19. Fitting of Negative binomial and geometric distribution

1. Negative binomial distribution:

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

Ex1: A blood bank collects B-negative blood samples only. The probability of getting B-negative blood is P and is treated as success. It takes only one bottle of blood from one person and purchases '5' bottles per day. The failures of 400 days before getting 5th bottle of blood of this kind were recorded as follows. Fit a negative binomial distribution to the following data.

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

```
#H0: Fit of negative binomial distribution is good.
#H1: Fit of negative binomial distribution is not good.
x=0:7
f=c(131, 131, 79, 37, 14, 5, 2, 1)
meanx=sum(x*f)/sum(f)
meanx
```

$f = (131, 131, 79, 37, 14, 5, 2, 1)$

```
[1] 1.25
```

```
vr=(sum(f*(x-meanx)^2)/sum(f))
vr
```

```
[1] 1.5625
```

```
p=meanx/vr
p
```

```
[1] 0.8
```

```
q=1-p
q
```

```
[1] 0.2
```

```
n=meanx*q/p
n
```

```
[1] 0.3125
```

```
px=dnbinom(x,1,p)
px
```

```
[1] 8.000e-01 1.600e-01 3.200e-02 6.400e-03 1.280e-03 2.560e-04 5.120e-05
[8] 1.024e-05
```

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

```
  x   f        px  ex
1 0 131 8.000e-01 320
2 1 131 1.600e-01  64
3 2  79 3.200e-02  13
4 3  37 6.400e-03   3
5 4  14 1.280e-03   1
6 5   5 2.560e-04   0
7 6   2 5.120e-05   0
8 7   1 1.024e-05   0
```

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

```
[1] 131 131 138
```

```
e=c(c(ex[1:2]),sum(ex[3:8]))
e
```

```
[1] 320  64  17
```

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

```
[1] 1043.004
```

```
df=length(o)-2-1
df
```

```
[1] 0
```

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

```
[1] 0
```

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

```
Negative binomial distribution does not fit the data
```

Ex2: Fit a Negative binomial distribution to the following data.

$x = (0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$

```r
#H0: Fit of negative binomial distribution is good.
#H1: Fit of negative binomial distribution is not good.
x=0:9
f=c(177, 87, 50, 38, 21, 7, 2, 2, 0, 1)
meanx=sum(x*f)/sum(f)
meanx
```

$f = (177, 87, 50, 38, 21, 7, 2, 2, 0, 1)$

```
[1] 1.181818
```

```r
vr=(sum(f*(x-meanx)^2)/sum(f))
vr
```

```
[1] 2.216293
```

```r
p=meanx/vr
p
```

```
[1] 0.533241
```

```r
q=1-p
q
```

```
[1] 0.466759
```

```r
n=meanx*q/p
n
```

```
[1] 1.034475
```

```r
px=dnbinom(x,1,p)
px
```

```
 [1] 0.5332409972 0.2488950361 0.1161739988 0.0542252598 0.0253101282
 [6] 0.0118137302 0.0055141649 0.0025737861 0.0012013378 0.0005607353
```

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

```
   x   f           px  ex
1  0 177 0.5332409972 205
2  1  87 0.2488950361  96
3  2  50 0.1161739988  45
4  3  38 0.0542252598  21
5  4  21 0.0253101282  10
6  5   7 0.0118137302   5
7  6   2 0.0055141649   2
8  7   2 0.0025737861   1
9  8   0 0.0012013378   0
10 9   1 0.0005607353   0
```

```
o=c(c(f[1:5]),sum(f[6:10]))
o
```

```
[1] 177  87  50  38  21  12
```

```
e=c(c(ex[1:5]),sum(ex[6:10]))
e
```

```
[1] 205  96  45  21  10   8
```

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

```
[1] 33.0856
```

```
df=length(o)-2-1
df
```

```
[1] 3
```

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

```
[1] 7.814728
```

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

```
Geometric distribution does not fit the data
```

2. Geometric distribution:

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

Ex1: A thief steals stereo-system from the parked cars. Every morning he starts searching parked cars until he gets a car with stereo system. Then, he steals the stereo-system and calls it a day. The probability that a parked car has a stereo-system is P and it is treated as success. The distribution of failures before getting a stereo-system for 100 days is obtained as

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

```
#H0: Fit of Geometric distribution is good.
#H1: Fit of Geometric distribution is not good.
x=c(0:8)
f=c(40,24,15,9,5,3,2,1,1)
meanx=sum(x*f)/sum(f)
vr=(sum(f*(x-meanx)^2)/sum(f))
vr
```

$f = (40, 24, 15, 9, 5, 3, 2, 1, 1)$

```
[1] 3.0051
```

```
p=meanx/vr
px=dgeom(x,p)
ex=round(px*sum(f),0)
fr.dist=data.frame(x,f,px,ex)
fr.dist
```

```
  x  f          px ex
1 0 40 0.475857709 48
2 1 24 0.249417150 25
3 2 15 0.130730076 13
4 3  9 0.068521162  7
5 4  5 0.035914839  4
6 5  3 0.018824486  2
7 6  2 0.009866709  1
8 7  1 0.005171560  1
9 8  1 0.002710633  0
```

```
o=c(c(f[1:4]),sum(f[5:9]))
o
```

```
[1] 40 24 15  9 12
```

```
e=c(c(ex[1:4]),sum(ex[5:9]))
e
```

```
[1] 48 25 13  7  8
```

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

```
[1] 4.252454
```

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

```
[1] 3
```

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

```
[1] 7.814728
```

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

```
Geometric distribution fits the data
```

Ex2. Fit a Geometric distribution to the following data.

$x = (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13)$

```
#H0: Fit of Geometric distribution is good.
#H1: Fit of Geometric distribution is not good.
x=c(0:13)
f=c(4990,2515,1235,635,315,161,75,44,18,12,0,0,0,0)
meanx=sum(x*f)/sum(f)
vr=(sum(f*(x-meanx)^2)/sum(f))
vr
```

$f = (4990, 2515, 1235, 635, 315, 161, 75, 44, 18, 12, 0, 0, 0, 0)$

```
[1] 1.928488
```

```
p=meanx/vr
px=dgeom(x,p)
ex=round(px*sum(f),0)
fr.dist=data.frame(x,f,px,ex)
fr.dist
```

```
    x    f           px   ex
1   0 4990 5.167261e-01 5167
2   1 2515 2.497202e-01 2497
3   2 1235 1.206833e-01 1207
4   3  635 5.832307e-02  583
5   4  315 2.818602e-02  282
6   5  161 1.362157e-02  136
7   6   75 6.582947e-03   66
8   7   44 3.181366e-03   32
9   8   18 1.537471e-03   15
10  9   12 7.430196e-04    7
11 10    0 3.590820e-04    4
12 11    0 1.735349e-04    2
13 12    0 8.386490e-05    1
14 13    0 4.052972e-05    0
```

```
o=c(c(f[1:9]),sum(f[10:14]))
o
```

```
 [1] 4990 2515 1235  635  315  161   75   44   18   12
```

```
e=c(c(ex[1:9]),sum(ex[10:14]))
e
```

```
 [1] 5167 2497 1207  583  282  136   66   32   15   14
```

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

```
[1] 26.55094
```

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

```
[1] 8
```

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

```
[1] 15.50731
```

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

```
Geometric distribution does not fit the data
```