

Prashant Sharma

Grading:

20 marks- Complete submission of worksheet4
40 marks- Problem 1 and 40 marks- Problem 2

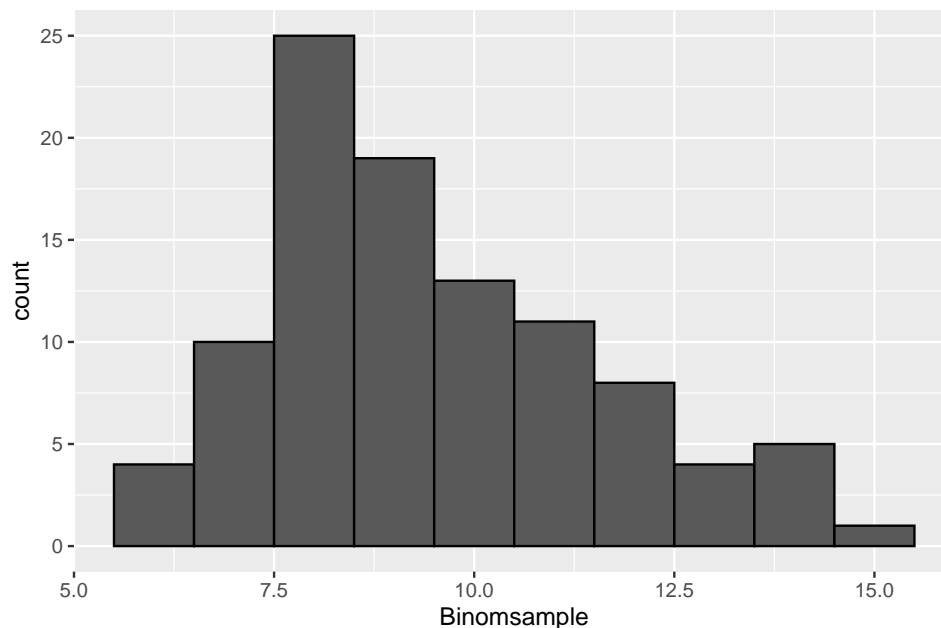
Worksheet-4:

Problem 1(a):

```
> #Generating random samples for Binomial distribution.  
> Binomsample=rbinom(100,20,0.5)  
> Binomsample
```

```
[1] 11 14 11 14 9 8 8 10 8 8 12 8 14 9 9 8 10 10 11 9 9 8 8 12 8  
[26] 9 8 10 8 9 11 11 7 8 15 7 12 10 12 6 9 9 10 12 7 7 10 9 8 8  
[51] 12 14 8 8 8 10 8 9 10 9 10 7 6 7 8 13 9 11 8 8 9 10 13 10 9  
[76] 6 8 13 9 13 11 12 9 12 7 8 6 9 8 7 7 8 11 9 14 7 11 11 10 11
```

```
> #Plotting the Histogram  
> library(ggplot2)  
> Df_binom=data.frame(Binomsample)  
> Plot_binom=ggplot(Df_binom)+geom_histogram(mapping=aes(x=Binomsample),color= "black",binwidth=1)  
> Plot_binom
```

