应用统计学1

郭一懋

2022-09-14

question 1 Introduce the distributions with an example

define function to analysis the sample data

I find some example data about the distributions, and make some code (include calculating var and expect value and draw the plot of mass function and CPF) to analysis it.

there is the link of my code and database https://github.com/gym2220013626/applied-static-assignment1.git (https://github.com/gym2220013626/applied-static-assignment1.git)。

```
setwd("C:/Users/20530/Desktop/")
getwd()
```

```
## [1] "C:/Users/20530/Desktop"
```

```
myfun <- function(name){#用于测试数据的函数
  #读取
  data <- read.table(name,sep = '\n',header = TRUE)
  value = ts(data)
  plot(value,xlab = name,ylab = "value")
  #方差
  varvalue = var(value)
  print(paste('varvalue =',as.character(varvalue[1,1])))
  expectvalue = mean(value)#均值
  print(paste('expectvalue =',as.character(expectvalue)))
  #密度函数
  density(value)
  plot(density(value),xlab = name)
  #概率
  court = table(value)
  prob = court/100
  plot(prob,xlab = name)
  #累计概率
  x <- seq(0.1, 50, by=0.1)
  if(name=='exp.txt'){
    plot(x, pexp(x, 1), type="l",xlab = name,ylab ="cum")
    #browser()
  }
  else if(name=='gamma.txt'){
    plot(x, pexp(x, 1), type="l",xlab = name,ylab ="cum")
  }else{
    cum <- cumsum(prob)
    plot(cum,xlab = name,ylab ="cum")
  }
  #browser()
  return(0)
}
first = c('binom','pois','exp','gamma','hyper')
second = 'txt'
```

Binomial Distribution

In probability theory and statistics, the binomial distribution with parameters n and p is the discrete probability distribution of the number of successes in a sequence of n independent experiments, each asking a yes–no question, and each with its own Boolean-valued outcome: success (with probability p) or failure (with probability q = 1 − p). A single success/failure experiment is also called a Bernoulli trial or Bernoulli experiment, and a sequence of outcomes is called a Bernoulli process; for a single trial, i.e., n = 1, the binomial distribution is a Bernoulli distribution. The binomial distribution is the basis for the popular binomial test of statistical significance.

Probability mass function:

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Probability mass function

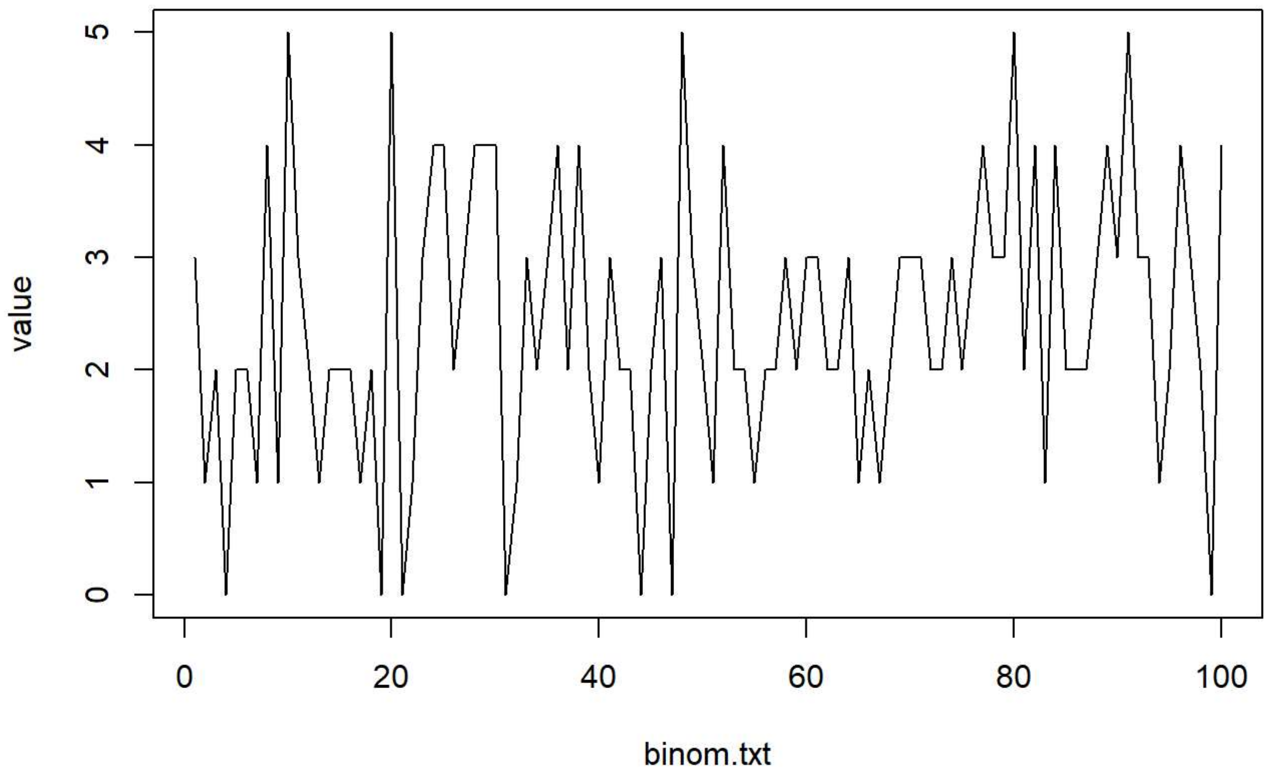
Cumulative distribution function:

$$F(k;n,p) = \Pr(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1-p)^{n-i},$$

