Name:Meet Popat
Roll No.202032
SPA Assignment -1

Ques1: The magnitude of various earthquakes for a site

Ques1a)
The record of various magnitudes from 2000-2022 can be modeled as a Bernoulli process as the following assumptions of the Bernoulli process hold true:

1. There are two possible outcomes for each trial (where each trial is an instance of earthquake of magnitude >= 6) i.e success or a failure. Success indicates that the earthquake is having a magnitude greater than 6 and failure indicates that there is no earthquake or the earthquake has a magnitude less than 6. Hence, only <u>2 possible outcomes</u>.
2. Each occurrence of each earthquake is <u>independent</u> of each other.
3. The probability of success i.e The probability of having an earthquake of magnitude greater than 6 (=p) is same for each trial/instance i.e <u>identical</u> process.
4. We are studying the process over discrete time (i.e no. of instances of each earthquake is discrete)
5. The process can start at time t=0, but the arrival/success (i.e having an earthquake of magnitude >=6) can only occur at time>=1. (Time denotes the instance of earthquake with magnitude >= 6)
6. Only one earthquake can happen at one time (i.e one arrival at one time)

The scatter plot is shown for the Bernoulli process with success probability (p = 0.4), the x-axis denotes the instances of earthquake, the success/ failure of the earthquake having magnitude >=6 is taken on the y-axis.

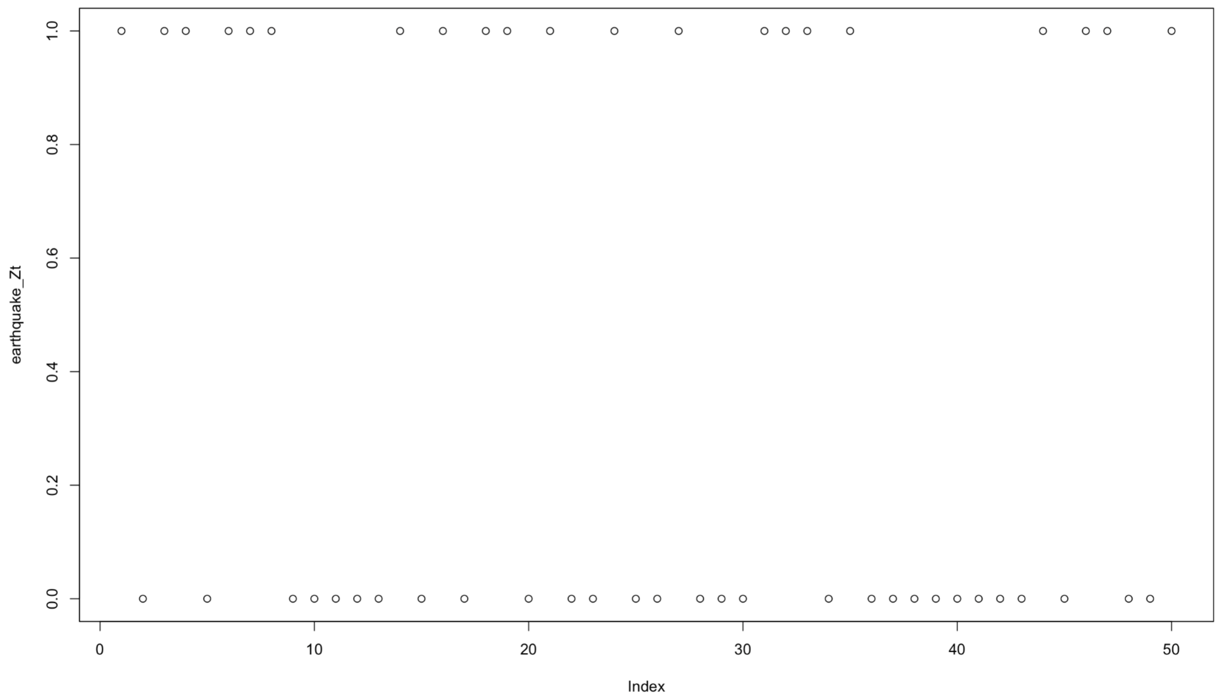
Graph of Bernoulli distribution is scatter points either at 1 or 0 denoting success and failure respectively.