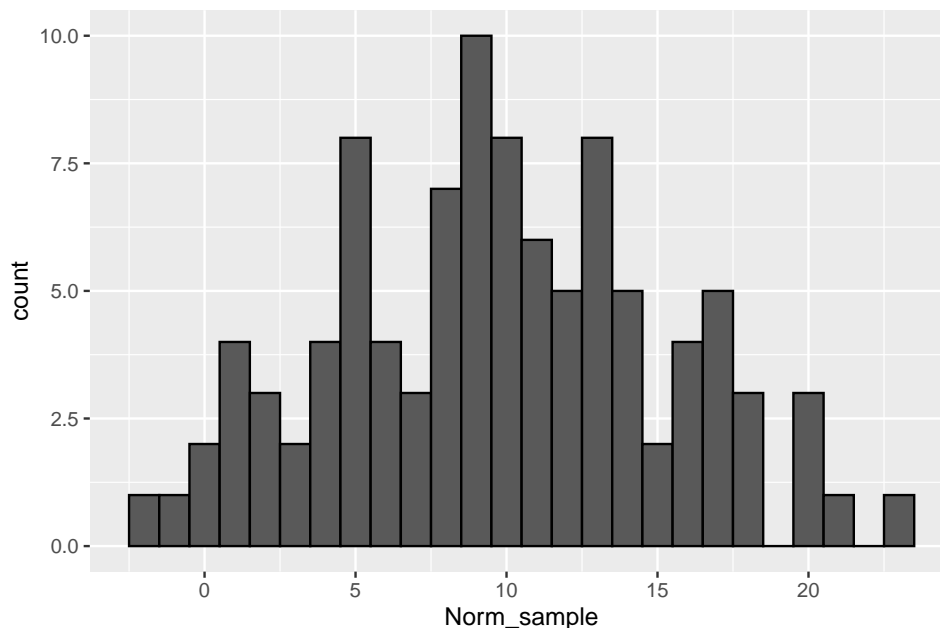


Problem 1(b):

```
> #Generating random samples for Normal distribution.  
> Norm_sample=rnorm(100,10,5)  
> Norm_sample
```

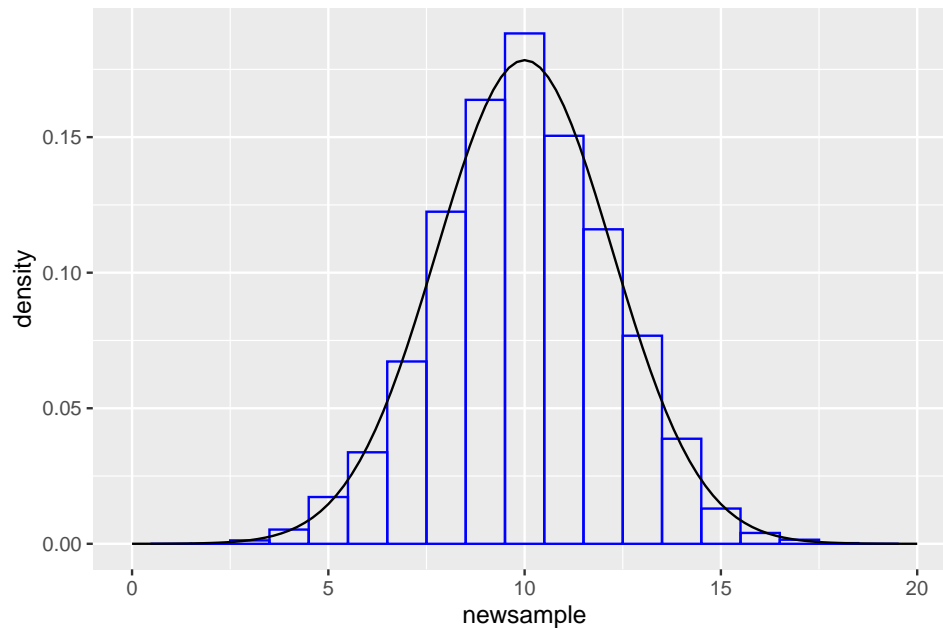
```
[1] 14.0223718 11.8615055 13.3158525 16.3714272  4.8495027 13.1256263  
[7]  9.2169129  7.1070811 20.3951563  9.5597353 13.5064729  8.9578831  
[13]  6.4368200  9.7356973  8.8685195  5.2299831  7.4568250  9.5671881  
[19]  5.2152304 12.8624465  7.5186220 13.3871745  0.1429472  9.7899044  
[25] 12.1083675  4.9570456  9.7226263 16.1585088 13.5368420  5.3099300  
[31]  9.1462326 20.5291740 -2.0281752 17.7715717 10.5378032 13.1780546  
[37] 12.5438178  4.5531456 17.9940652  8.0026207  1.4514650 12.3144308  
[43]  0.3540616  3.9320644 10.5892894 15.3565459  1.0461221  7.1994184  
[49]  5.0014879 14.3232813 12.5714835  3.7504372 10.7400389  9.6967380  
[55]  4.6229344  8.2059836 17.8811239  6.1113988 11.5713738  1.4449460  
[61] 16.4961895 16.5679047  3.5132723  7.8914668  9.2761906 16.9345561  
[67] 10.4557895 11.2589628 12.5010918  2.8238811  1.8102961  8.8954103  
[73]  3.8172852 13.9137238 -1.3777295  1.7275095  1.8048235 11.0264772  
[79] 19.5164570  0.8545873  5.9363183 11.4880325 19.5894484  9.2320778  
[85]  9.2495376  7.8936461  9.4938994 11.5534788  8.1864801 14.5997669  
[91]  2.6156206 10.1312918 16.1524895  9.0775134 16.5416503  7.9730726  
[97] 23.2902863 16.6527147  6.2530773 17.3370827
```

```
> #Plotting the Histogram  
> Df_norm=data.frame(Norm_sample)  
> Plot_norm=ggplot(Df_norm)+geom_histogram(mapping=aes(x=Norm_sample),color= "black",binwidth=1)  
> Plot_norm
```