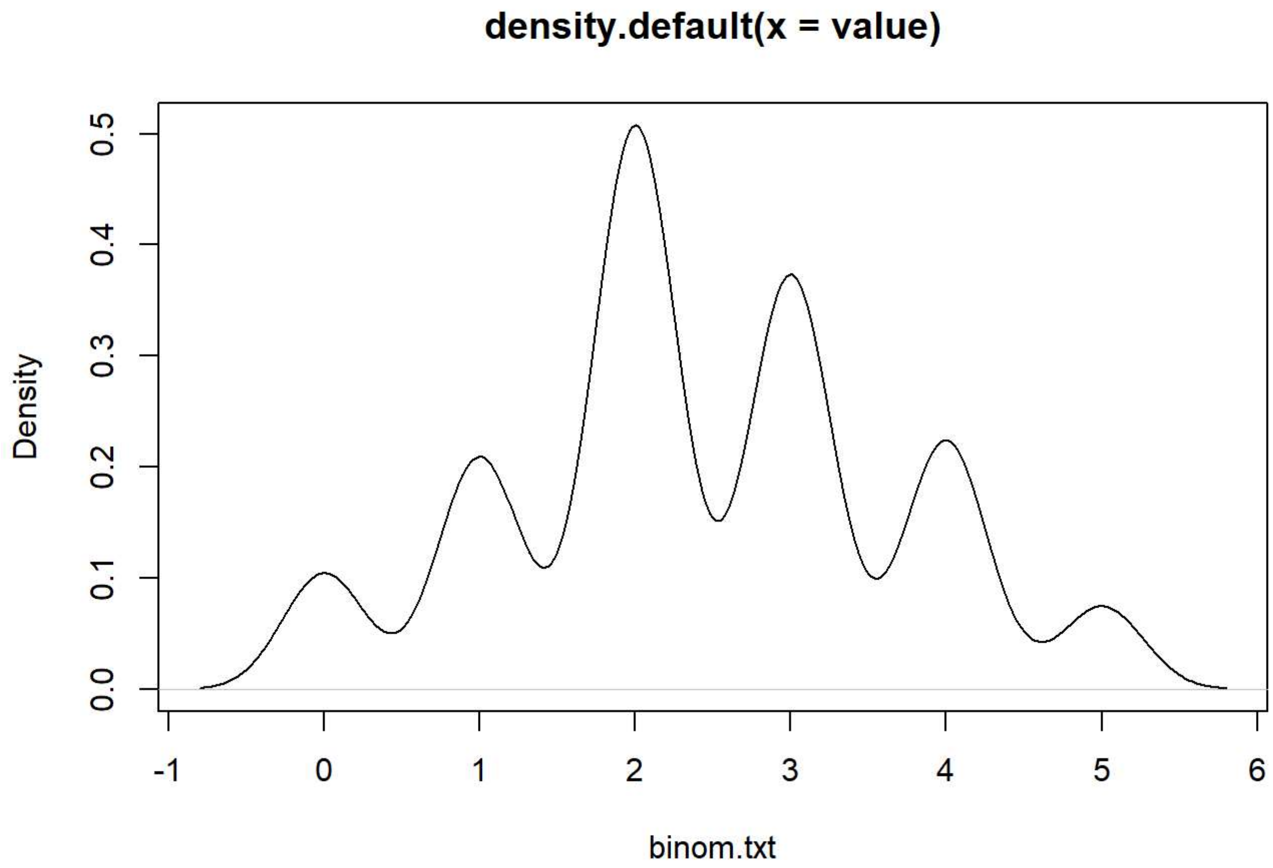
Cumulative distribution function

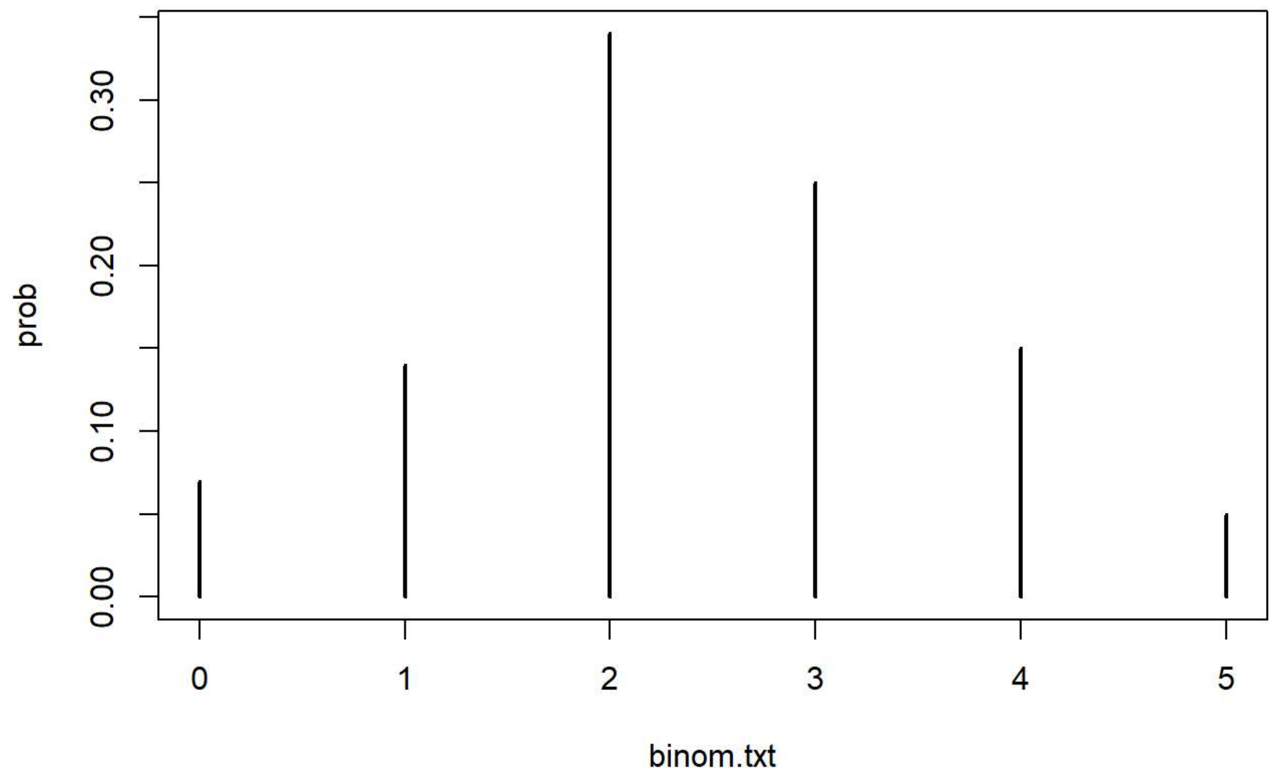
example data analysis result(coin data)

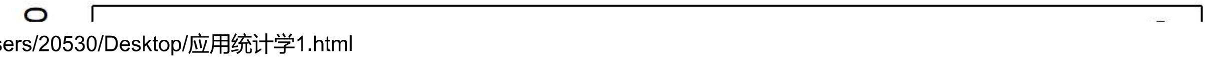
```
for (f in first) {
  #分析指定样本
  if(f=='binom'){
    name = paste(f,second,sep = '.')
    a = myfun(name = name)
  }
}
```

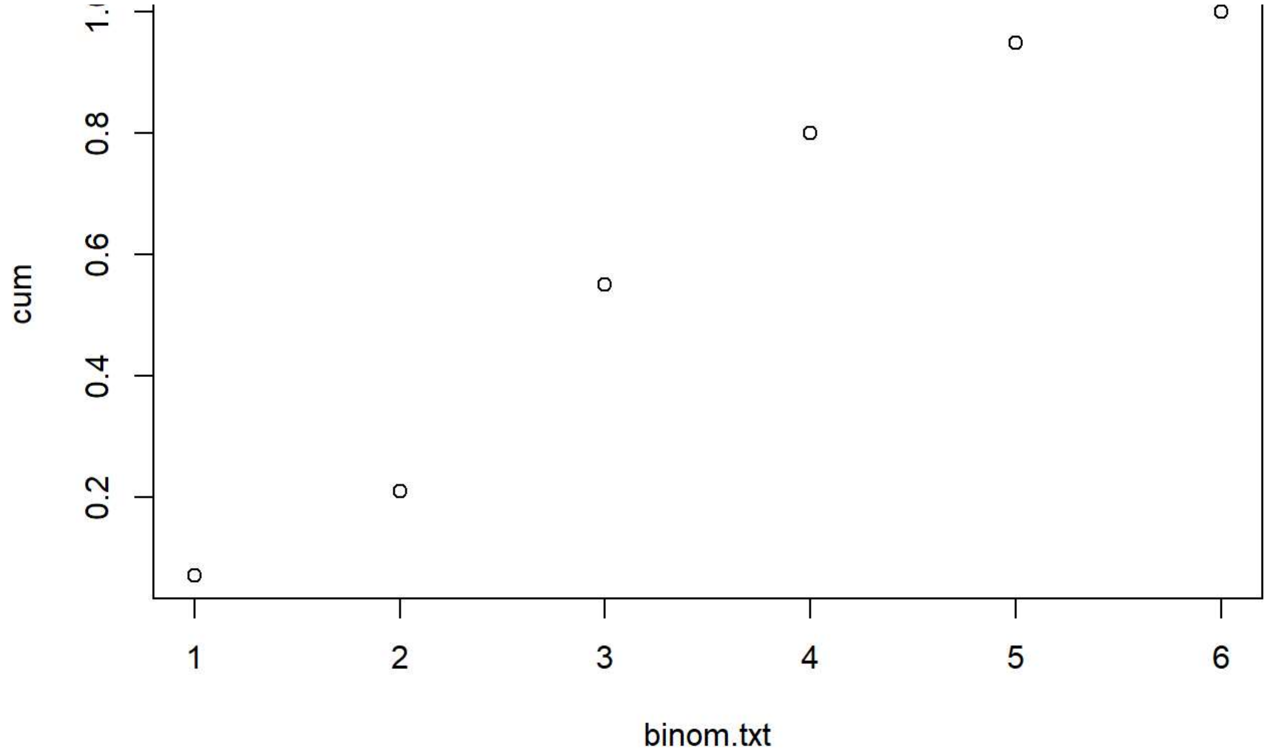



```
## [1] "varvalue = 1.55919191919192"
## [1] "expectvalue = 2.42"
```


density.default(x = value)
Density
0.0
0.1
0.2
0.3
0.4
0.5
-1
0
1
2
3
4
5
6
binom.txt


prob
0.00
0.10
0.20
0.30
0
1
2
3
4
5
binom.txt





Poisson Distribution

In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate and independently of the time since the last event.[1] It is named after French mathematician Siméon Denis Poisson (/ˈpwɑːsɒn/; French pronunciation: [pwasɔ̃]). The Poisson distribution can also be used for the number of events in other specified interval types such as distance, area or volume.

