11. Hyper geometric, Binomial, Poisson, Negative binomial, Geometric and Normal distribution

Frequency table: Discrete random variable

```r
n=100
x1=sample(1:6,n,replace = T)
x1
```

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

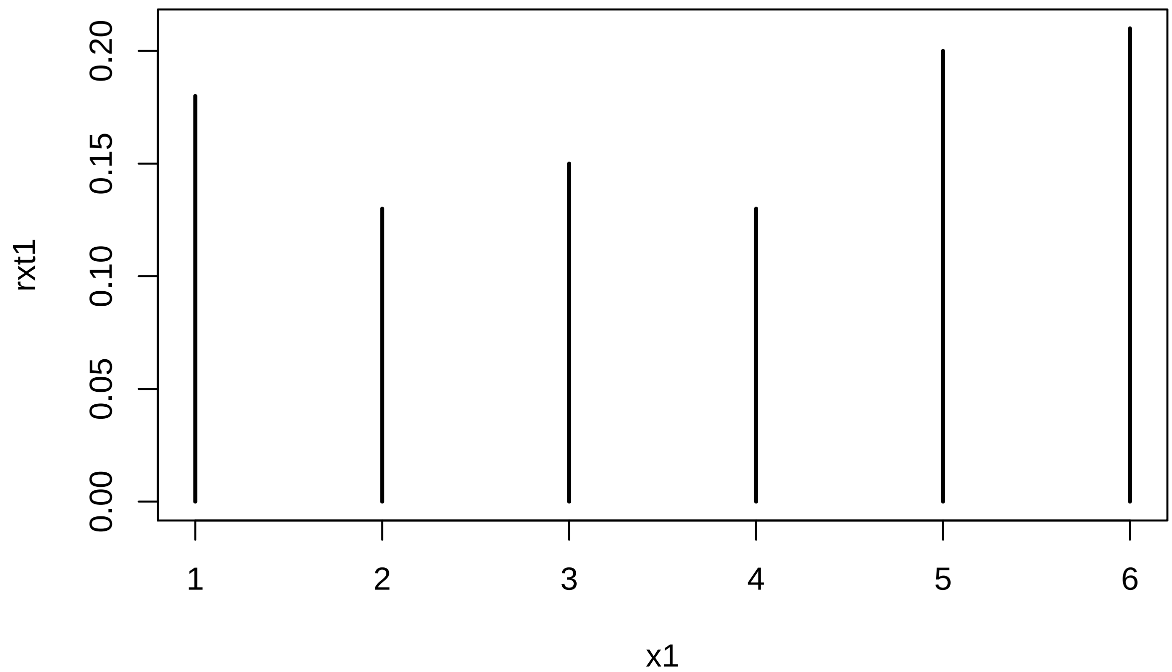
```r
xt1=table(x1)
xt1
```

```
x1
 1  2  3  4  5  6 
18 13 15 13 20 21 
```

```r
# Relative frequency table
rxt1=xt1/n
rxt1
```

```
x1
   1    2    3    4    5    6 
0.18 0.13 0.15 0.13 0.20 0.21 
```

```r
plot(rxt1)
```




```
#Find frequency table for 200 and 500 sample sizes.
#As the sample sizes increases, probability of each face approaches 1/6=0.1667
```

Cumulative distribution Function: discrete data

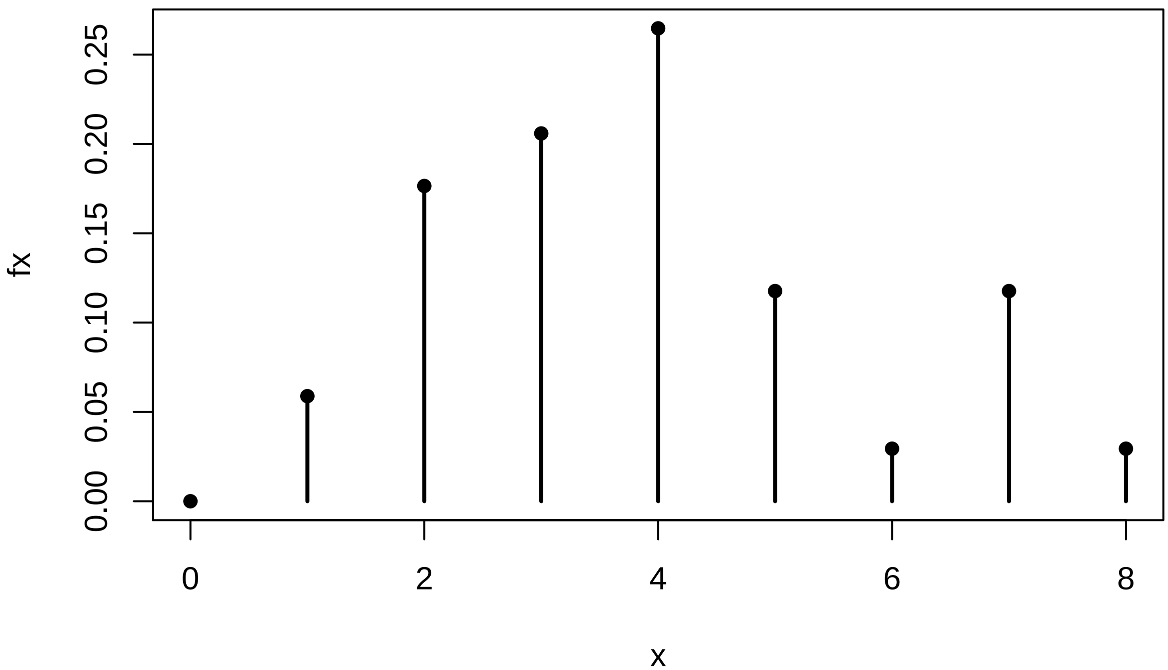
```
x=c(0:8)
x
```

```
[1] 0 1 2 3 4 5 6 7 8
```

```
fx=c(0,2/34,6/34,7/34,9/34,4/34,1/34,4/34,1/34)
fx
```

```
[1] 0.00000000 0.05882353 0.17647059 0.20588235 0.26470588 0.11764706 0.02941176
[8] 0.11764706 0.02941176
```

