Solutions:

1a

For both type of distribution as random variable was not mentioned I considered 10 for both.

```r

for (i in 1:10000){
    poission <- rpois(n = 10, lambda =1.5)
    mean(poission)
    sd(poission)
}
```

```r
for (i in 1:10000){
    binom <- rbinom( n=5, size = 10, prob = 0.15)
    mean(binom)
    sd(binom)
}
```

I have simulated sample of 10,000 for both poisson and Binomial distribution. Here poisson distribution with lambda as 1.5 and bionomial with probability of 0.15

1b.

```r
sd(poission)
```

1.43372087784044

```r
summary(poission)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00    0.25    1.00    1.50    2.75    4.00 
```

```r
summary(binom)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    0.0     1.0     1.0     1.2     2.0     2.0 
```

```r
sd(binom)
```

0.836660026534076

In case of both Poisson and Binomial distribution the Mean is only 0.3 different.

But while comparing to the mathematical calculation, I don’t see much variation.

Standard deviation is 1.12 whereas through function I got 0.83 for binomial

Standard deviation is 1.22 whereas through function I got 1.43 for Poisson.

Binomial distribution

Mean $\mu = n \cdot p = (10) \cdot \left(\frac{3}{20}\right) = \frac{3}{2}.$

Variance $\sigma^2 = n \cdot p \cdot (1-p) = (10) \cdot \left(\frac{3}{20}\right) \cdot \left(1 - \left(\frac{3}{20}\right)\right) = \frac{51}{40} \approx 1.275.$

Standard deviation

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{(10) \cdot \left(\frac{3}{20}\right) \cdot \left(1 - \left(\frac{3}{20}\right)\right)} = \frac{\sqrt{510}}{20} \approx 1.12915897906362.$$

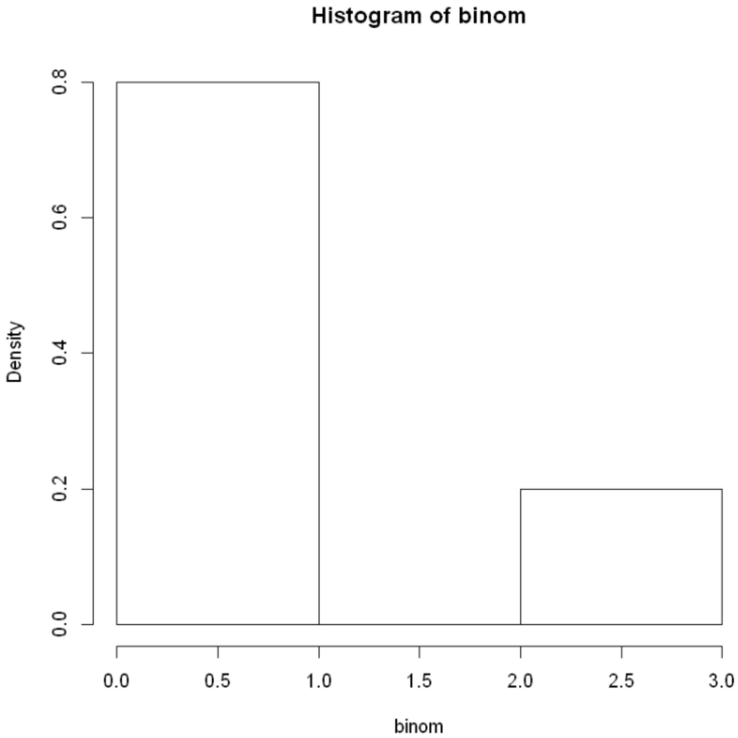
Poission distribution

Mean $\mu = \lambda = \frac{3}{2}$.

Variance $\sigma^2 = \lambda = \frac{3}{2}$.

Standard deviation $\sigma = \sqrt{\lambda} = \sqrt{\frac{3}{2}} \approx 1.22474487139159$.