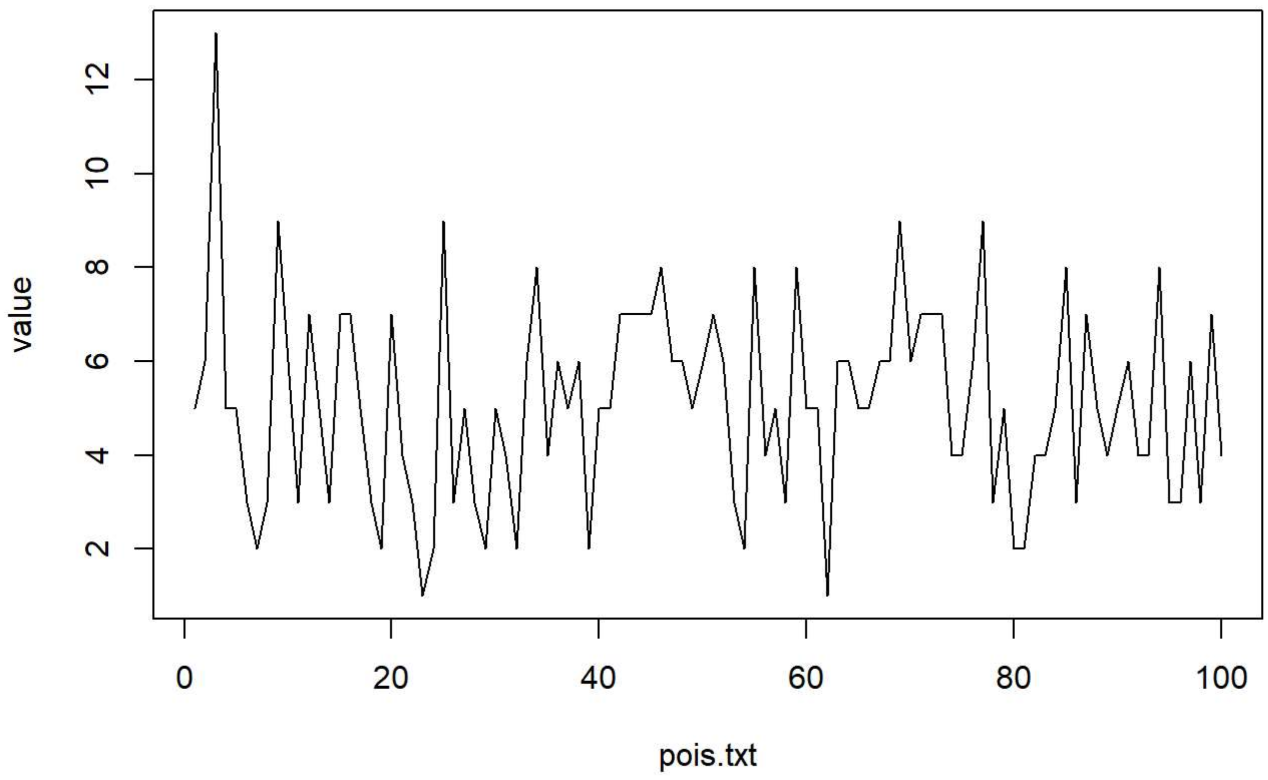
Probability mass function:

$$f(k;\lambda) = \Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

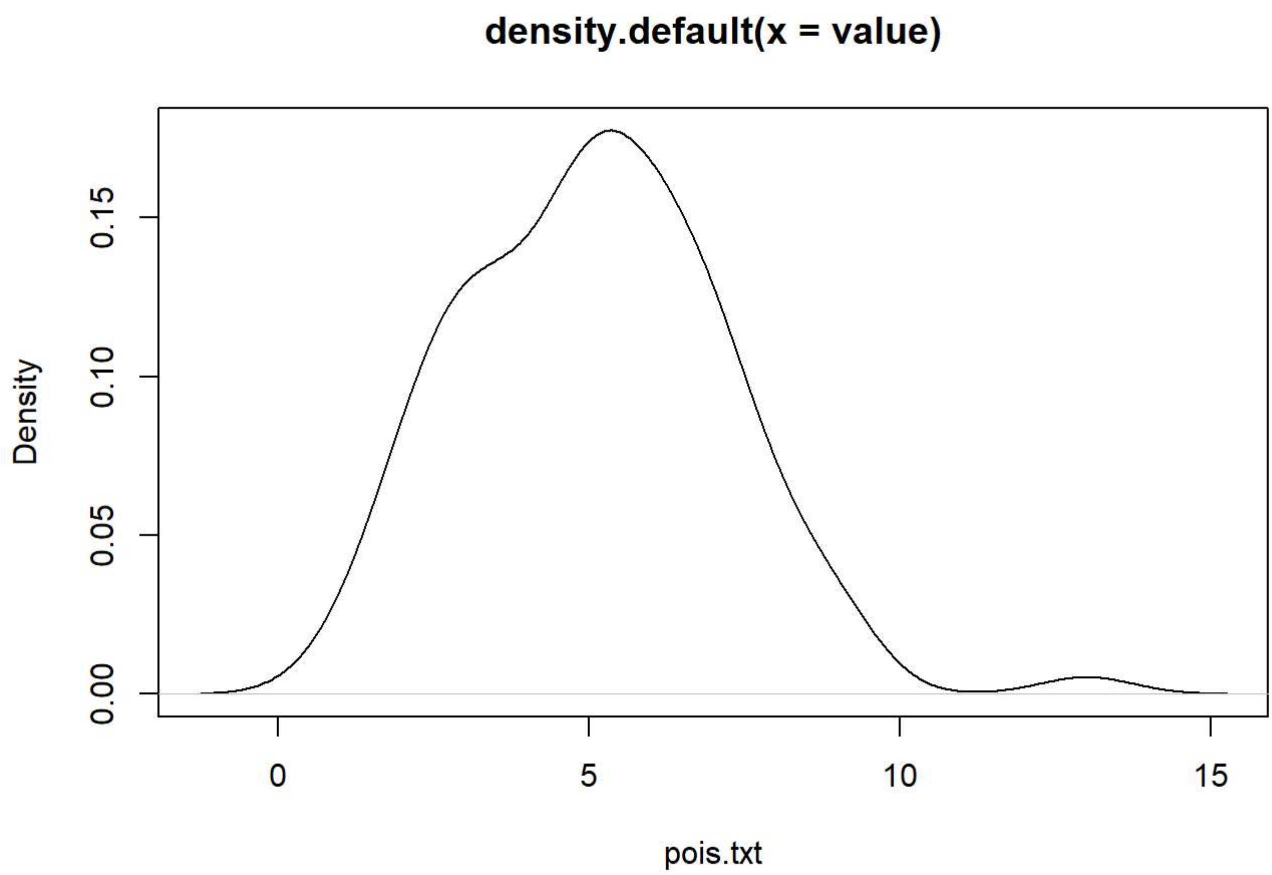
Probability mass function

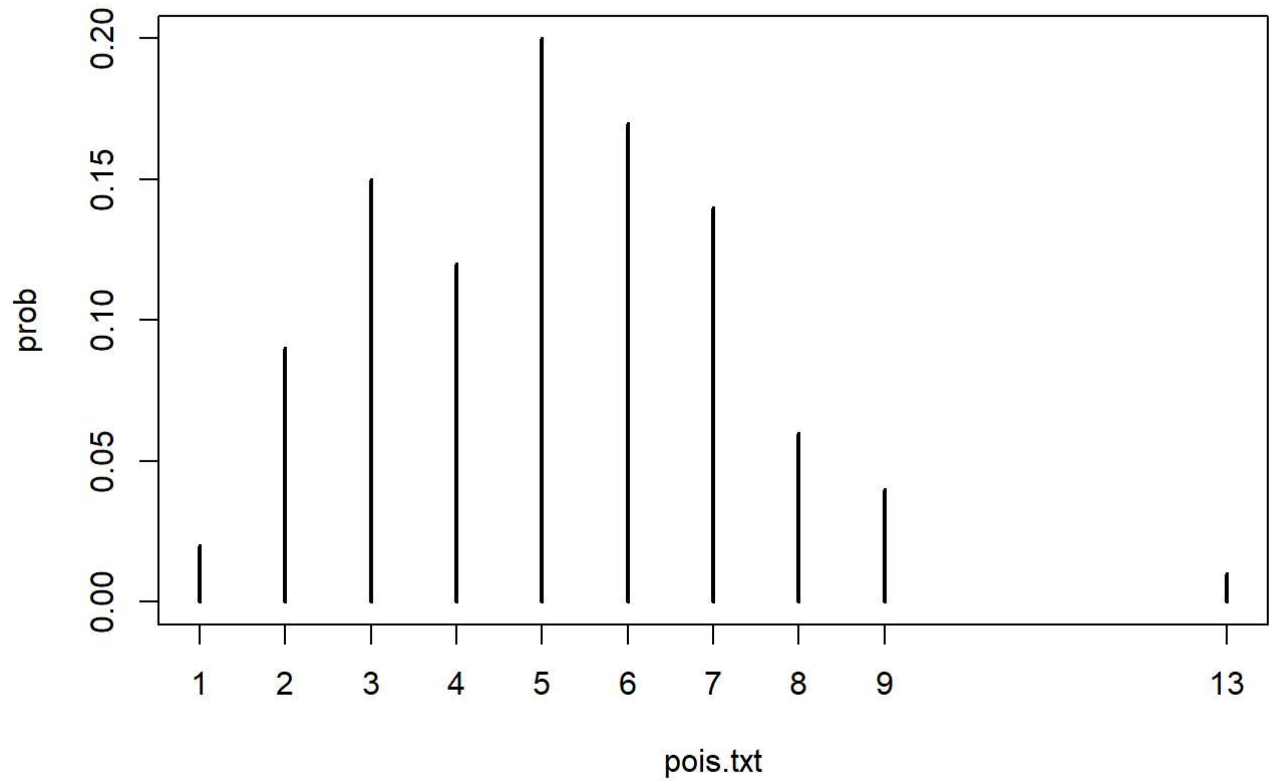
example data analysis result(laser photons data)