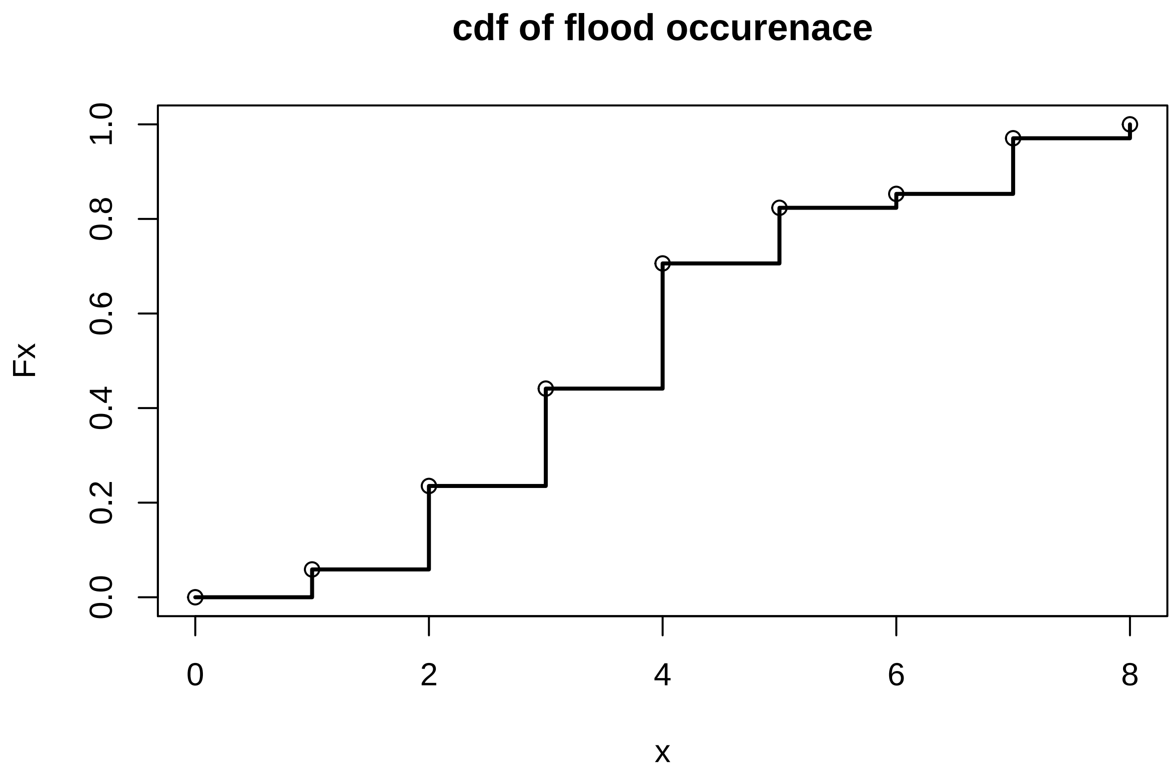
```
plot(x,fx,type='h',lwd=2,main = "")
points(x,fx,pch=16)
```




```r
Fx=cumsum(fx)
cdf=data.frame(x,Fx)
cdf
```

```
  x         Fx
1 0 0.00000000
2 1 0.05882353
3 2 0.23529412
4 3 0.44117647
5 4 0.70588235
6 5 0.82352941
7 6 0.85294118
8 7 0.97058824
9 8 1.00000000
```

```r
plot(x,Fx,type="s",main = "cdf of flood occurenace",lwd=2)
points(x,Fx)
```



Some special discrete distributions:

1. Hypergeometric distribution:

$$f(x) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}, \qquad x = 0, 1, 2, ...$$

$$mean = \frac{(m)(k)}{m+n}, \qquad variance = \frac{(mn)(m+n)(m+n-k)}{(m+n)^2(m+n-1)}$$

Ex1: A box contains 4 red & 5 white flowers. A random sample of 6 flowers is drawn without replacement from the box. Find the probability that the sample contains 3 red flowers.

```
x=3 #success for red flowers
m=4 #No. of red flowers in the population
n=5 #No. of white flowers in the population
k=6 #Sample size
ans1=(choose(4,3)*choose(5,3))/choose(9,6)
ans1
```

```
## [1] 0.4761905
```

```r
#OR
ans2=dhyper(x,m,n,k)
ans2
```

```
## [1] 0.4761905
```

Ex2: As part of a pollution survey, an investigator decides to inspect the exhaust of 8 trucks out of a company's 16 trucks. He suspects that 5 of the trucks omit excessive amounts of pollutants. What is the probability that, if his suspicion is correct, his sample will catch at least 3 of these 5 trucks?

```r
#x=3,4,5
m=5
n=11
k=8
x=3
f3=(choose(m,x)*choose(n,k-x))/choose(m+n,k)
x=4
f4=(choose(m,x)*choose(n,k-x))/choose(m+n,k)
x=5
f5=(choose(m,x)*choose(n,k-x))/choose(m+n,k)
ans2=f3+f4+f5
ans2
```

```
[1] 0.5
```

```r
#OR
ans3=1-phyper(2,m,n,k,lower.tail = T)
ans3
```

```
[1] 0.5
```

Ex3: A bag of marbles contains 5 red marbles and 16 blue marbles, sample size 13. If five red marbles are drawn from the bag, what is the resulting hypergeometric distribution?

```r
x=c(0:5)
m=5
n=16
k=13
fx=dhyper(x,m,n,k)
fx
```

```
## [1] 0.002751978 0.044719642 0.214654283 0.393532852 0.281094894 0.063246351
```

```r
pr.dist=data.frame(x,fx)
pr.dist
```

```
##   x          fx
## 1 0 0.002751978
## 2 1 0.044719642
## 3 2 0.214654283
## 4 3 0.393532852
## 5 4 0.281094894
## 6 5 0.063246351
```

```
pr.dist1=transform(pr.dist,Fx=cumsum(fx))
pr.dist1
```

