18. Fitting of Hyper, Binomial and Poisson distribution

1. Hypergeometric distribution:

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

Ex1.Out of 20 packages to be dispached by a mail-room clerk eight are to be by air mail and the rest by surface mail. The package got mixed thoroughly. Five of the packages are selected randomly. The distribution of packages marked for air mail into the chosen five packages, observed over period of 100 days is given below. Fit hyper geometric distribution to the given data.

$x = (0, 1, 2, 3, 4, 5)$

```
#H0: Fit of hypergeometric distribution is good.
#H1: Fit of hypergeometric distribution is not good.
m = 8 #  No. of packages to be sent by air mail
n = 12 # No. of packages to be sent by surface mail
k = 5 # No. of packages to be selected
x=0:5
f=c(7,22,45,20,5,1)
px=dhyper(x,m,n,k)
px=round(px,6)
ef=round(100*px,0)
fr.dist=data.frame(x,f,ef)
fr.dist
```

$f = (7, 22, 45, 20, 5, 1)$

```
  x  f ef
1 0  7  5
2 1 22 26
3 2 45 40
4 3 20 24
5 4  5  5
6 5  1  0
```

```
o=c(c(f[1:4]),sum(f[5:6]))
o
```

```
[1]  7 22 45 20  6
```

```
e=c(c(ef[1:4]),sum(ef[5:6]))
e
```

```
[1]  5 26 40 24  5
```

```
chcal=sum((o-e)^2/e)
chcal
```

```
[1] 2.907051
```

```
df=length(o)-0-1
df
```

```
[1] 4
```

```
chtab=qchisq(0.95,df)
chtab
```

```
[1] 9.487729
```

```
if (chcal <= chtab) {
  cat("hypergeometric distribution fits the data")
} else {
  cat("hypergeometric distribution does not fit the data")
}
```

```
hypergeometric distribution fits the data
```

2. Binomial distribution:

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

Ex1: Fit a binomial distribution for the following data.

$x = (0, 1, 2, 3, 4, 5, 6, 7)$

```
#H0: Binomial distribution fits the data.
#H1: Binomial distribution does not fit the data.
x=0:7
n=max(x)
f=c(0,4,13,28,42,20,6,2)
N=sum(f)
N
```

$f = (0, 4, 13, 28, 42, 20, 6, 2)$

```
[1] 115
```

```
smean=sum(f*x/N)
smean
```

```
[1] 3.756522
```

```
p=smean/n
p
```

```
[1] 0.536646
```

```
px=dbinom(x,n,p)
px
```

```
[1] 0.004585542 0.037176082 0.129169442 0.249335117 0.288774183 0.200670961
[7] 0.077470827 0.012817846
```

```
ef=round(px*N,0)
ef
```

```
[1]  1  4 15 29 33 23  9  1
```

```
fr.dist=data.frame(x,f,px,ef)
fr.dist
```

```
  x  f          px ef
1 0  0 0.004585542  1
2 1  4 0.037176082  4
3 2 13 0.129169442 15
4 3 28 0.249335117 29
5 4 42 0.288774183 33
6 5 20 0.200670961 23
7 6  6 0.077470827  9
8 7  2 0.012817846  1
```

```
o=c(sum(f[1:3]),c(f[4:6]),sum(f[7:8]))
o
```

```
[1] 17 28 42 20  8
```

```
e=c(sum(ef[1:3]),c(ef[4:6]),sum(ef[7:8]))
e
```

```
[1] 20 29 33 23 10
```

```
chcal=sum((o-e)^2/e)
chcal
```

```
[1] 3.730333
```

```
df=length(o)-1-1
df
```

```
[1] 3
```

```
chtab=qchisq(0.95,df)
chtab
```

```
[1] 7.814728
```