The code for the plot on R is :

```
p<-0.4
t<-50

earthquake_Zt<-sample(0:1,t,replace=T,prob=c((1-p),p))
plot(earthquake_Zt)
```

Scatter plot for bernoulli process with success probability = 0.4 :



Ques1b)
Since the process is a Bernoulli process, the interarrival times follow a geometric distribution.
Proof: Let X1 denote the first interarrival time.
P(X1 = 1) = P(1st earthquake have magnitude >= 6) = p = 0.4
P(X1 = 2) = P(1st earthquake<6 & 2nd earthquake have magnitude >= 6) = (1-p)*p
.
.
P(X1 = k) = P(no earthquake >=6 for k-1 instances, arrival at the kth instance of earthquake >=6) = (1-p)^(k-1) * p
Thus, we observe that X1 follows a geometric distribution(p).
By the fresh start property, it can be similarly shown that all X2, X3, X4, ... follow geometric distribution(p).

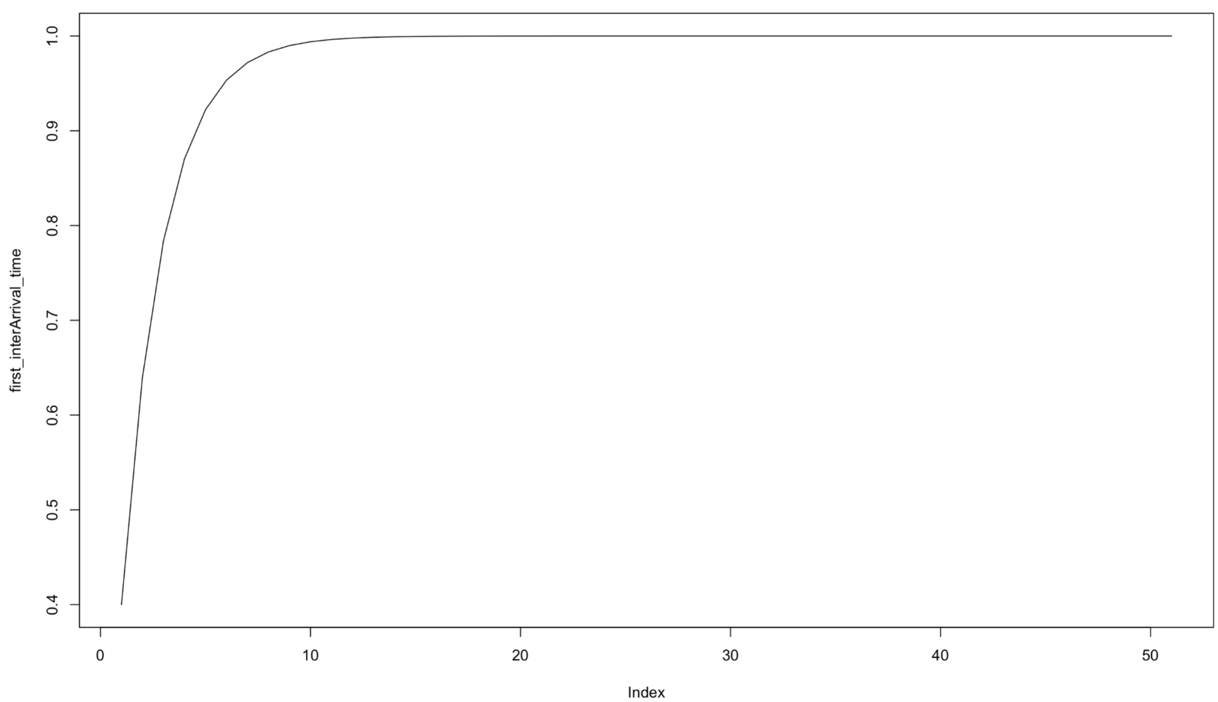
The graph for distribution (CDF) of geometric distribution is plotted with values for first interarrival time on the x-axis and corresponding CDF on the y-axis.