Now,

```
> density = function(x,a,s){ (1/((2*pi)^(0.5)*s ))* exp(-(x-a)^2/(2*s^2))}
> newsample=rbinom(4000,20,0.5)
> df=data.frame(newsample)
> newplot=ggplot(df)+geom_histogram(mapping=aes(x=newsample,y=.density..), color= "Blue", fill="NA",
+ binwidth=1)+xlim(0,20)+geom_function(fun=density,args=list(a=10,s=(5)^(0.5)))
> newplot
```



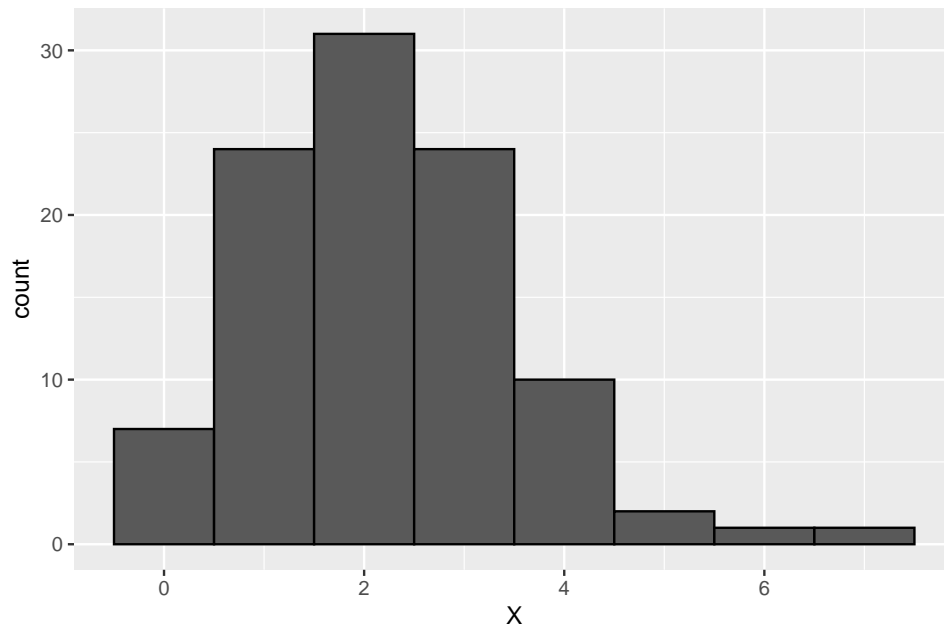
Here, we have generated 4000 samples(for large n) from Binomial(20,0.5).So for this sampling, population mean is $a=20*0.5=10$ and population variance is $=20*0.5*(1-0.5)=5$. Thus, for $n=4000$ normal curve fits the histogram very well.

Problem 2(a):

```
> X=rbinom(100, 2000, 0.001)
> X
```

```
[1] 1 1 1 2 3 4 2 1 2 1 2 1 2 1 2 3 1 2 3 3 3 2 2 2 2 4 3 1 1 2 2 1 2 2 4 2 3
[38] 4 3 2 1 5 1 3 3 3 3 2 0 4 0 3 4 2 4 1 2 1 1 2 3 6 2 3 3 3 1 0 3 1 3 0 1 1
[75] 2 1 2 2 2 4 1 3 2 4 2 2 3 1 3 0 3 5 0 4 3 0 2 7 1 2
```

```
> dfbinomial=data.frame(X)
> Plot_X=ggplot(dfbinomial)+geom_histogram(mapping=aes(x=X), color="black",binwidth=1)
> Plot_X
```