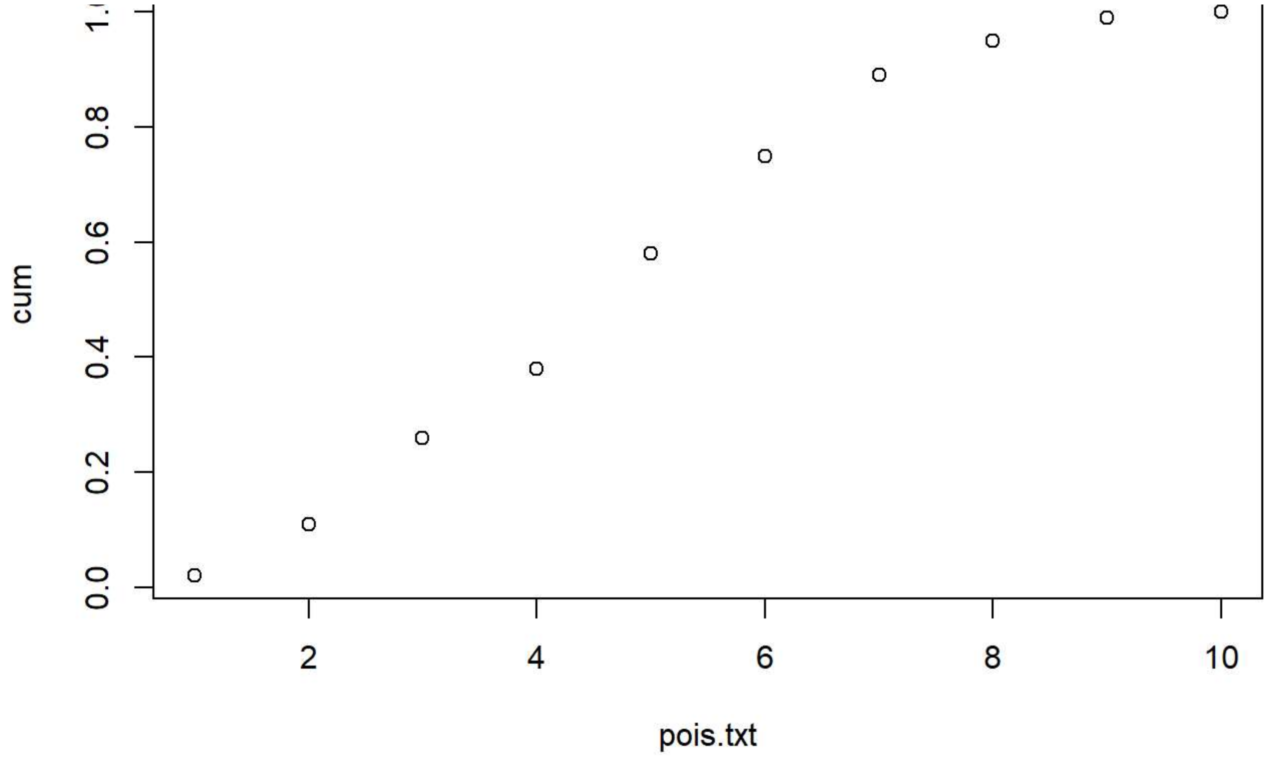
```
for (f in first) {
  #分析指定样本
  if(f=='pois'){
    name = paste(f,second,sep = '.')
    a = myfun(name = name)
  }
}
```




```
## [1] "varvalue = 4.43434343434343"
## [1] "expectvalue = 5.1"
```


density.default(x = value)
Density
0.00
0.05
0.10
0.15
0
5
10
15
pois.txt


prob
0.00
0.05
0.10
0.15
0.20
1
2
3
4
5
6
7
8
9
13
pois.txt



Hypergeometric distribution

In probability theory and statistics, the hypergeometric distribution is a discrete probability distribution that describes the probability of {k}k successes (random draws for which the object drawn has a specified feature) in {n}n draws, without replacement, from a finite population of size {N}N that contains exactly {K}K objects with that feature, wherein each draw is either a success or a failure. In contrast, the binomial distribution describes the probability of {k}k successes in {n}n draws with replacement.

Probability mass function:

$$p_X(k) = \Pr(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}},$$

Probability mass function

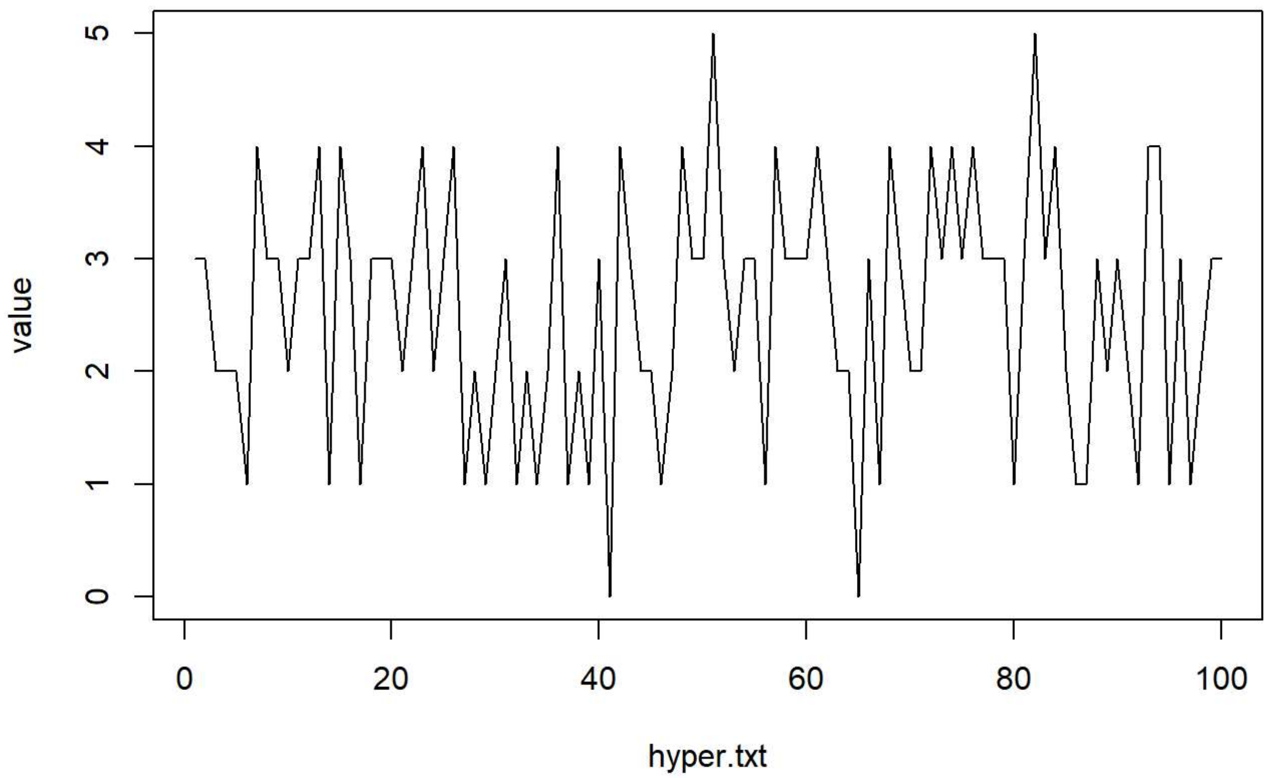
Combinatorial identities :

$$\sum_{0\le k\le min(n,K)} \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = 1,$$