We know that as the number of trials increases the CDF tends to 1. We can confirm that from the graph.

The code for it on R:

```
x <- seq(0, 22, by = 1)
first_interArrival_time<- pgeom(x, prob = 0.4)
plot(first_interArrival_time,type = "l")
```

The plot of CDF of 1st interarrival time:(X-axis: Time and Y-axis:1st interarrival time)



Ques1c)

Nk is the number of earthquakes of magnitude>=6 till time k.

Assuming we need to find the total no. of occurrences in 50 earthquake,

To study the change in behavior for 2 different p(0.4 and 0.9) we use the Moment generating function (MGF) to derive the distribution of the process.

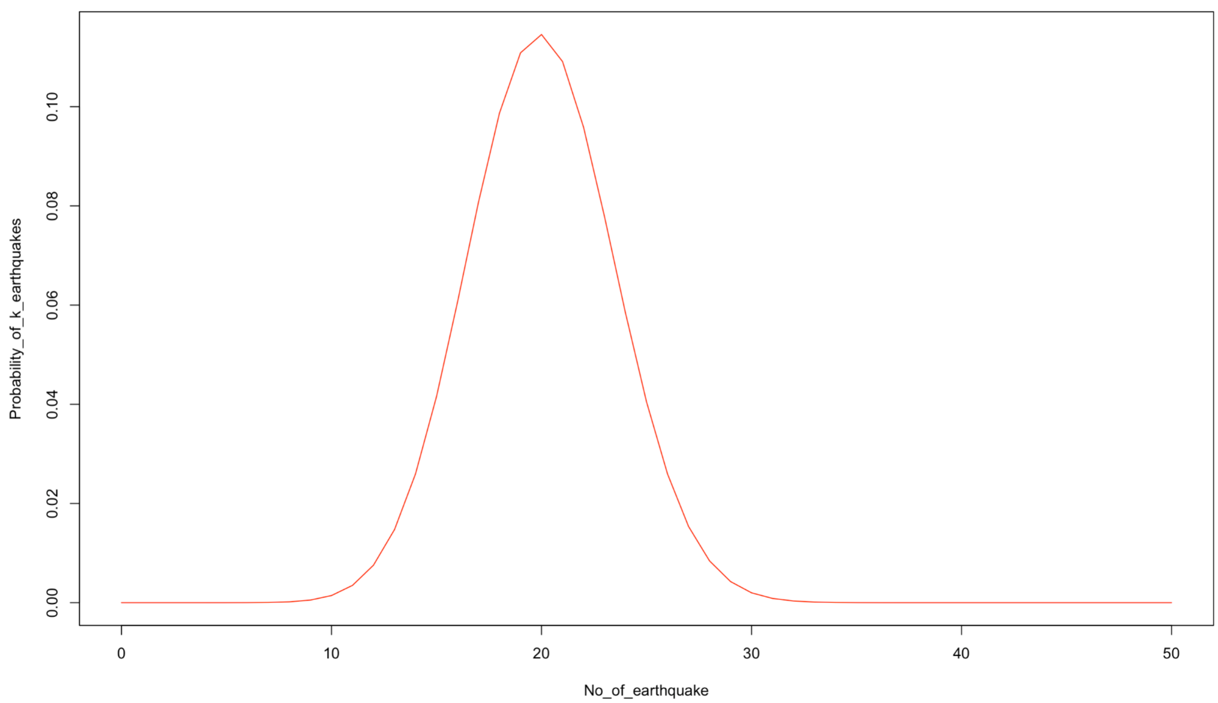
MN1(s) = E(exp(sN1))