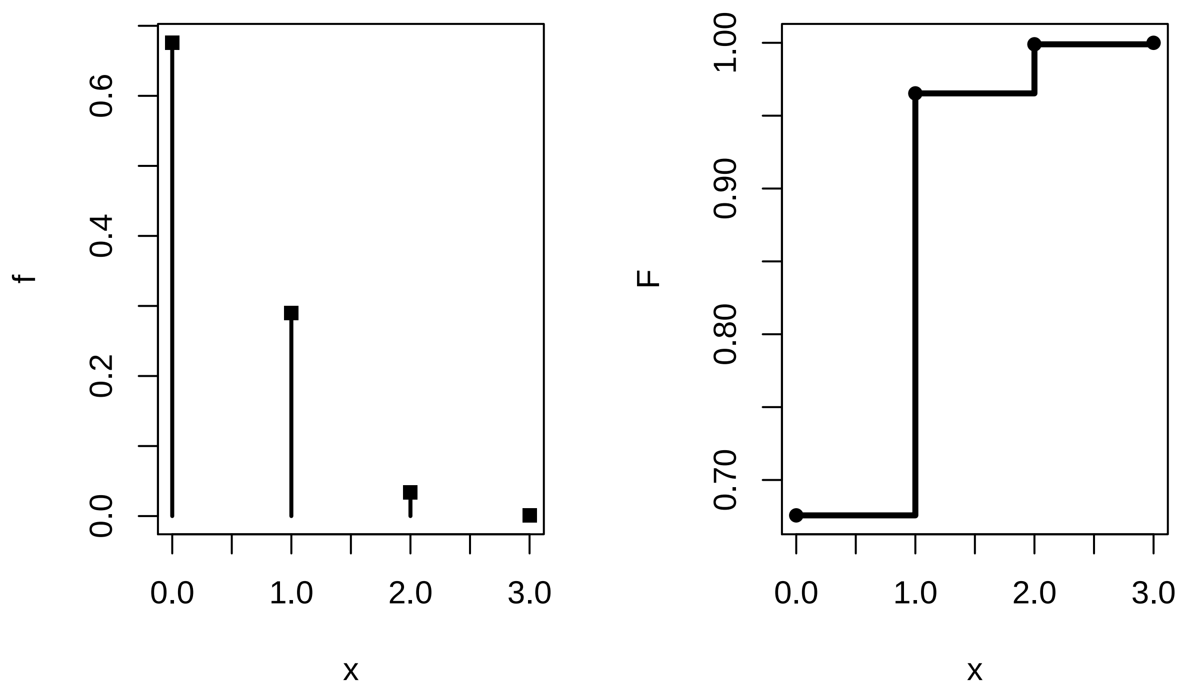
```
##   x          fx          Fx
## 1 0 0.002751978 0.002751978
## 2 1 0.044719642 0.047471620
## 3 2 0.214654283 0.262125903
## 4 3 0.393532852 0.655658755
## 5 4 0.281094894 0.936753649
## 6 5 0.063246351 1.000000000
```

Ex4: A lot of 50 chickens consists of 6 females. If 3 chickens are selected at random without replacement, plot the probability distribution and cumulative distribution function of number of female chickens in the sample.

```
x=c(0:3)
m=6
n=44
k=3
f=dhyper(x,m,n,k)
F=phyper(x,m,n,k)
fr.dist=data.frame(x,f,F)
fr.dist
```

```
##   x           f         F
## 1 0 0.675714286 0.6757143
## 2 1 0.289591837 0.9653061
## 3 2 0.033673469 0.9989796
## 4 3 0.001020408 1.0000000
```

```
par(mfrow=c(1,2))
plot(x,f,"h",lwd=2)
points(x,f,pch=15)
plot(x,F,"s",lwd=3)
points(x,F,pch=16)
```



Ex5: A taxi cab company has 12 Ambassadors and 8 Fiats. If 5 of these taxi cabs are in the shop for repairs and Ambassadors is as likely to be in for repairs as a Fiat, what is the probability that (1) 3 of them are Ambassordors (2) at least 3 of them are Ambassadors (3) at the most 2 of them are Ambassadors.

```
#(1)
m=12
n=8
k=5
x=3
#(1)
ans1=dhyper(x,m,n,k)
ans1
```

```
[1] 0.3973168
```

```
#(2)
x=3:5
ans2=sum(dhyper(x,m,n,k))
ans2
```

```
[1] 0.7038184
```

```
#(3)
x=0:2
ans3=sum(dhyper(x,m,n,k))
ans3
```

```
[1] 0.2961816
```

Random number generator

```
#To simulate a random sample from hypergeometric distribution
#rhyper(nn,m,n,k)
#First argument represents the number of data points to be generated.
#Simulate a random sample of size nn = 10 from hypergeometric distirubtion
#with parameters m=12, n=8, k=5
x=rhyper(10,12,8,5)
x
```

```
##  [1] 4 1 3 2 3 3 3 2 3 5
```

```
new=dhyper(x,m,n,k)
new
```

```
##  [1] 0.25541796 0.05417957 0.39731682 0.23839009 0.39731682 0.39731682
##  [7] 0.39731682 0.23839009 0.39731682 0.05108359
```

2. Binomial distribution:

$f(x) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0, 1, 2, ...$

$mean = np, \qquad variance = npq$

Ex1: According to the Mendelian theory of inheritance, a cross fertilization of related species of red and white flowered plants produces offspring of which 25% are red flowered plants.Suppose that a horticulture wishes to cross 5 pairs of red and white flowered plants. Of the 5 offspring, what is the probability that (1) there will be no red flowered plants? (2) there will be 4 or more red flowered plants?

```
n=5
p=0.25
#(1)
p1=dbinom(0,n,p)
p1
```

```
[1] 0.2373047
```

```r
#(2)
p2=dbinom(4,n,p)+dbinom(5,n,p)
p2
```

```
[1] 0.015625
```

```r
#OR
ans2=1-pbinom(3,n,p,lower.tail = T)
ans2
```

```
[1] 0.015625
```