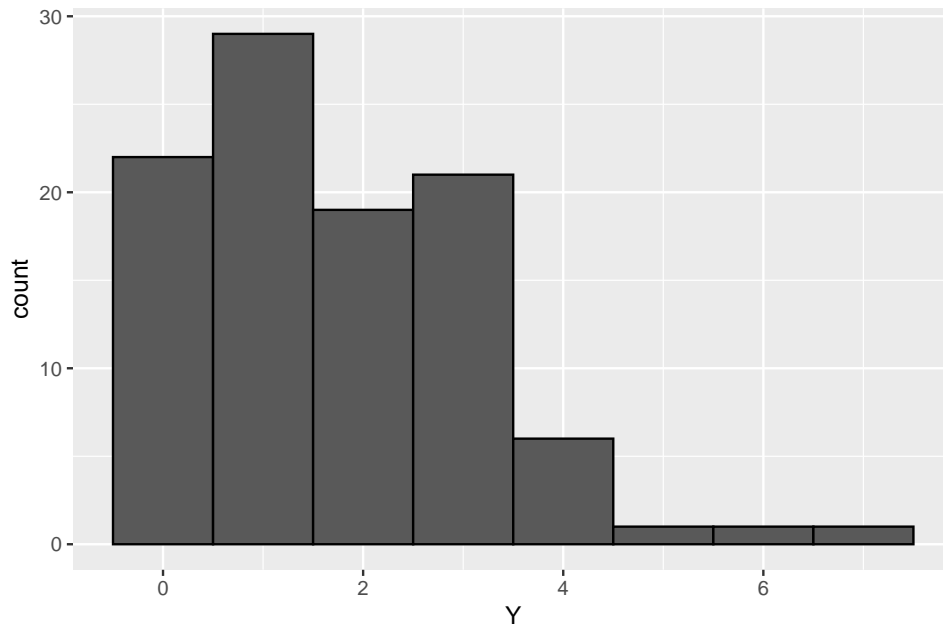


Problem 2(b):

```
> Y=rpois(100,2)
> Y
```

```
[1] 1 1 2 2 5 0 4 1 1 3 4 3 2 3 2 0 3 1 0 1 0 1 2 3 0 2 0 2 0 3 3 1 0 1 1 2 1
[38] 7 3 3 0 2 1 1 2 1 2 0 3 3 1 1 3 3 4 1 0 3 6 2 1 3 0 0 3 0 1 1 4 2 1 4 0 1
[75] 1 0 1 0 1 1 1 1 2 0 3 3 0 3 0 2 2 1 2 3 4 0 2 0 2 3
```

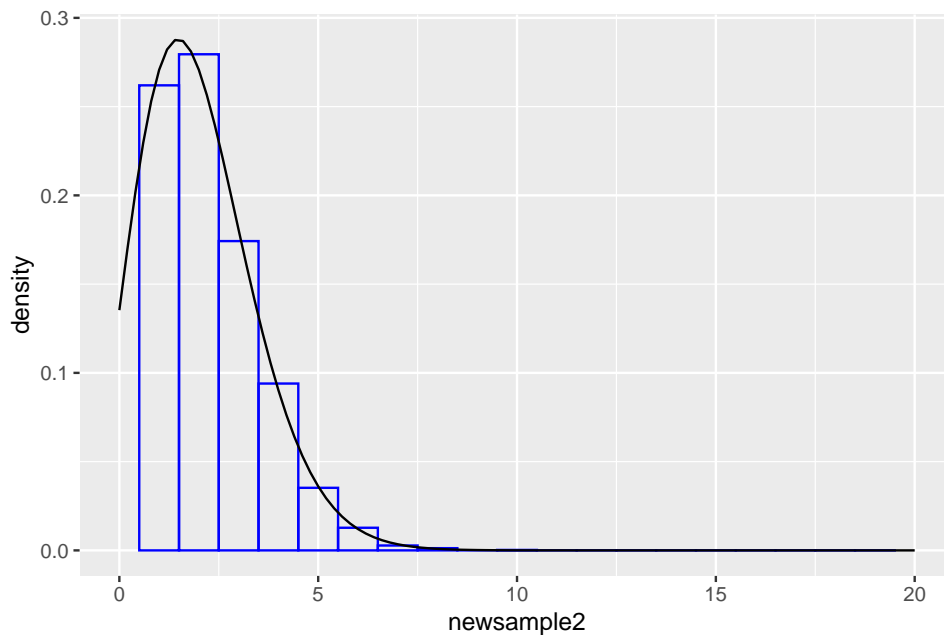
```
> dfnormal=data.frame(Y)
> Plot_Y=ggplot(dfnormal)+geom_histogram(mapping=aes(x=Y), color="black",binwidth=1)
> Plot_Y
```



```

> density=function(a,x){exp(-a)*(a^x)/factorial(x)}
> newsample2=rbinom(4000,2000,0.001)
> df2=data.frame(newsample2)
> newplot2=ggplot(df2)+geom_histogram(mapping=aes(x=newsample2,y=..density..),color="Blue", fill="NA",
+ binwidth=1)+xlim(0,20)+geom_function(fun=density,args=list(a=2))
> newplot2