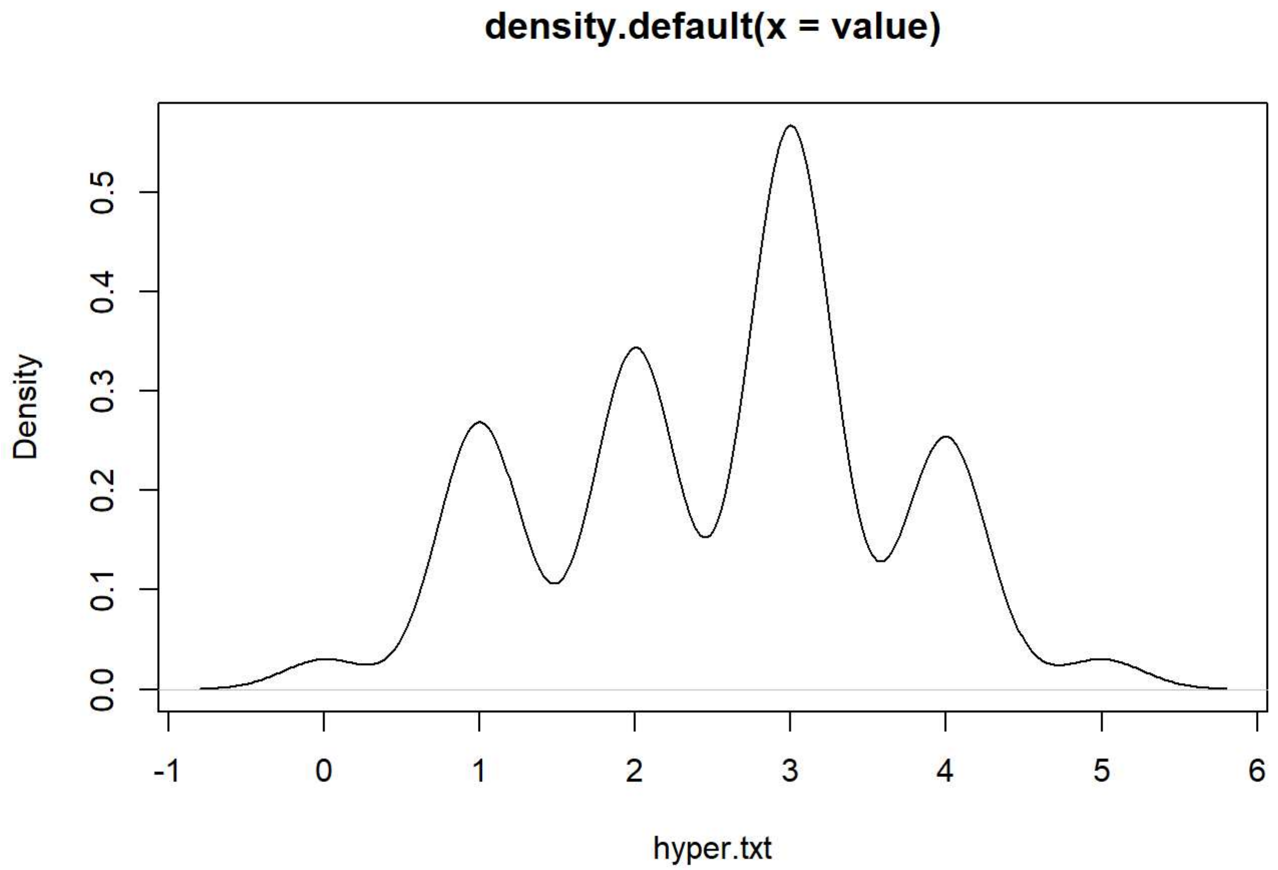
Combinatorial identities

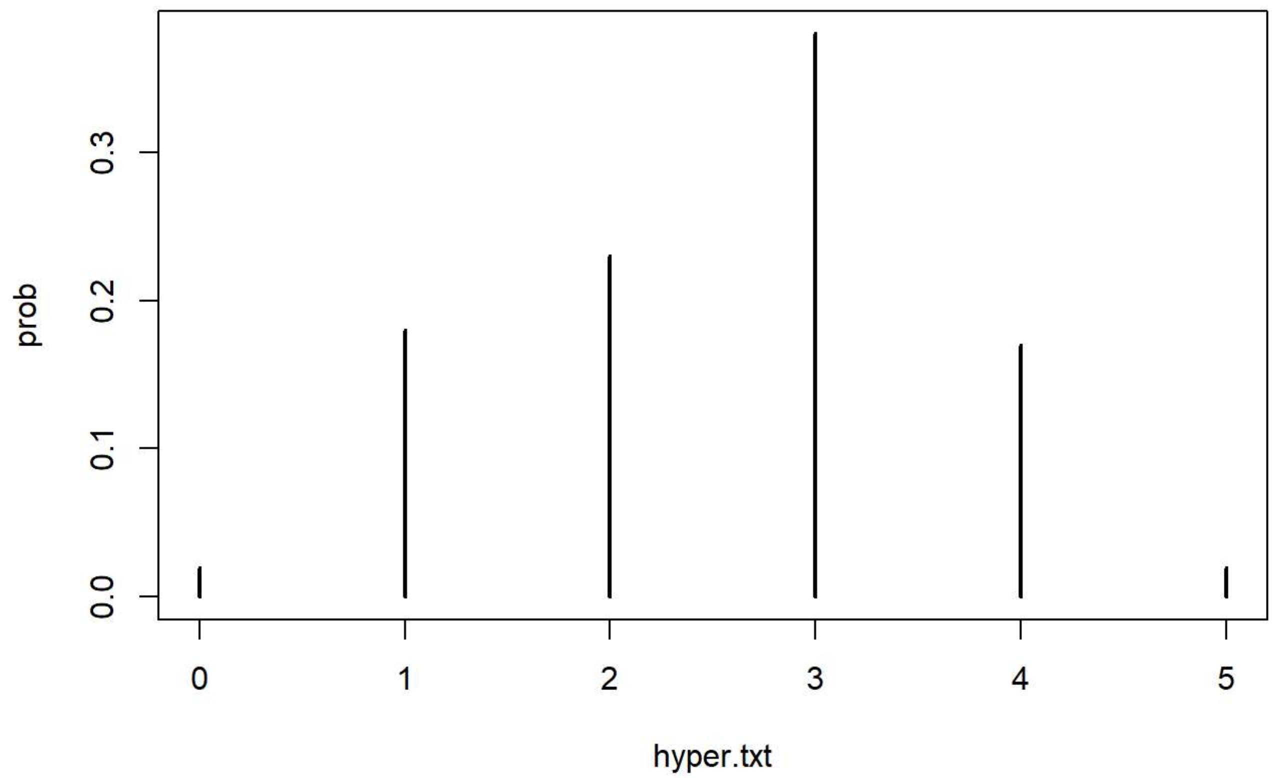
example data analysis result(marble data)

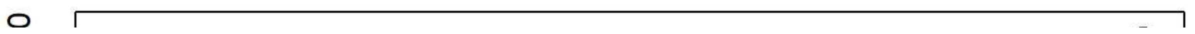
```
for (f in first) {
  #分析指定样本
  if(f=='hyper'){
    name = paste(f,second,sep = '.')
    a = myfun(name = name)
  }
}
```

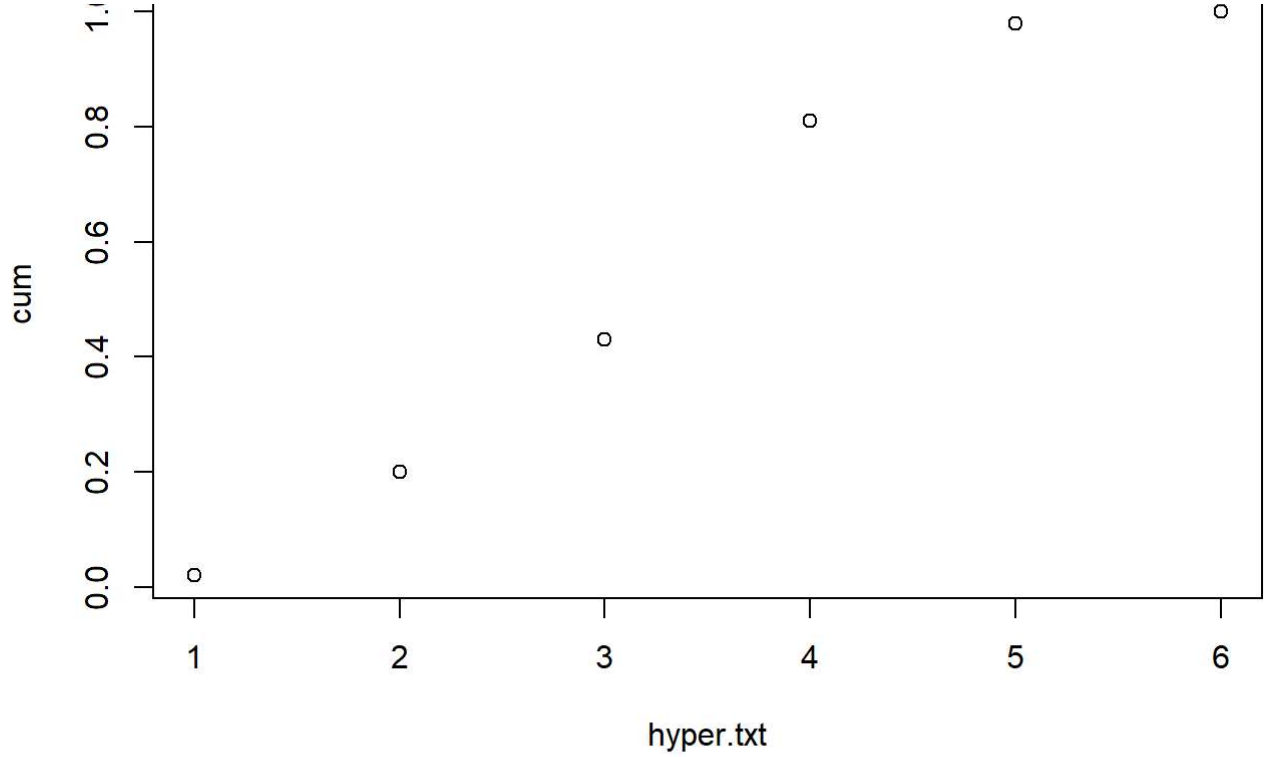



```
## [1] "varvalue = 1.19838383838384"
## [1] "expectvalue = 2.56"
```


density.default(x = value)
Density
0.0
0.1
0.2
0.3
0.4
0.5
-1
0
1
2
3
4
5
6
hyper.txt


prob
0.0
0.1
0.2
0.3
0
1
2
3
4
5
hyper.txt





Exponential distribution

In probability theory and statistics, the exponential distribution is the probability distribution of the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate. It is a particular case of the gamma distribution. It is the continuous analogue of the geometric distribution, and it has the key property of being memoryless. In addition to being used for the analysis of Poisson point processes it is found in various other contexts.