1c

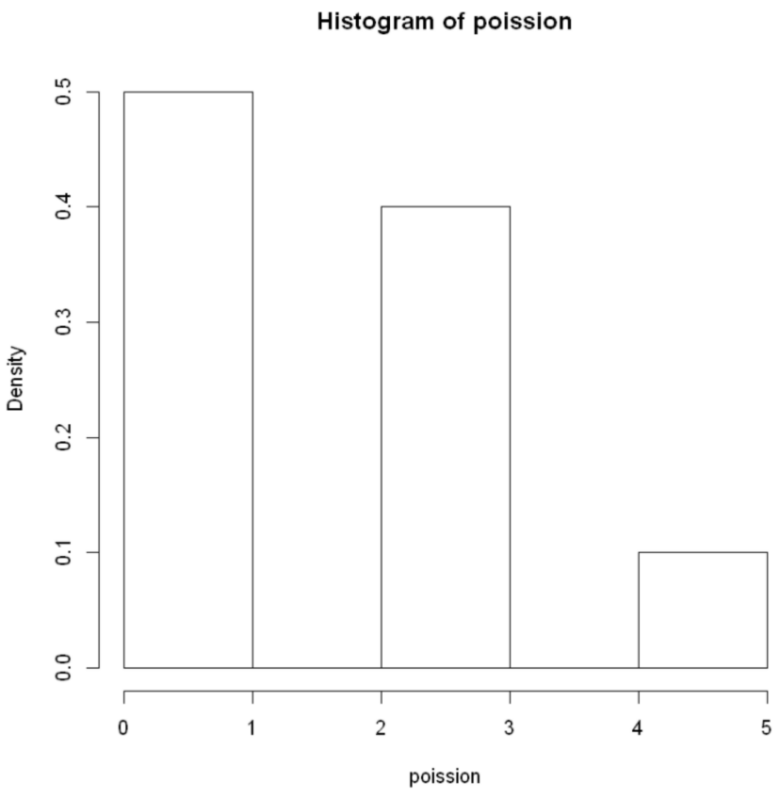


This graph shows the clear 2 distriniction based on the probability distribution of 0.15

The density is varing from 0.2 to 0.8 with distinct diference between sample.

If we consider the numbers as population then with exact 0.15 probability its distributed.

The distinction made it easy to interpret.



Here the linear change can be seen based on the sample number. For higher numbers the density is lower which seems to be the ideal case.

As Lambda which can be ‘expected number events in the interval’ i.e. 1.5 here,

Initially for lesser sample size of 0 to 1 the density is 0.5 which is low, slowly from 2 to 3 block the density slightly increased compare to 1^{st} block. Finally at the end its 0.1

2a

Binomial

```
sample <- 120
x1 <- 0.025 * 120
p <- 0.01
B_dist <- dbinom( size=sample, x=x1, prob=p)
print(B_dist)
```

```
[1] 0.08665163
```

Poisson

```
sample <- 120
x1 <- 0.025 * 120
p <- 0.01

lambda <- sample*p
P_dist <- dpois(x=x1, lambda=lambda)

print(P_dist)
```

```
[1] 0.08674393
```

Difference

```
d <- (P_dist - B_dist)
d
```

9.23038976786972e-05

The Binomial distribution approximately close to Poisson distribution

2b

By conisdering the sample hint the sample size will vary between 80 and 400 (by an increment of 40), while the population prevalence varies between 0.25% and 2.5% (by an increment of 0.25%).12 The prevalence is still 2.5%

Hint solution

```
n=seq(80,400,by=40)
p=seq(0.0025,.025,by=.0025)
np<-expand.grid(n=n,p=p)
```