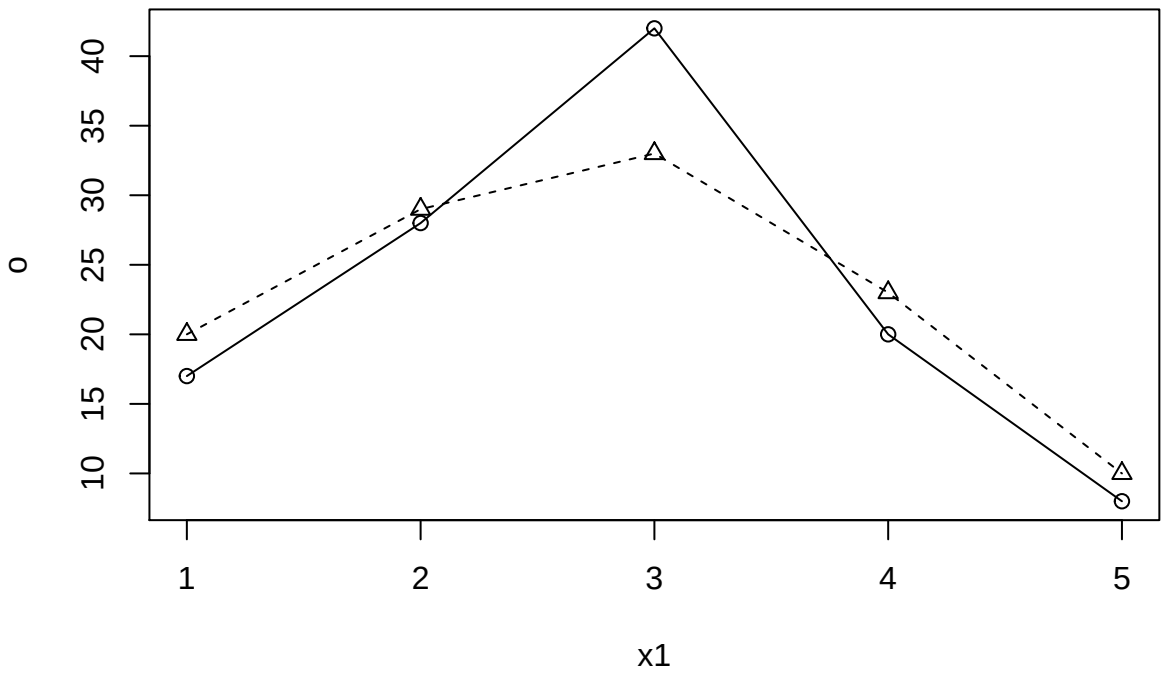
```
if (chcal <= chtab) {
  cat("Binomial distribution fits the data")
} else {
  cat("Binomial distribution does not fit the data")
}
```

```
Binomial distribution fits the data
```

```
x1=1:5
fit=data.frame(x1,o,e)
fit
```

```
  x1  o  e
1  1 17 20
2  2 28 29
3  3 42 33
4  4 20 23
5  5  8 10
```

```
plot(x1,o,"o",pch=1,lty=1)
lines(x1,e,"o",pch=2,lty=2)
```




```
#Conclusion
#This is a line throught origin with slope 1, which is obtained by abline
#If model is adequate then the points will be close to the line of reference.
#From figure binomial model seems to be adequate.
```

Ex2: Fit a binomial distribution for the following data.

$x = (0, 1, 2, 3, 4, 5, 6, 7, 8)$

```
#H0: Binomial distribution fits the data.
#H1: Binomial distribution does not fit the data.
x=0:8
n=max(x)
f=c(5,9,22,29,36,25,10,3,1)
N=sum(f)
N
```