Probability mass function:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Probability mass function

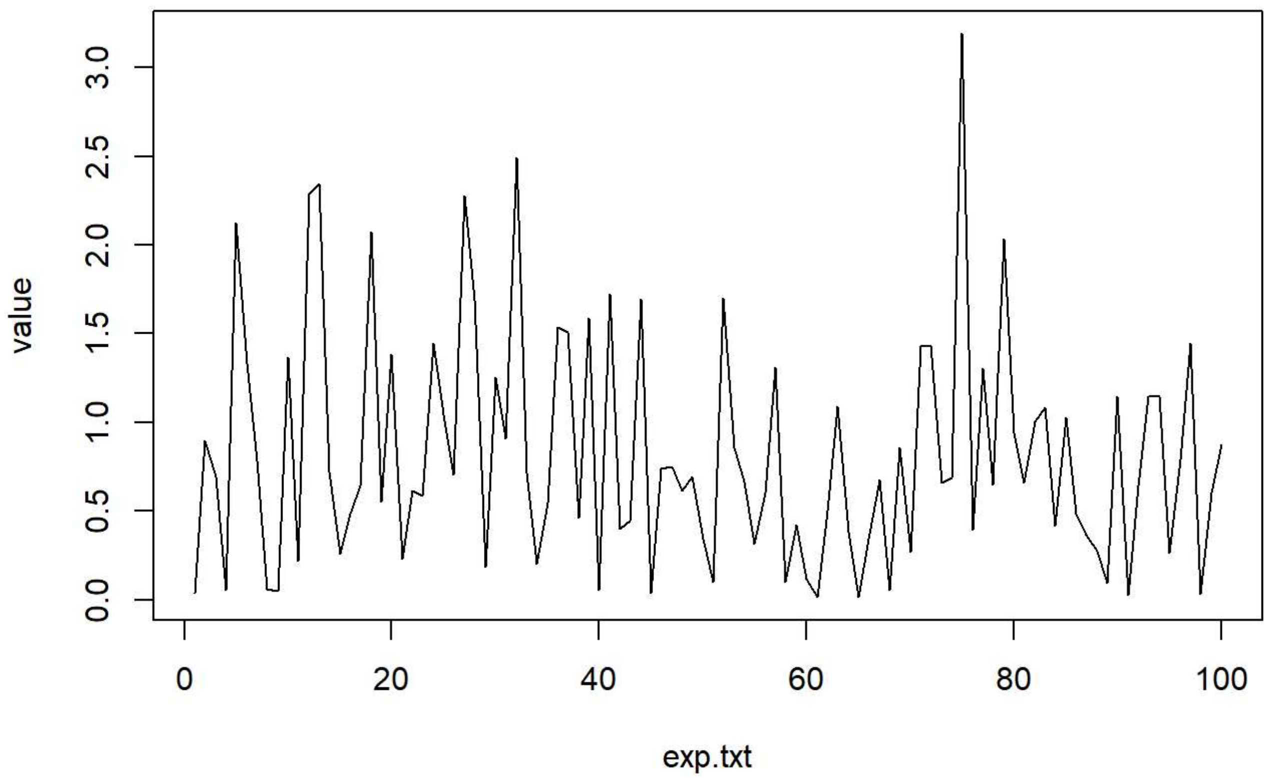
Cumulative distribution function:

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

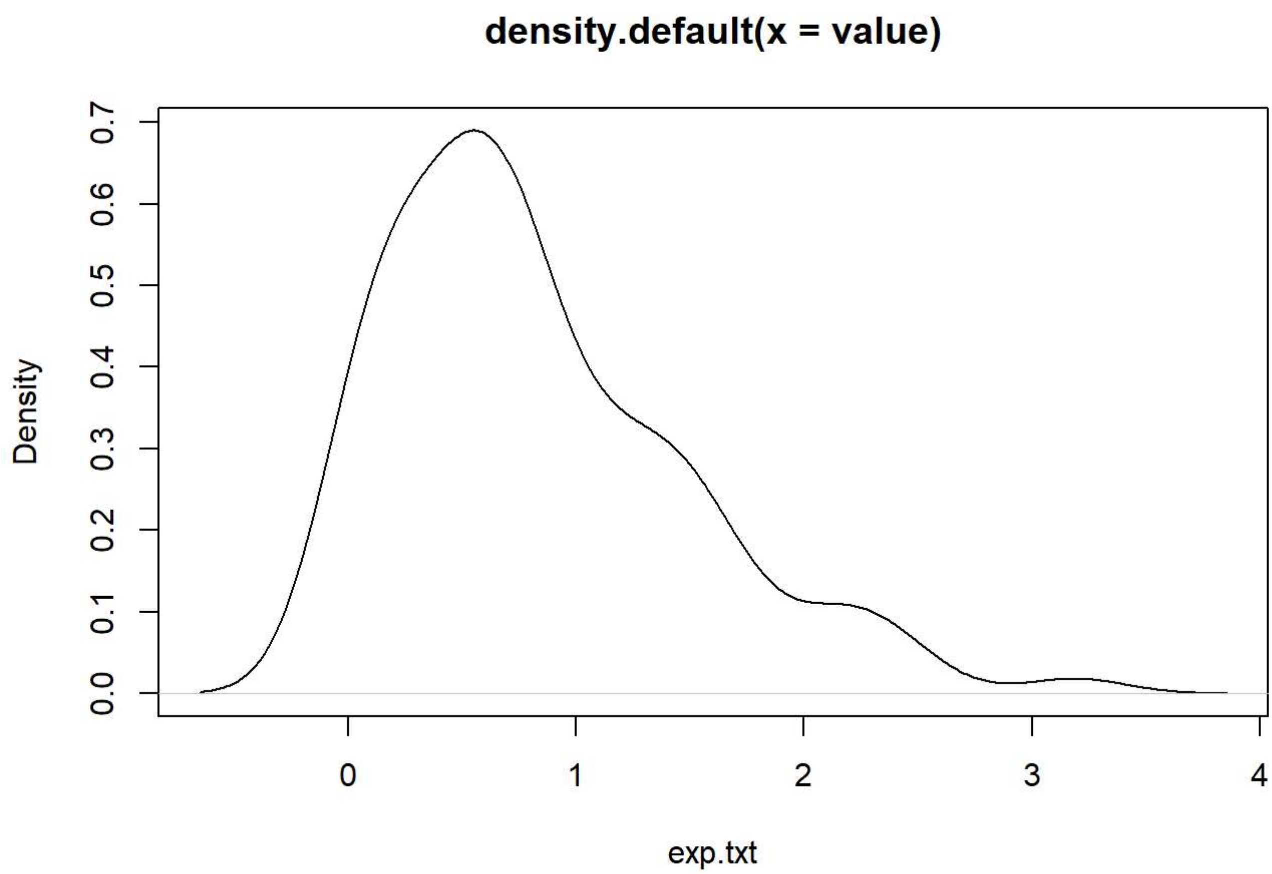
Cumulative distribution function

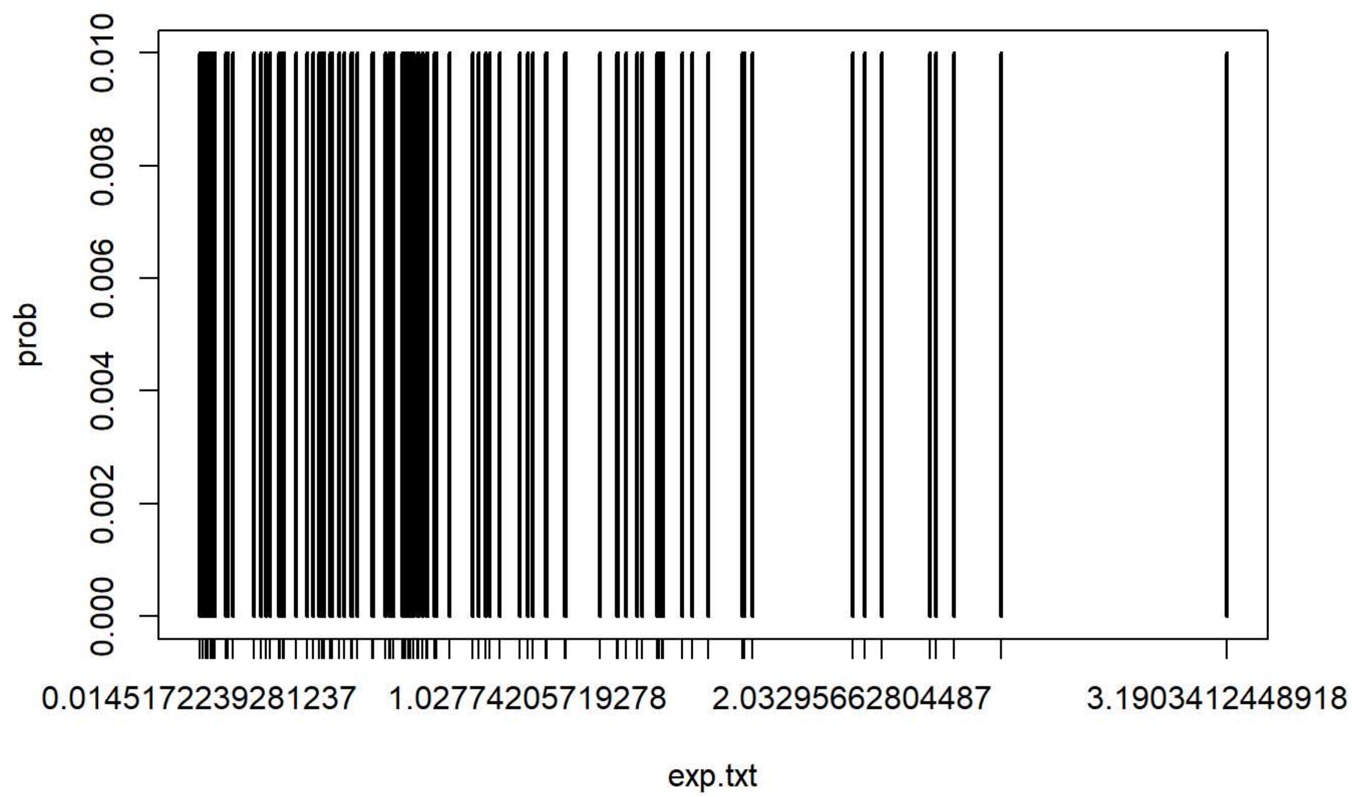
example data analysis result(telephone interval data)

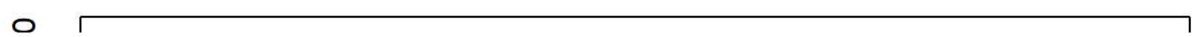
```
for (f in first) {
  #分析指定样本
  if(f=='exp'){
    name = paste(f,second,sep = '.')
    a = myfun(name = name)
  }
}
```

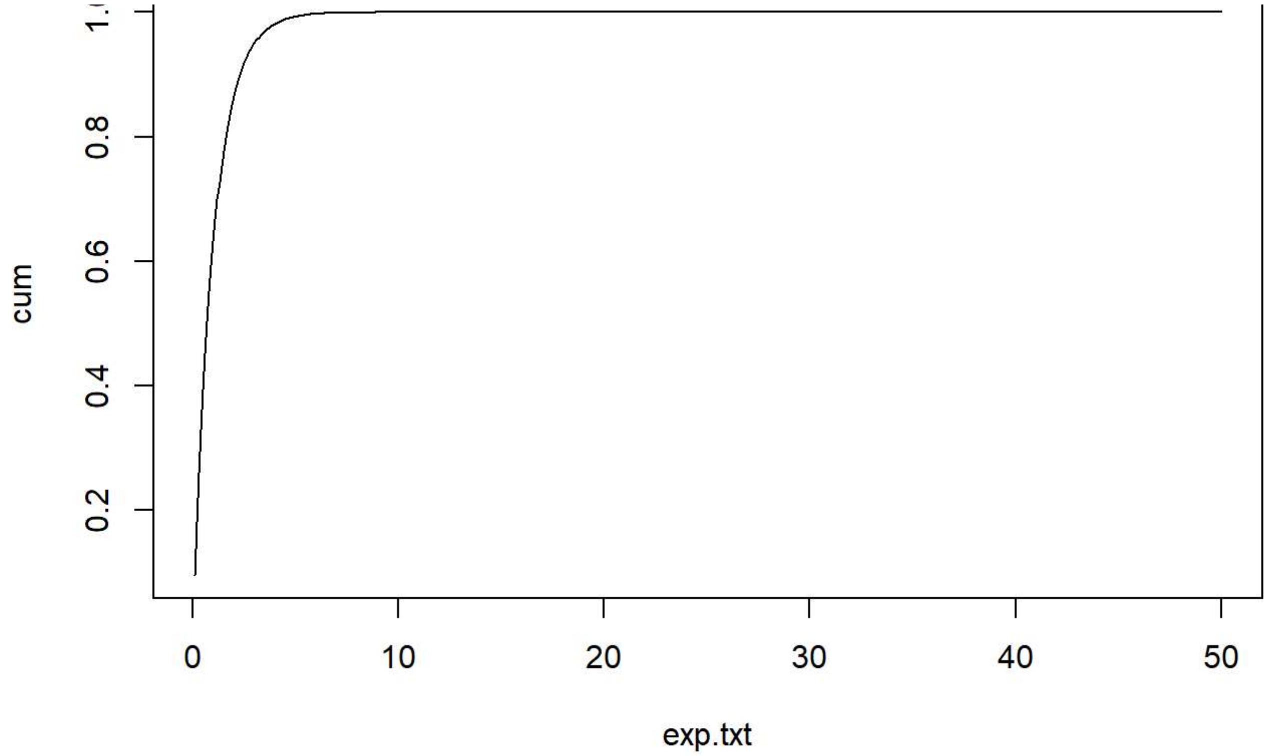



```
## [1] "varvalue = 0.430801850741277"
## [1] "expectvalue = 0.82526300875478"
```


density.default(x = value)
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3
4
exp.txt


prob
0.000
0.002
0.004
0.006
0.008
0.010
0.0145172239281237
1.02774205719278
2.03295662804487
3.1903412448918
exp.txt


0



Gamma distribution

In probability theory and statistics, the gamma distribution is a two-parameter family of continuous probability distributions. The exponential distribution, Erlang distribution, and chi-square distribution are special cases of the gamma distribution.

Probability mass function:

$$f(x;\alpha,\beta)=\frac{x^{\alpha-1}e^{-\beta x}\beta^{\alpha}}{\Gamma(\alpha)}\quad \text{for } x>0\quad \alpha,\beta>0,$$

Probability mass function

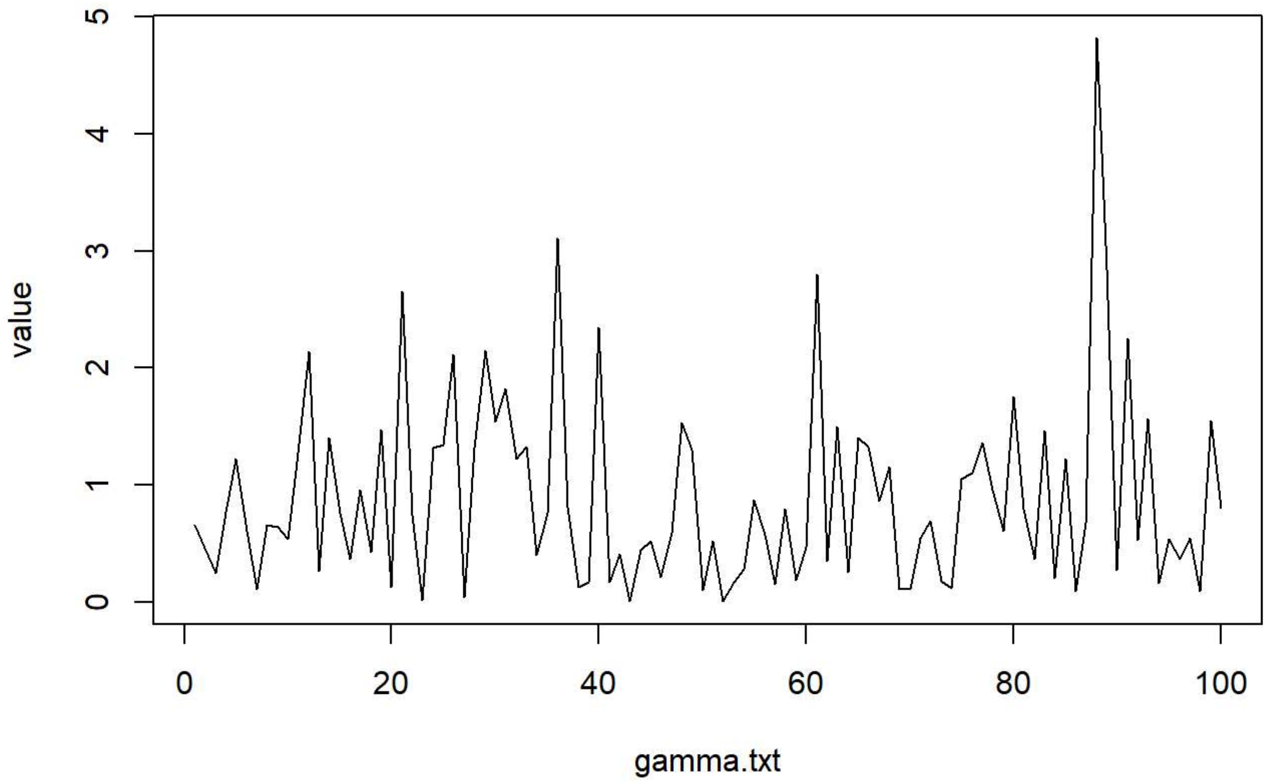
Cumulative distribution function:

$$F(x;\alpha,\beta)=\int_0^x f(u;\alpha,\beta)\,du=\frac{\gamma(\alpha,\beta x)}{\Gamma(\alpha)},$$