Ex2: A university found that 20% of its students withdraw without completing the introductory statistics course. Assume that 20 students register for the course this quarter. (1) Compute the probability that two or fewer will withdraw. (2) Compute the probability that exactly four will withdraw. (3) Compute the probability that more than three will withdraw. (4) Compute the expected number of withdrawls.Also, find standard deviation.

```r
n=20
p=0.20
#(1)
x=0:2
pbinom(2,n,p)
```

```
[1] 0.2060847
```

```r
#(2)
dbinom(4,n,p)
```

```
[1] 0.2181994
```

```r
#(3)
1-pbinom(3,n,p,lower.tail = TRUE)
```

```
[1] 0.5885511
```

```r
#(4)
mn=n*p
mn
```

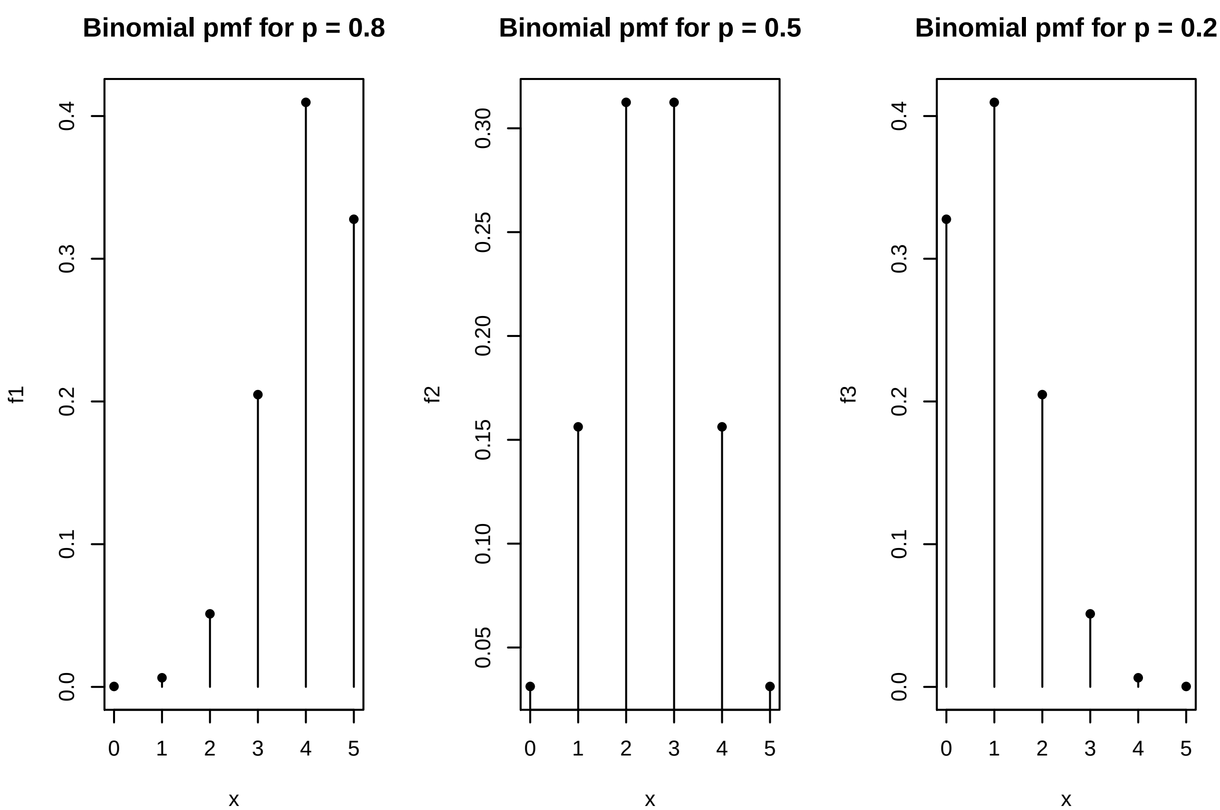
```
[1] 4
```

```r
sd=sqrt(n*p*(1-p))
sd
```

```
[1] 1.788854
```

Ex3: (1) consider 3 binomial distributions with n = 5 and p = 0.8, 0.5 and 0.2. (2) Plot spike plots to represent pmf of the three distributions. (3) Draw random sample of size 100 form each of the three distributions.(4) and plot spike plots for the relatiove frequency distributions.

```
x=c(0:5)
f1=dbinom(x,5,0.8)
f2=dbinom(x,5,0.5)
f3=dbinom(x,5,0.2)
par(mfrow=c(1,3))
plot(x,f1,"h", main = "Binomial pmf for p = 0.8")
points(x,f1,pch=16)
plot(x,f2,"h", main = "Binomial pmf for p = 0.5")
points(x,f2,pch=16)
plot(x,f3,"h", main = "Binomial pmf for p = 0.2")
points(x,f3,pch=16)
```



Ex4: We simulate the three binomial distributions for p = 0.8, 0.5 and 0.2. for the same value of n.

```
x=rbinom(100,5,0.8)
x1=sort(unique(x))
xt=table(x)
```

```r
rxt=xt/length(x)
rxt
```

```
x
   0    2    3    4    5 
0.01 0.05 0.15 0.42 0.37 
```

```r
y=rbinom(100,5,0.5)
y
```

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

```r
y1=sort(unique(y))
yt=table(y)
ryt=yt/length(y)
ryt
```

```
y
   0    1    2    3    4    5 
0.03 0.12 0.35 0.29 0.17 0.04 
```

```r
z=rbinom(100,5,0.2)
z
```

```
  [1] 2 2 0 2 0 1 0 1 1 2 2 1 1 0 1 2 1 0 0 1 0 1 1 3 0 1 2 1 1 0 0 1 0 0 1 0 1
 [38] 0 0 1 2 1 1 1 1 1 1 0 2 0 0 0 1 2 0 2 1 0 0 2 1 0 0 1 3 0 0 1 0 0 3 1 1 2
 [75] 1 2 1 2 0 0 2 0 1 2 1 2 2 0 1 0 0 2 1 0 1 1 3 1 1 2
```