$f = (5, 9, 22, 29, 36, 25, 10, 3, 1)$

```
[1] 140
```

```
smean=sum(f*x/N)
smean
```

```
[1] 3.557143
```

```
p=smean/n
p
```

```
[1] 0.4446429
```

```
px=dbinom(x,n,p)
px
```

```
[1] 0.009048548 0.057957259 0.162410777 0.260066131 0.260275188 0.166710024
[7] 0.066737614 0.015266574 0.001527885
```

```
ef=px*N
ef
```

```
[1]  1.2667967  8.1140163 22.7375088 36.4092584 36.4385263 23.3394033  9.3432660
[8]  2.1373204  0.2139038
```

```
fr.dist=data.frame(x,f,px,ef)
fr.dist
```

```
  x  f          px         ef
1 0  5 0.009048548  1.2667967
2 1  9 0.057957259  8.1140163
3 2 22 0.162410777 22.7375088
4 3 29 0.260066131 36.4092584
5 4 36 0.260275188 36.4385263
6 5 25 0.166710024 23.3394033
7 6 10 0.066737614  9.3432660
8 7  3 0.015266574  2.1373204
9 8  1 0.001527885  0.2139038
```

```
o=c(sum(f[1:2]),c(f[3:5]),sum(f[6:8]))
o
```

```
[1] 14 22 29 36 38
```

```
e=c(sum(ef[1:2]),c(ef[3:5]),sum(ef[6:8]))
o
```

```
[1] 14 22 29 36 38
```

```
chcal=sum((o-e)^2/e)
chcal
```

```
[1] 4.101924
```

```
df=length(o)-1-1
df
```

```
[1] 3
```

```
chtab=qchisq(0.95,df)
chtab
```

```
[1] 7.814728
```

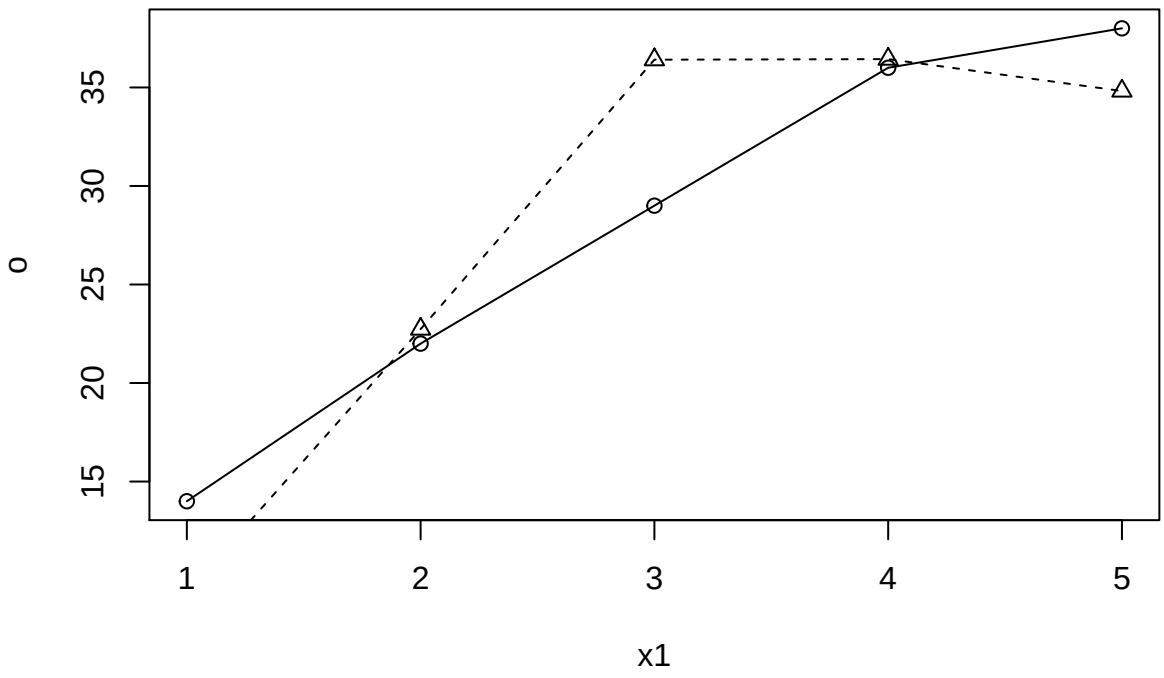
```
if (chcal <= chtab) {
  cat("Binomial distribution fits the data")
} else {
  cat("Binomial distribution does not fit the data")
}
```

```
Binomial distribution fits the data
```

```
x1=1:5
fit=data.frame(x1,o,e)
fit
```

```
  x1  o         e
1  1 14  9.380813
2  2 22 22.737509
3  3 29 36.409258
4  4 36 36.438526
5  5 38 34.819990
```

```
plot(x1,o,"o",pch=1,lty=1)
lines(x1,e,"o",pch=2,lty=2)
```




```
#Conclusion
#This is a line throught origin with slope 1, which is obtained by abline
#If model is adequate then the points will be close to the line of reference.
#From figure binomial model seems to be adequate.
```