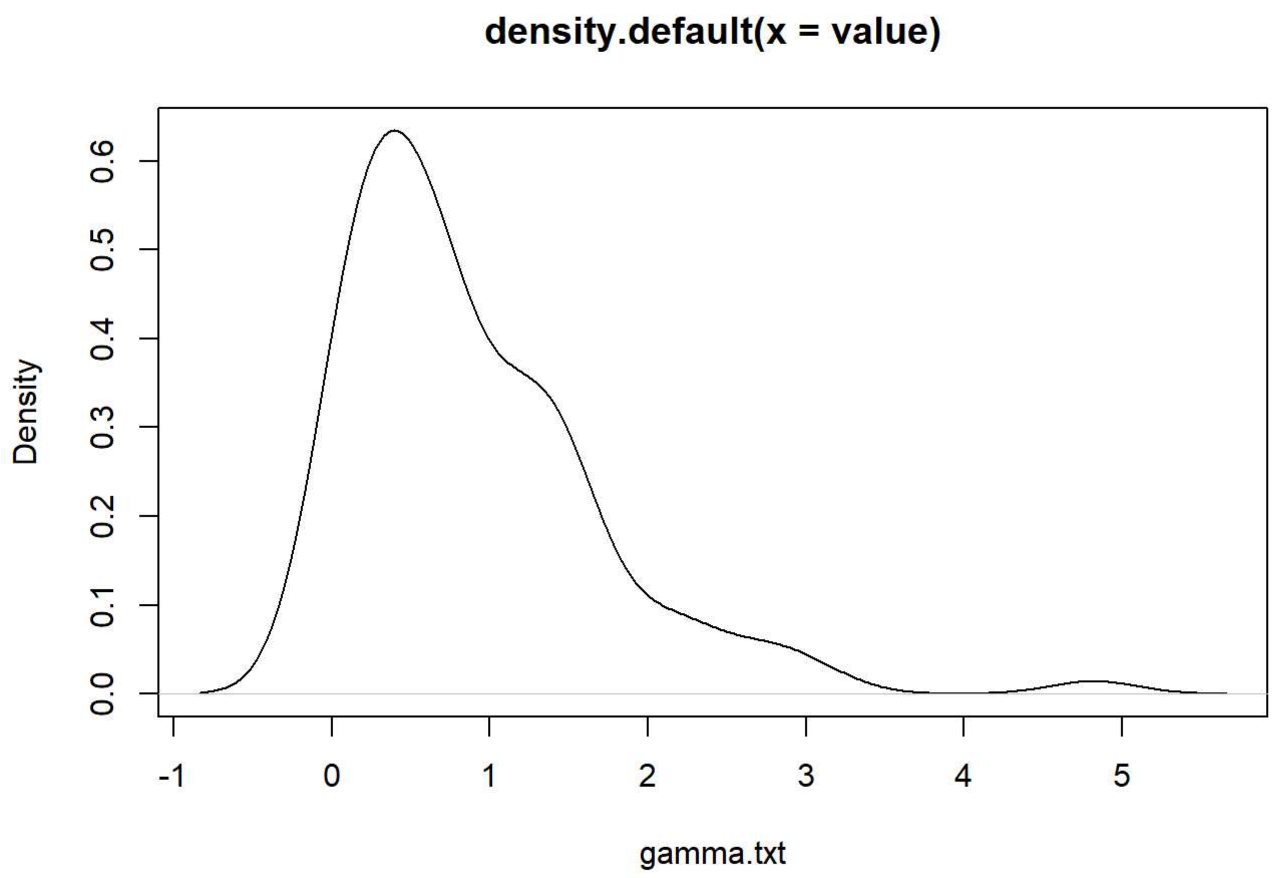
Cumulative distribution function

example data analysis result(rainfalls data)

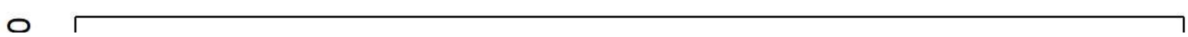
```
for (f in first) {
  #分析指定样本
  if(f=='gamma'){
    name = paste(f,second,sep = '.')
    a = myfun(name = name)
  }
}
```

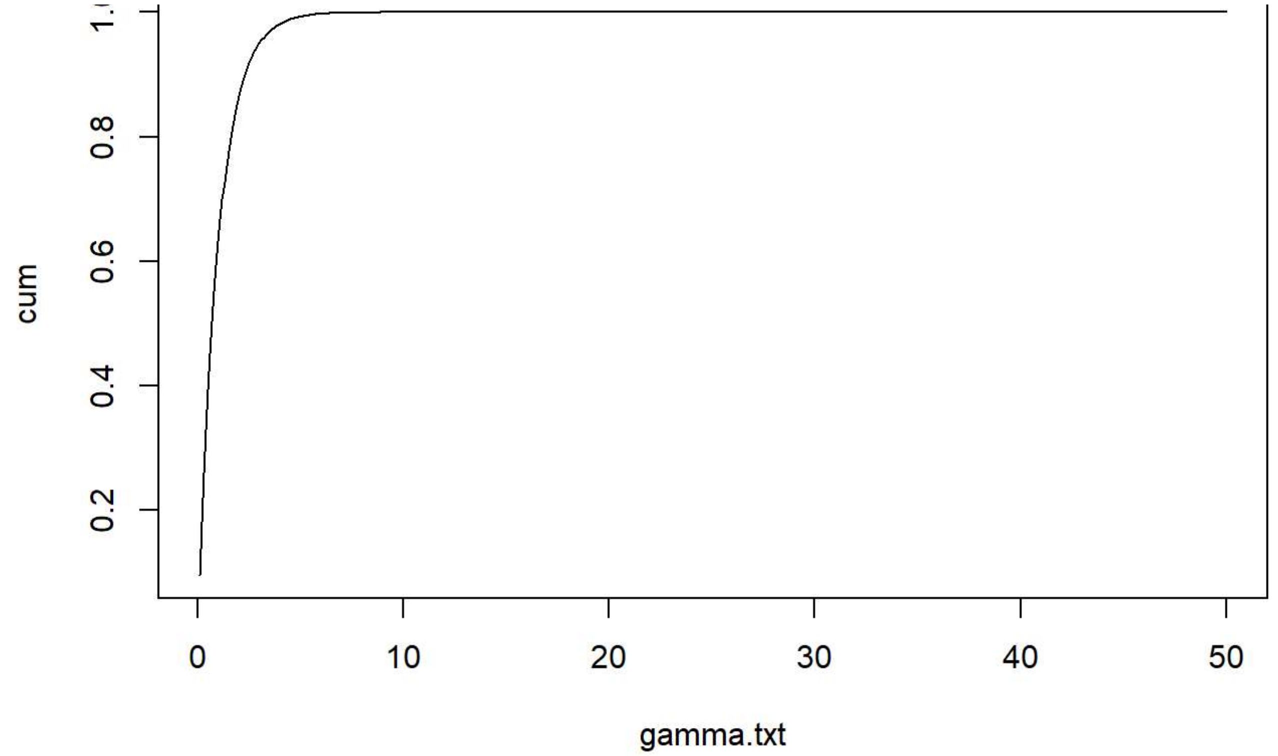



```
## [1] "varvalue = 0.659927173065152"
## [1] "expectvalue = 0.892255993392107"
```


density.default(x = value)
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
-1
0
1
2
3
4
5
gamma.txt


prob
0.000
0.002
0.004
0.006
0.008
0.010
0.00713647622689042
1.5467920018709
3.10743967818846
4.8168177258592
gamma.txt





question 2 the relationship between Binomial, Poisson and Normal Distribution

the relationship between Binomialand Poisson Distribution

The Poisson distribution is actually a limiting case of a Binomial distribution when the number of trials, n, gets very large and p, the probability of success, is small. As a rule of thumb, if n≥100 and np≤10, the Poisson distribution (taking λ=np) can provide a very good approximation to the binomial distribution.

the relationship between Binomial and Normal Distribution

the mean of the distribution is nP and the variance nP(1 - P). It turns out that as n gets larger, the Binomial distribution becomes approximately the same as a Normal distribution with mean nP and variance nP(l - P). This approximation is sufficiently accurate as long as nP > 5 and n(1 - P) > 5, so the approximation may not be very good (even for large values of n) if P is very close to zero or one. For the coin tossing experiment, where P = 0.5, 10 tosses should be sufficient. Note that this approximation is good enough with only l0 observations even though the underlying probability distribution is nothing like a Normal distribution.