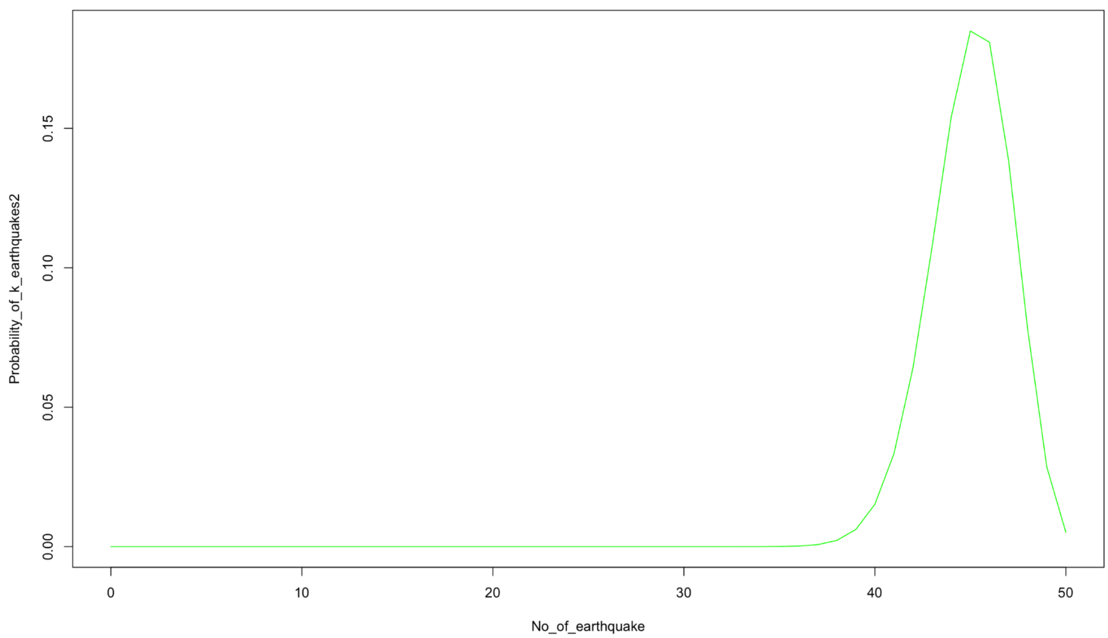
MN2(s) = E(exp(sN2))

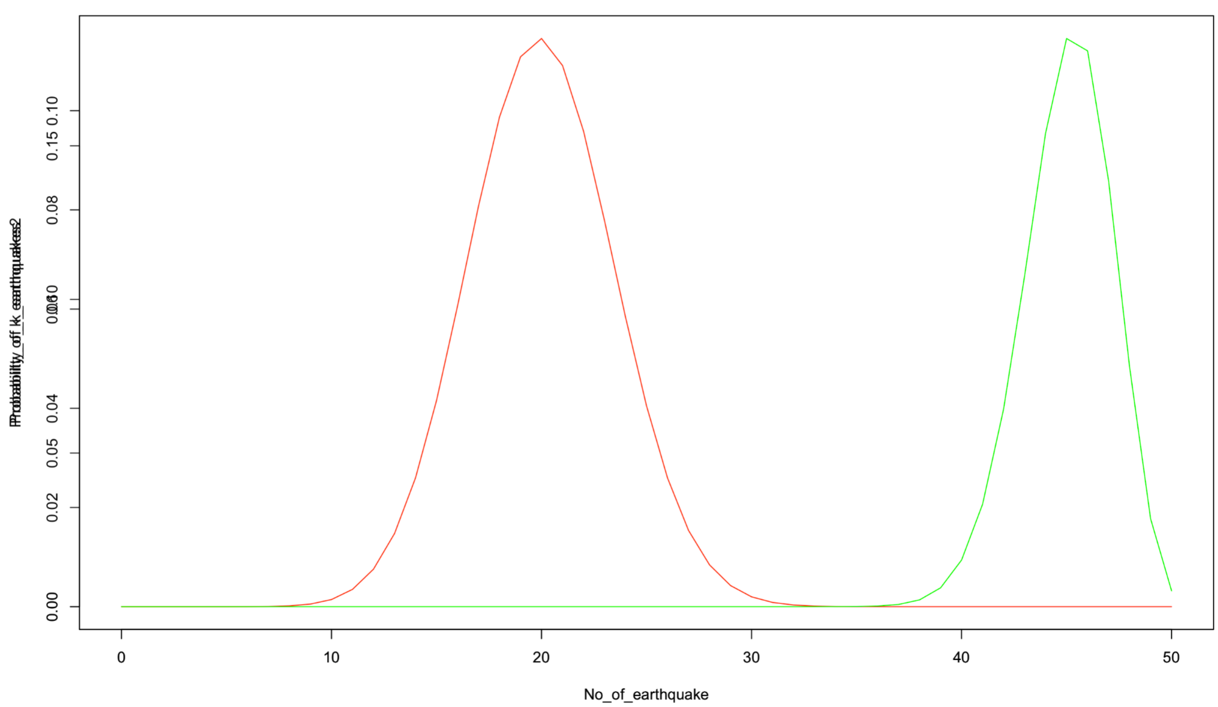
N1 follows a binomial distribution (n1, p), Mgf of N1 (MN1(s))= (p*exp(t) + (1-p))^ n1