Approximation equ.

```
np$lambda <- np$n*np$p
np$k <- 0.025*np$n
```

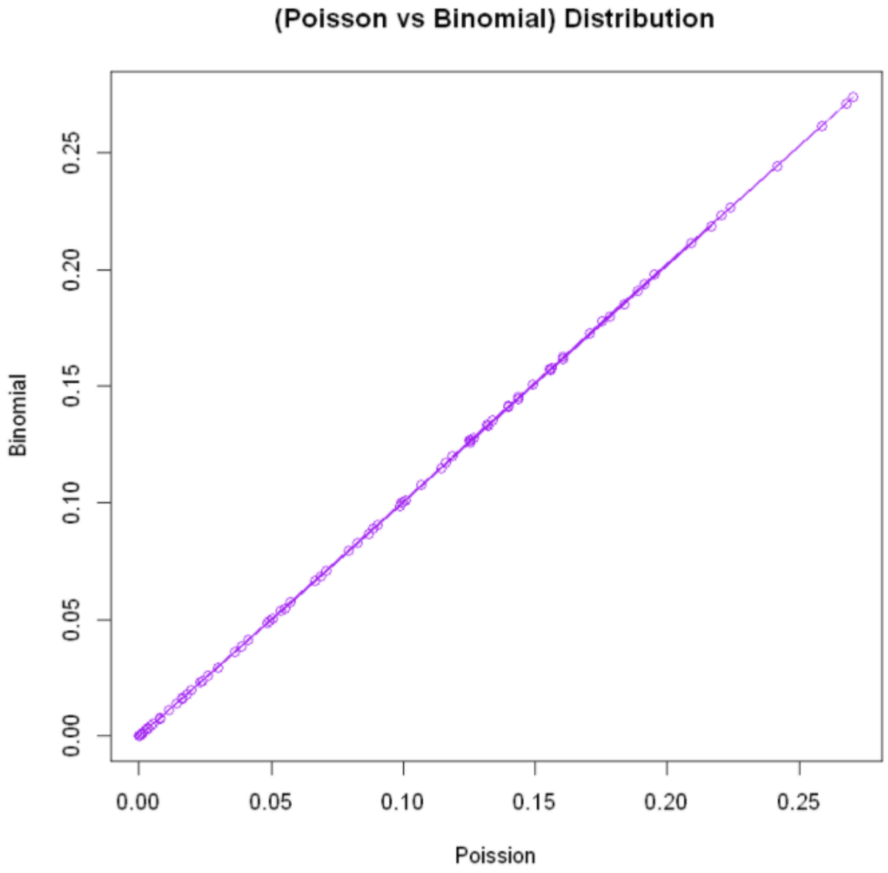
```
np$binomial <- dbinom(size=np$n, x= np$k, p =np$p)
np$poission  <- dpois(x = np$k, lambda = np$lambda)
np$difference <- (np$binomial - np$poission)
```

```
diff <- np$difference
```

```
np
```

200	0.0050	1.0	5	2.981536e-03	3.065662e-03	-8.412585e-05
240	0.0050	1.2	6	1.205044e-03	1.249113e-03	-4.406911e-05
280	0.0050	1.4	7	4.935597e-04	5.157669e-04	-2.220718e-05

```
y2 <- np$binomial
y3 <- np$poission
plot(x=y3,y=y2, main= '(Poisson vs Binomial) Distribution', xlab = 'Poission', ylab= 'Binomial', type="o", col="purple")
```