Ex3: Fit a Binomial Distribution to the following data.

$x = (0, 1, 2, 3, 4)$

```
#H0: Fit of binomial distribution is good.
#H1: Fit of binomial distribution is not good.
x=0:4
f=c(5,20,45,20,10)
n=max(x)
N=sum(f)
smean=sum(f*x)/sum(f)
smean
```

$f = (5, 20, 45, 20, 10)$

```
[1] 2.1
```

```
p=smean/n
p
```

```
[1] 0.525
```

```
px=dbinom(x,n,p)
px=round(px,4)
ex=round(px*N,2)
ex
```

```
[1]  5.09 22.51 37.31 27.49  7.60
```

```
fr.dist=data.frame(x,f,px,ex)
fr.dist
```

```
  x  f     px    ex
1 0  5 0.0509  5.09
2 1 20 0.2251 22.51
3 2 45 0.3731 37.31
4 3 20 0.2749 27.49
5 4 10 0.0760  7.60
```

```
o=c(sum(f[1:2]),c(f[3:5]))
o
```

```
[1] 25 45 20 10
```

```
e=c(sum(ex[1:2]),c(ex[3:5]))
e
```

```
[1] 27.60 37.31 27.49  7.60
```

```
chcal=sum((o-e)^2/e)
chcal
```

```
[1] 4.628561
```

```
df=length(o)-1-1
df
```

```
[1] 2
```

```
chtab=qchisq(0.95,df)
chtab
```

```
[1] 5.991465
```

```r
if (chcal <= chtab) {
  cat("Binomial distribution fits the data")
} else {
  cat("Binomial distribution does not fit the data")
}
```

```
Binomial distribution fits the data
```

3. Poisson distribution:

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

Ex1: Fit a poisson distribution to the following data with respect to the number of red blood corpuscles(x). Find expected value also.

$x = (0, 1, 2, 3, 4, 5, 6, 7, 8)$

```r
#H0: Fit of poisson distribution is good.
#H1: Fit of poisson distribution is not good.
x=0:8
f=c(162,193,115,83,44,24,19,8,2)
l=sum(f*x)/sum(f)
l
```

$f = (162, 193, 115, 83, 44, 24, 19, 8, 2)$

```
[1] 1.775385
```

```r
px=dpois(x,l)
ex=round(px*sum(f),0)
fr.dist=data.frame(x,f,px,ex)
fr.dist
```

```
  x   f           px  ex
1 0 162 0.1694182761 110
2 1 193 0.3007826009 196
3 2 115 0.2670024011 174
4 3  83 0.1580106518 103
5 4  44 0.0701324200  46
6 5  24 0.0249024039  16
7 6  19 0.0073685575   5
8 7   8 0.0018688605   1
9 8   2 0.0004147433   0
```

```r
o=c(c(f[1:6]),sum(f[7:9]))
o
```

```
[1] 162 193 115  83  44  24  29
```

```
e=c(c(ex[1:6]),sum(ex[7:9]))
e
```

```
[1] 110 196 174 103  46  16   6
```

```
chcal=sum((o-e)^2/e)
chcal
```

```
[1] 140.7706
```

```
df=length(o)-1-1
df
```

```
[1] 5
```

```
chtab=qchisq(0.95,df)
chtab
```

```
[1] 11.0705
```

```
if (chcal <= chtab) {
  cat("Poission distribution fits the data")
} else {
  cat("Poission distribution does not fit the data")
}
```

```
Poission distribution does not fit the data
```