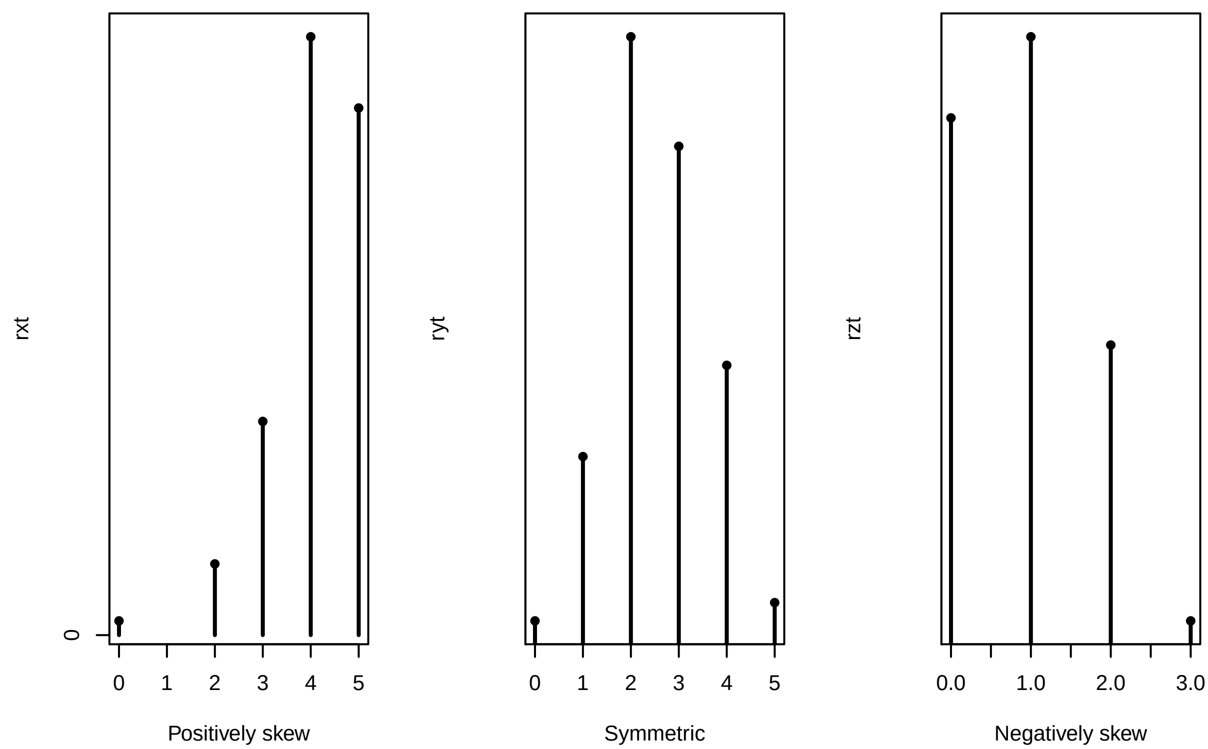
```r
z1=sort(unique(z))
zt=table(z)
rzt=zt/length(z)
rzt
```

```
z
   0    1    2    3 
0.35 0.40 0.21 0.04 
```

```r
par(mfrow=c(1,3))
plot(x1,rxt,"h",lwd=2,xlab="Positively skew")
points(x1,rxt,pch=16)

plot(y1,ryt,"h",lwd=2,xlab="Symmetric")
points(y1,ryt,pch=16)

plot(z1,rzt,"h",lwd=2,xlab="Negatively skew")
points(z1,rzt,pch=16)
```



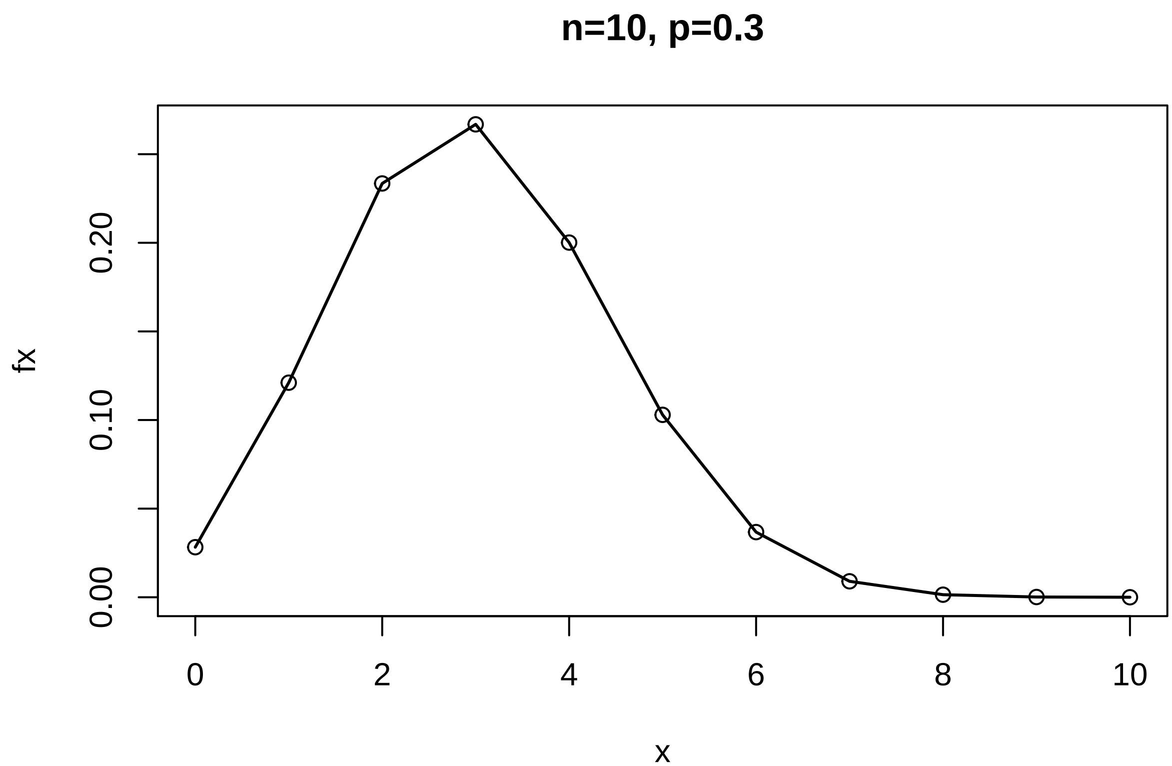

```
#The distribution is positively skew for p = 0.8, symmetric for p = 0.5,Negatively skew for p = 0.2
```