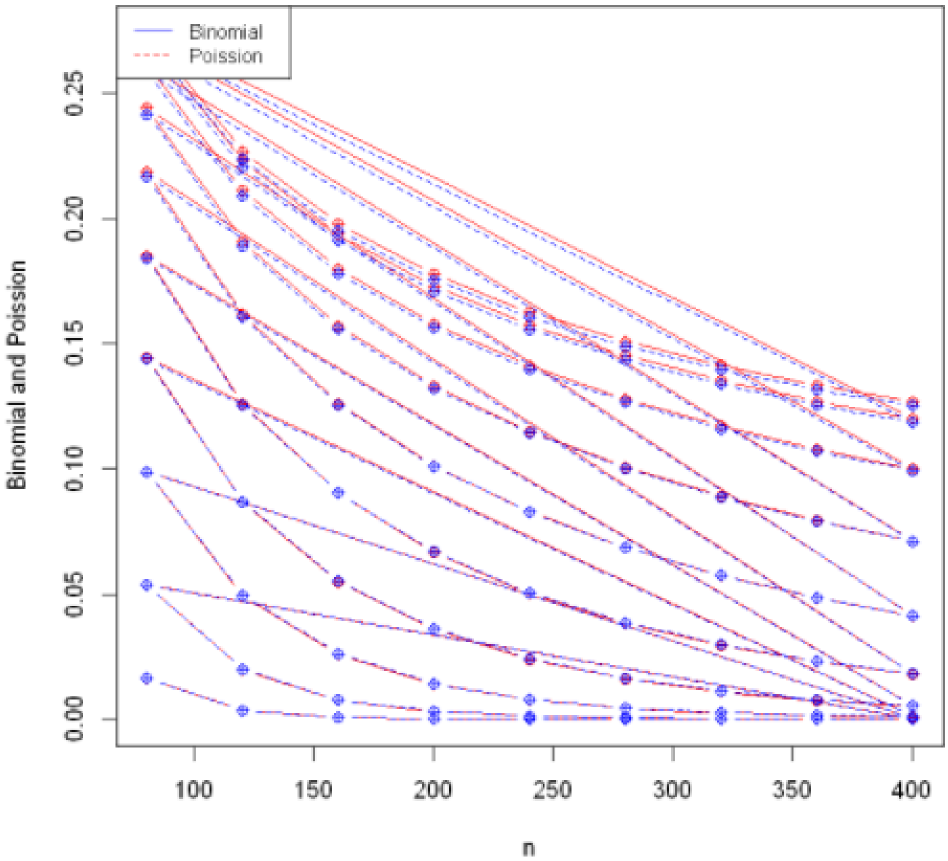


1. Here also Binomial approximation is very very close to the Poisson at each and every sample size

2. This plot shows a one to one relationship, a y=x slope of the line.

3. both of the distribution has minimal difference.

```r
x <- np$n
y1 <- np$binomial
y2  <- np$poission
# Create a first line
plot(x, y1, type = "b", frame = TRUE, pch = 19,
      col = "red", xlab = "n", ylab = "Binomial and Poission")
# Add a second line
lines(x, y2, pch = 18, col = "blue", type = "b", lty = 2)
# Add a legend to the plot
legend("topleft", legend=c( "Binomial","Poission"),
        col=c( "blue","red"), lty = 1:2, cex=0.8)
```



The x axis shows the sample from 80 and 400 (by an increment of 40) where I calculated Poisson and binomial distribution.

1. for higher sample size from 250 to 400, passion distribution varies from 0.15 to 0.25
2. for sample 100 to 250 the binomial distribution varies from 0.00 to 0.15
3. This is a clear distinction between sample size and the results of both type of distribution
4.

```r
np$difference <- (np$binomial - np$poission)
diff <- np$difference
summary(diff)
```

```
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max. 
-3.778e-04 -1.456e-04  1.069e-05  6.553e-04  1.536e-03  3.441e-03 
```

5. from the graph as well as the summary, its clearly visible that both type of distribution for this case is almost similar with less differences.

2c

Ideally For Large sample size i.e. (> 15) and small p (< 0.1) value Poisson distribution is a better option compare to binomial.

So for n in between 150 to 400 with probability with p less than 0.1, Poisson distribution will be better.

Sample size in between 100-150 but greater probability, binomial is better.

Difference between the two distribution can be,

1. measures the number of certain random events (or "successes") within a certain frame,

2. Binomial is based on discrete events (have a certain number i.e. 'N' of "attempts)

   N.p

3. Poisson is based on continuous events. (infinity attempts of infinite chance to get a success)

   But if N → ∞, p → 0 like [Np → λ], then the distribution approaches to passion

   Hence Poisson distribution can happen large No. of times with less occurrence