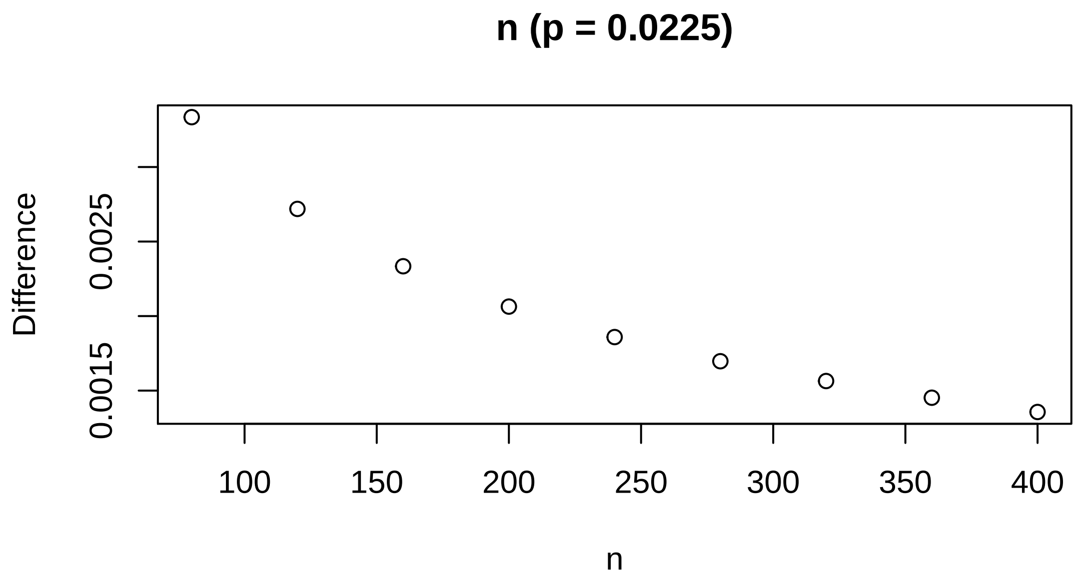
n (p = 0.0125)
Difference
0.0000
0.0006
0.0012
100
150
200
250
300
350
400
n

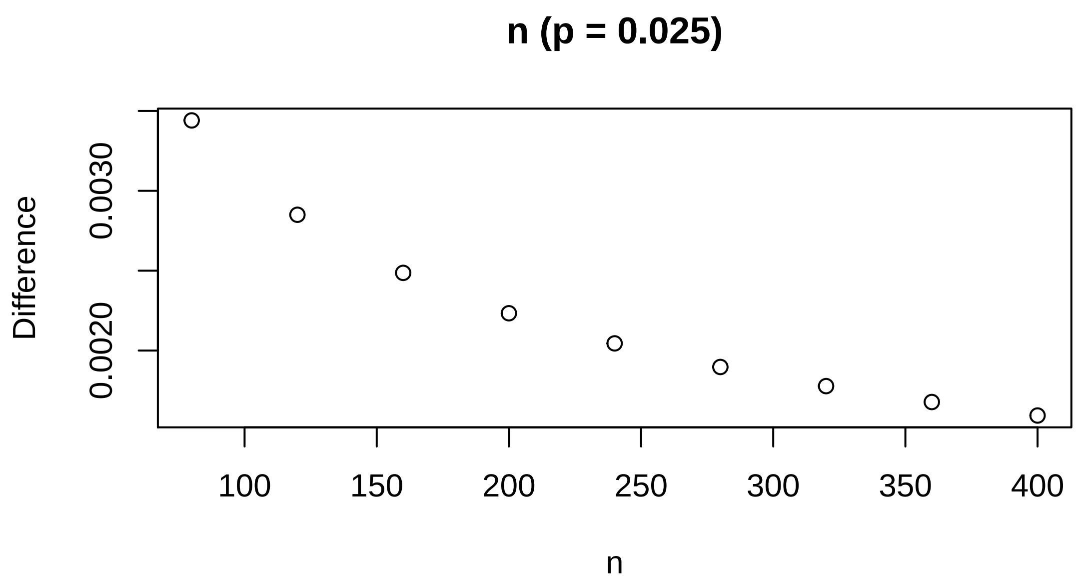


n (p = 0.015)
Difference
0.0000
0.0010
100
150
200
250
300
350
400
n


n (p = 0.0175)
Difference
0.0000
0.0015
100
150
200
250
300
350
400
n


n (p = 0.02)
Difference
0.0010
0.0020
0.0030
100
150
200
250
300
350
400
n


n (p = 0.0225)
Difference
0.0015
0.0025
100
150
200
250
300
350
400
n


n (p = 0.025)
Difference
0.0020
0.0030
100
150
200
250
300
350
400
n

c.

Looking at the plot of everything together, it looks like the differences start to go a little wild at around index 30 (most likely a little before). Index 30 is p = 0.01 and n = 160. The recommendation depends on how conservative you want to be, but I agree with Rosner that p = 0.01 is the maximum probability you'd want to use the Poisson approximation for. I also think that his rule of n >= 100 makes sense based on this plot, since for p = 0.01, n = 80 looks pretty bad.

Exercise 2