Ex5: For n = 10, 20, 50, 99. Plot pmf of binomial distribution for p = 0.3.

```
x=0:10
fx=dbinom(x,10,0.3)
fx
```

```
 [1] 0.0282475249 0.1210608210 0.2334744405 0.2668279320 0.2001209490
 [6] 0.1029193452 0.0367569090 0.0090016920 0.0014467005 0.0001377810
[11] 0.0000059049
```

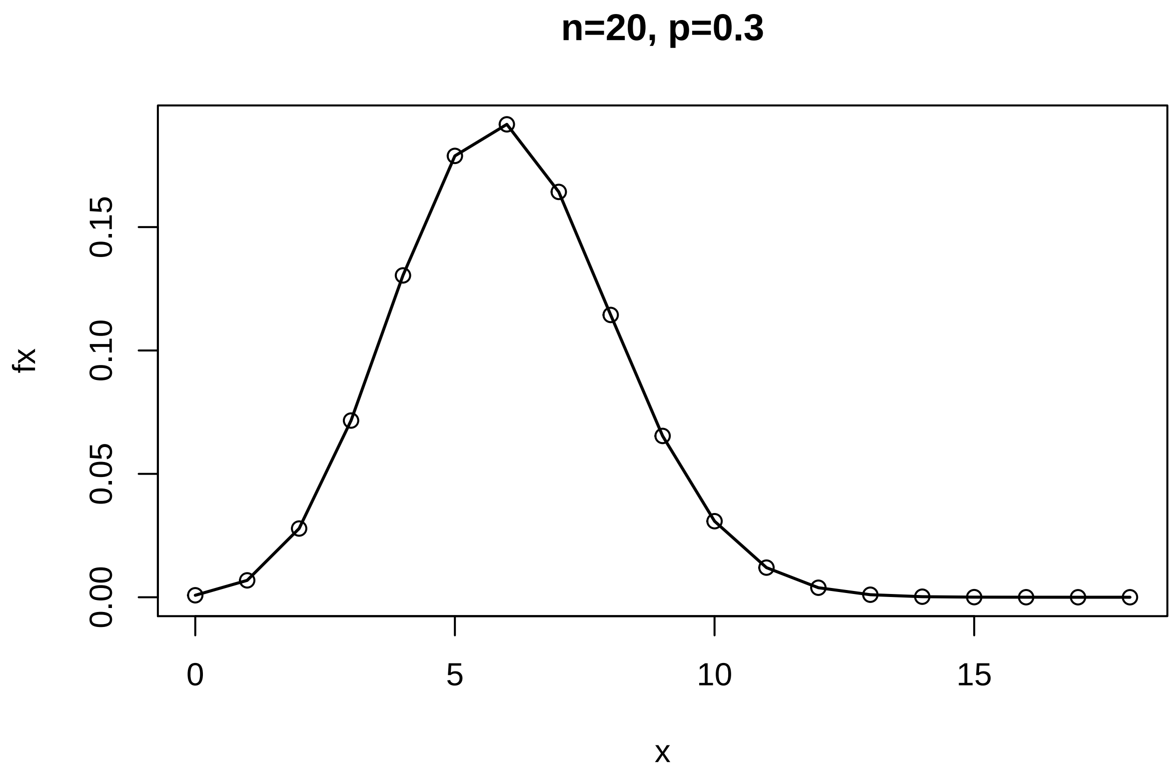
```
plot(x,fx,"l",lwd=1.5,main ="n=10, p=0.3")
points(x,fx,pch=1)
```




```
x=0:18
fx=dbinom(x,20,0.3)
fx
```

```
 [1] 7.979227e-04 6.839337e-03 2.784587e-02 7.160367e-02 1.304210e-01
 [6] 1.788631e-01 1.916390e-01 1.642620e-01 1.143967e-01 6.536957e-02
[11] 3.081708e-02 1.200665e-02 3.859282e-03 1.017833e-03 2.181070e-04
[16] 3.738977e-05 5.007558e-06 5.049639e-07 3.606885e-08
```