```



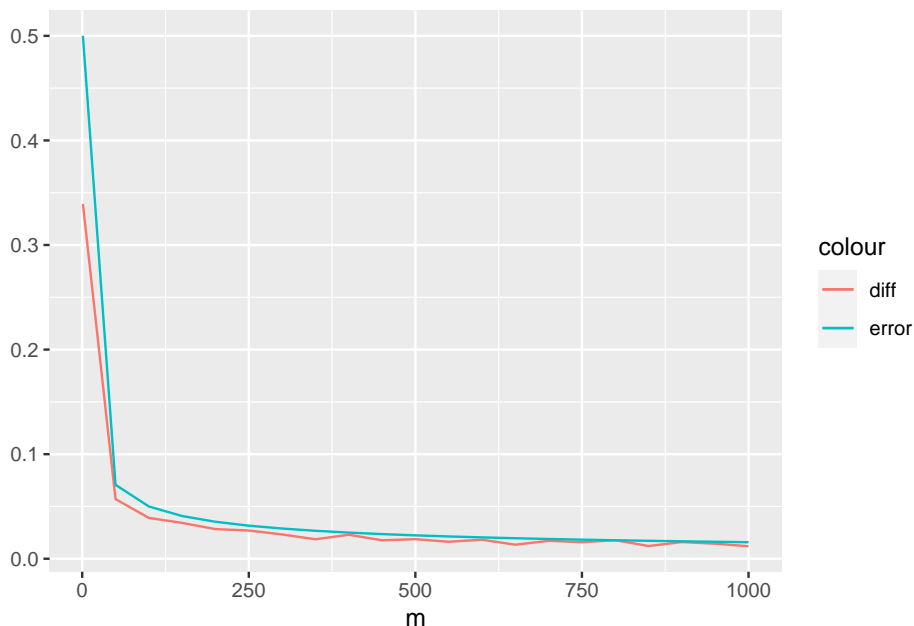
Here, we have generated 4000 (which is large) samples from Binomial(2000, 0.001) so for this sampling, binomial population mean is $=2000 \times 0.001 = 2$.

Here for the binomial population no of Bernoulli trials is 2000 that is too large and probability of success

in each trial is 0.001 that is too small and it is approximating to a Poisson distribution with parameter $a=2000*0.001=2$.

Problem 3:

```
> x= seq(-2,2,0.1)
> m= c(1, seq(50,1000,50))
> p=0.5
> error = c()
> diff = c()
> for(i in m){
+ absdiff = c()
+ for(j in x){
+ zx = pnorm(j)
+ sum = 0
+ for(k in 1:100){
+ B = rbinom(1000,i,p)
+ SB = (B-i*p)/((i*p*(1-p))^(0.5))
+ y_x = sum(SB <= j)/length(SB)
+ sum = sum + y_x
+ }
+ myx = sum/100
+ absdiff = c(absdiff, abs(myx-zx))
+ }
+ diff = c(diff, max(absdiff))
+ error = c(error, (p^2+(1-p)^2)/(2*(i*p*(1-p))^0.5))
+ }
> df=data.frame(m,diff,error)
> d_e_plot=ggplot(df,aes(m))+geom_line(aes(y=diff,color="diff"))+geom_line(aes(y=error,
+ color="error"))+ylab("")
> d_e_plot
```



Thus, we can see that the Difference is always less or equal to the Error term. Hence, the result is verified

by the picture. To verify the Theorem we can do an additional thing that is varying the value of p . The value of p should be in $[0,1]$. Now,we can do the same for different values of p and also for different values of m .