Ex2: For the arrival of the patients at a doctor's clinic has obtained the following distribution for 445 days. Fit a poisson distribution for the following data.

$x = (0, 1, 2, 3, 4, 5, 6)$

```
#H0: Fit of poisson distribution is good.
#H1: Fit of poisson distribution is not good.
x=0:6
f=c(153,169,72,31,12,6,2)
l=sum(f*x)/sum(f)
l
```

$f = (153, 169, 72, 31, 12, 6, 2)$

```
[1] 1.114607
```

```r
px=dpois(x,l)
ex=round(px*sum(f),0)
fr.dist=data.frame(x,f,px,ex)
fr.dist
```

```
  x   f           px  ex
1 0 153 0.3280442597 146
2 1 169 0.3656403434 163
3 2  72 0.2037725959  91
4 3  31 0.0757087697  34
5 4  12 0.0210963763   9
6 5   6 0.0047028326   2
7 6   2 0.0008736348   0
```

```r
o=c(c(f[1:4]),sum(f[5:7]))
o
```

```
[1] 153 169  72  31  20
```

```r
e=c(c(ex[1:4]),sum(ex[5:7]))
e
```

```
[1] 146 163  91  34  11
```

```r
chcal=sum((o-e)^2/e)
chcal
```

```
[1] 12.15185
```

```r
df=length(o)-1-1
df
```

```
[1] 3
```

```r
chtab=qchisq(0.95,df)
chtab
```

```
[1] 7.814728
```

```r
if (chcal <= chtab) {
  cat("Poission distribution fits the data")
} else {
  cat("Poission distribution does not fit the data")
}
```

```
Poission distribution does not fit the data
```

Ex3: Fit a Poisson Distribution to the following data.

$x = (0, 1, 2, 3, 4, 5)$

```
#H0: Fit of poisson distribution is good.
#H1: Fit of poisson distribution is not good.
x=0:5
f=c(13,24,30,18,7,8)
smean=sum(x*f)/sum(f)
smean
```

$f = (13, 24, 30, 18, 7, 8)$

```
[1] 2.06
```

```
x=0:4
px=dpois(x,2.06)
px
```

```
[1] 0.12745397 0.26255518 0.27043183 0.18569653 0.09563371
```

```
px=c(px,1-sum(px))
px
```

```
[1] 0.12745397 0.26255518 0.27043183 0.18569653 0.09563371 0.05822878
```

```
ef=round(sum(f)*px,0)
ef
```

```
[1] 13 26 27 19 10  6
```

```
x=c(x,5)
fr.dist=data.frame(x,f,px,ef)
fr.dist
```

```
  x  f         px ef
1 0 13 0.12745397 13
2 1 24 0.26255518 26
3 2 30 0.27043183 27
4 3 18 0.18569653 19
5 4  7 0.09563371 10
6 5  8 0.05822878  6
```

```
chcal=sum((f-ef)^2/ef)
chcal
```

```
[1] 2.106478
```

```r
df=length(o)-1-1
df
```

```
[1] 3
```

```r
chtab=qchisq(0.95,df)
chtab
```

```
[1] 7.814728
```

```r
if (chcal <= chtab) {
  cat("Poission distribution fits the data")
} else {
  cat("Poission distribution does not fit the data")
}
```

```
Poission distribution fits the data
```