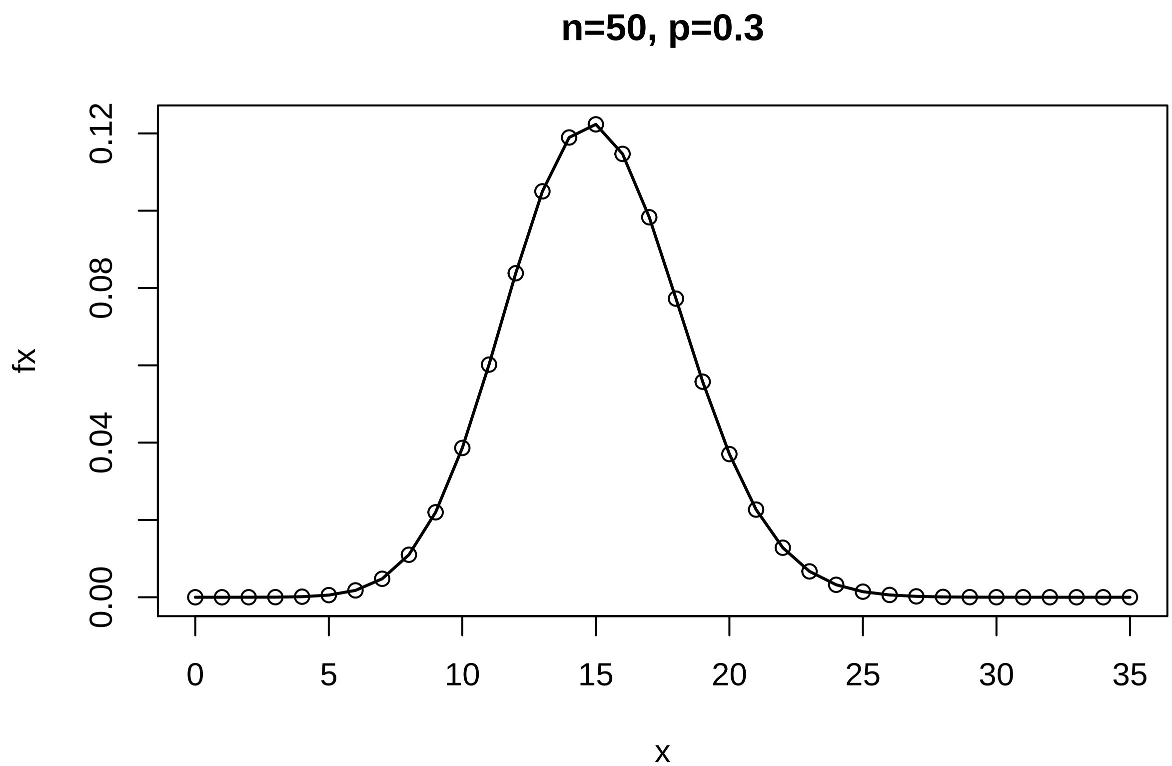
```
plot(x,fx,"l",lwd=1.5,main ="n=20, p=0.3")
points(x,fx,pch=1)
```



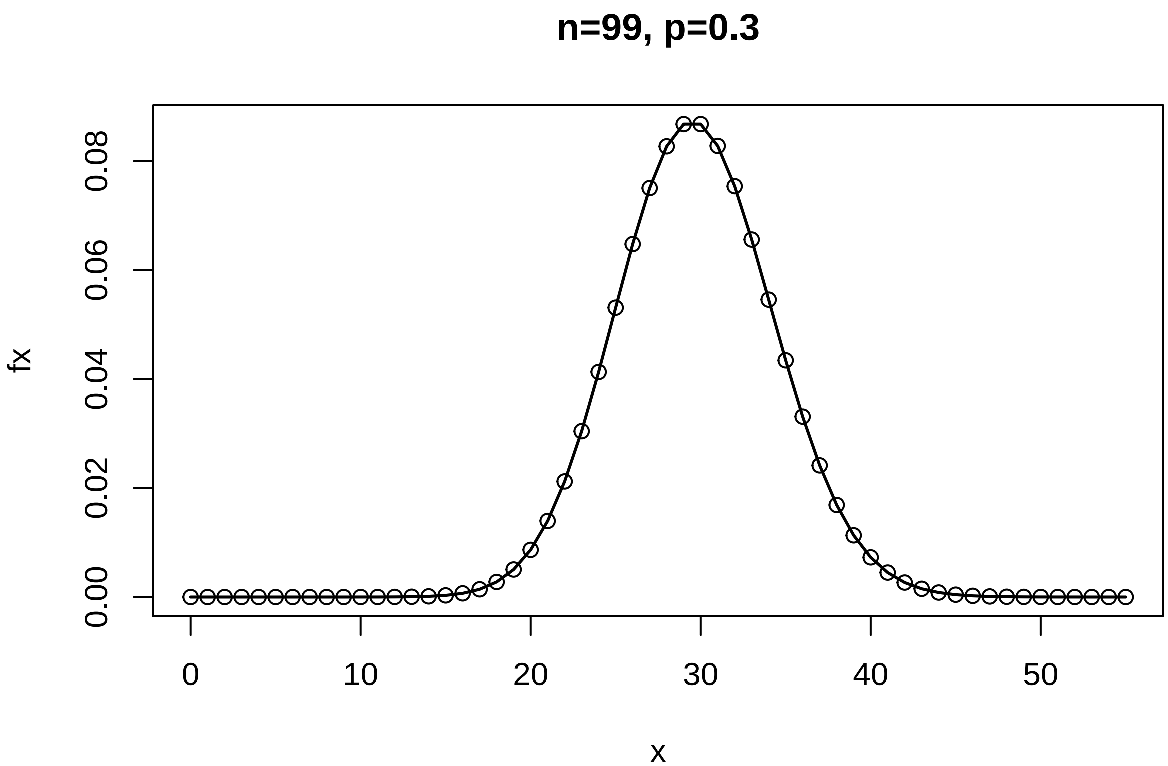

```
x=0:35
fx=dbinom(x,50,0.3)
fx
```

```
 [1] 1.798465e-08 3.853854e-07 4.046546e-06 2.774775e-05 1.397297e-04
 [6] 5.509343e-04 1.770860e-03 4.770481e-03 1.098914e-02 2.197829e-02
[11] 3.861899e-02 6.018544e-02 8.382972e-02 1.050175e-01 1.189483e-01
[16] 1.223469e-01 1.147002e-01 9.831444e-02 7.724706e-02 5.575728e-02
[21] 3.703876e-02 2.267679e-02 1.281092e-02 6.683956e-03 3.222622e-03
[26] 1.436369e-03 5.919101e-04 2.254896e-04 7.938153e-05 2.580877e-05
[31] 7.742632e-06 2.140820e-06 5.447622e-07 1.273470e-07 2.728865e-08
[36] 5.346347e-09
```

```
plot(x,fx,"l",lwd=1.5,main ="n=50, p=0.3")
points(x,fx,pch=1)
```




```
x=0:55
fx=dbinom(x,99,0.3)
plot(x,fx,"l",lwd=1.5,main ="n=99, p=0.3")
points(x,fx,pch=1)
```




```
#Figure reveals the effect of increasing values of the parameter n.
#As n increases the distribution becomes more and more symmetric.
```

3. Poisson distribution:

$$f(x) = \frac{e^{-}\lambda(\lambda)^x}{x!}, \qquad x = 0, 1, 2, ...$$

$$mean = \lambda, \qquad variance = \lambda$$

Ex1: If 5% of the electric bulbs manufactured by a company are defective. Use Poisson distribution to find the probability that in a sample of 100 bulbs (i) none is defective (ii) 5 bulbs are defective bulbs. (iii) at the most 2 defective bulbs.

```
p=0.05
n=100
l=n*p
l
```

```
[1] 5
```

```r
#(i)
dpois(0,l)
```

```
[1] 0.006737947
```

```r
#(ii)
dpois(5,l)
```

```
[1] 0.1754674
```

```r
#(iii)
ppois(2,l)
```

```
[1] 0.124652
```

Ex2: A manufacturer of copper pins knows that 2% of his product is defective. if he sells copper pins in boxes of 200 and guarantees that not more than 5 pins will be defective. what is the probability that a box will fail to meet the guaranteed?

```r
n=200
p=0.02
l=n*p
l
```

```
[1] 4
```

```r
1-ppois(5,l)
```

```
[1] 0.2148696
```

4. Negative binomial distribution:

$f(x) = \binom{n+x-1}{n-1} p^n q^x, \qquad x = 0, 1, 2, ...$

$$mean = \frac{nq}{p}, \qquad variance = \frac{nq}{p^2}$$

Ex1: The probability that a person can hit a target in any trial is 2/3. Find the probability that he will hit the target fourth time at the ninth trial.

```r
p=2/3
n=4
#n+x=9
x=5
dnbinom(x,n,p)
```

```
[1] 0.04552152
```

Ex2: The probability that a person can hit a target is 0.6. He is to be given a prize when he hits the target for the fourth time. Find the probability that he will require more than 8 trials to obtain the prize.

```
p=0.6
n=4
#n+x>=9
#x>=5 means p(5)+p(6)+p(7)+......
x=4
1-pnbinom(x,n,p)
```

```
[1] 0.1736704
```

5. Geometric distribution:

$f(x) = pq^x, \qquad x = 0, 1, 2, ...$

$mean = \frac{q}{p}, \qquad variance = \frac{q}{p^2}$

Ex1: 80% of the bulbs are defective in a big lot of bulbs. Bulbs are inspected one after the other from the lot. Find the probability that the first non-defective bulb will be otained on testing the tenth bult.

```
q=0.8 # Probability for defective bulb
p=1-q # Probability for non-defective bulb
x=9
dgeom(x,p)
```

```
[1] 0.02684355
```

Ex2: A die is thrown and getting 5 is regarded as success. Find the probability that more than 4 trials will be required before getting first success.

```
p=1/6
#x>=5 means 1-[p(0)+p(1)+p(2)+p(3)]
1-pgeom(3,p)
```

```
[1] 0.4822531
```

Continuous distribution:

Normal distribution:

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$

$mean = \mu, \qquad variance = \sigma^2$

Ex1: Normal distribution

```
x= rnorm(20,5,2)
x
```

```
 [1] 6.2989901 1.2571225 5.7532017 7.1149506 4.4517344 4.1852188 5.0985058
 [8] 0.5250631 6.1939099 3.3550470 9.9393480 4.3772720 3.8575346 4.3432568
[15] 9.7855419 4.3660742 4.1027876 5.2770150 4.7065910 5.1612780
```

```
smean=mean(x)
smean
```

```
[1] 5.007522
```

```
smedian=median(x)
smedian
```

```
[1] 4.579163
```

```
ssd=sd(x)
ssd
```

```
[1] 2.255809
```

```
#Mean and median are close, indicating symmetry of the distribution.
#The value of sample mean is close to the value of population mean.
#The value of sample sd is close to the value of population sd.
```

Ex2: Suppose X follows Standard Normal distribution. Determine the following probabilities (1) p[X<=2] (2) p[0.84<=X<=2.5] (3) p[X>=2].

```
pnorm(2)
```

```
[1] 0.9772499
```

```
pnorm(2.5)-pnorm(0.84)
```

```
[1] 0.1942445
```

```
1-pnorm(2)
```

```
[1] 0.02275013
```

Ex3: An intelligent test is conducted for 1000 children and it is found that the average marks are 42 and standard deviations is 24. If the marks obtained by the children is normally distributed, then (1) find the number of children getting marks more than 50.(2) Find the percentage of children getting marks between 30 and 54. (3) Find the minimum score of most intelligent 100 children.

```
N=1000
mu=42
sd=24
#Converting into standard normal variate
#(1) find the number of children getting marks more than 50.
x1=50
z1=(x1-mu)/sd
round(z1,2)
```

```
[1] 0.33
```

```
#p[Z>=z1]=p[Z>=0.33]
ans1=(1-pnorm(z1))*1000
round(ans1,0)
```

```
[1] 369
```

```
#(2) Find the percentage of children getting marks between 30 and 54.
x1=30
z1=(x1-mu)/sd
round(z1,2)
```

```
[1] -0.5
```

```
x2=54
z2=(x2-mu)/sd
round(z2,2)
```

```
[1] 0.5
```

```
#p[z1<=Z<=z2]=p[-0.5<=z<=0.5]
ans2=(pnorm(0.5)-pnorm(-0.5))*1000
round(ans2,0)
```

```
[1] 383
```

```
#(3) Find the minimum score of most intelligent 100 children.
qnorm(1-(100/1000),mu,sd,lower.tail = T)
```

```
[1] 72.75724
```