Similarly, N2 follows a binomial distribution (n2, p), Mgf of N2 (MN2(s))= (p*exp(t) +(1-p))^ n2

Thus N1 and N2 both follow a binomial distribution.
Nk ~ Binomial(k,p) where k is the time and p is the probability of having an earthquake of magnitude >= 6.
The following plot is obtained.(X-axis -> No. of earthquakes and Y-axis is probability of k earthquakes)
The red curve is with probability 0.4 and green curve with probability 0.9:







The code in R:

```
No_of_earthquake <- 0:50
Probability_of_k_earthquakes <- dbinom(success, size=50, prob=0.4)
plot(No_of_earthquake, Probability_of_k_earthquakes, type='l',col = 'red')
```

```
par(new = TRUE)
Probability_of_k_earthquakes2 <- dbinom(success, size=50, prob=0.9)
plot(No_of_earthquake, Probability_of_k_earthquakes2, type='l', col = 'green')
```

The change in behavior is due to the change in probability :
As we increase the probability we get the peak of the probability of arrivals shifts to right i.e for the same probability the number of arrivals increases at the peak, and hence the peak of the curve shifts to the right.

Ques 2 : Visitors on a website

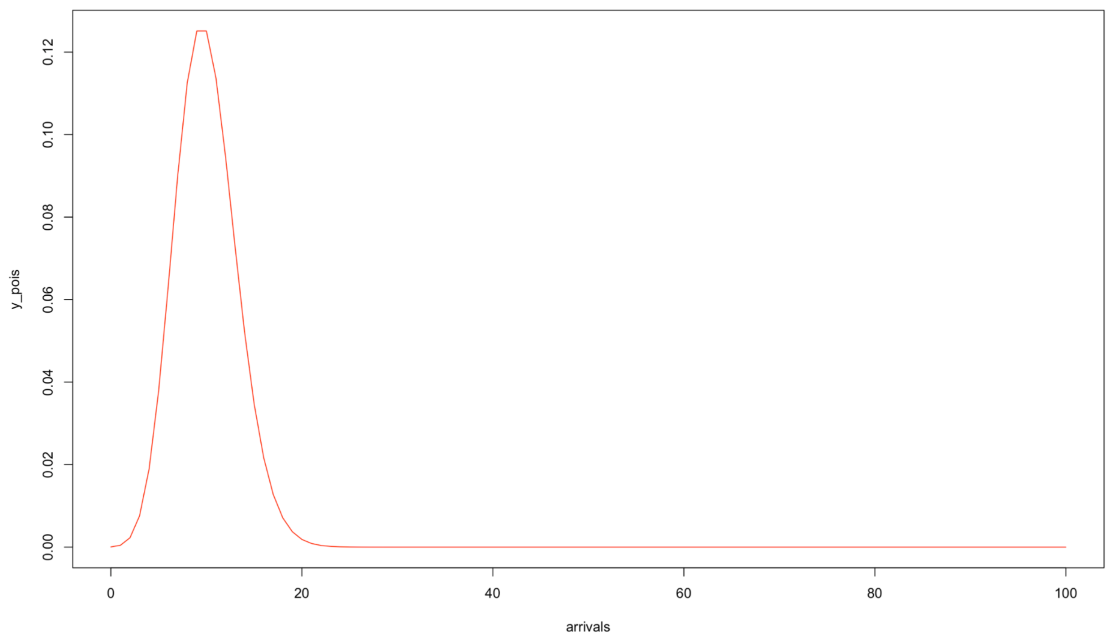
Ques2a):
The following is a poisson process with the average rate of number of visitors in an hour (lamba = 10) and hence the following assumptions hold true:

1. It is an arrival process i.e starts at t=0 and multiple arrivals don't occur.
2. It is a counting process
3. It is a renewal process i.e interarrival time is a sequence of IID random variables .
4. Stationary Increment Property
5. Independent Increment Property.

The density of the number of arrivals until time t is given by the following graph.(X-axis -> No. of arrivals and Y axis is the probability of the number of arrivals)
The function dpois() calculates the probability of a random variable :

The poisson process is given by
P(N(t)= n) = e^(-Lambda *t) *(Lambda*t)^n/n!

As the lambda is 10 the graph peaks towards 10.



Code on R:

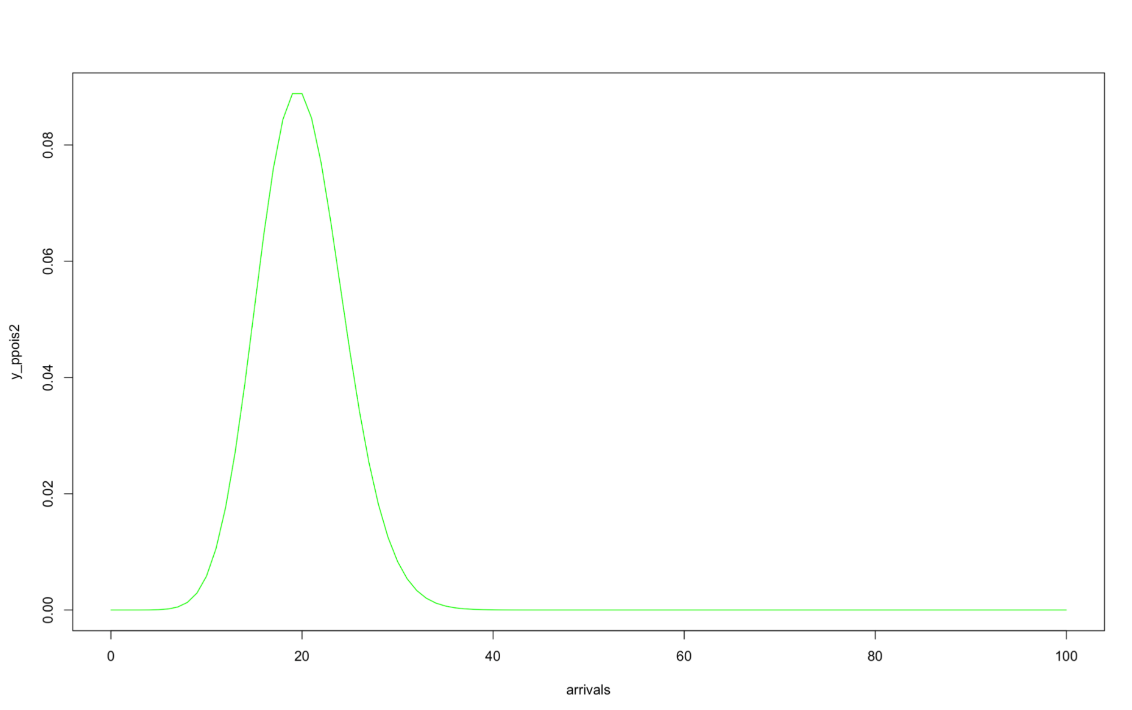
arrivals <- 0:100

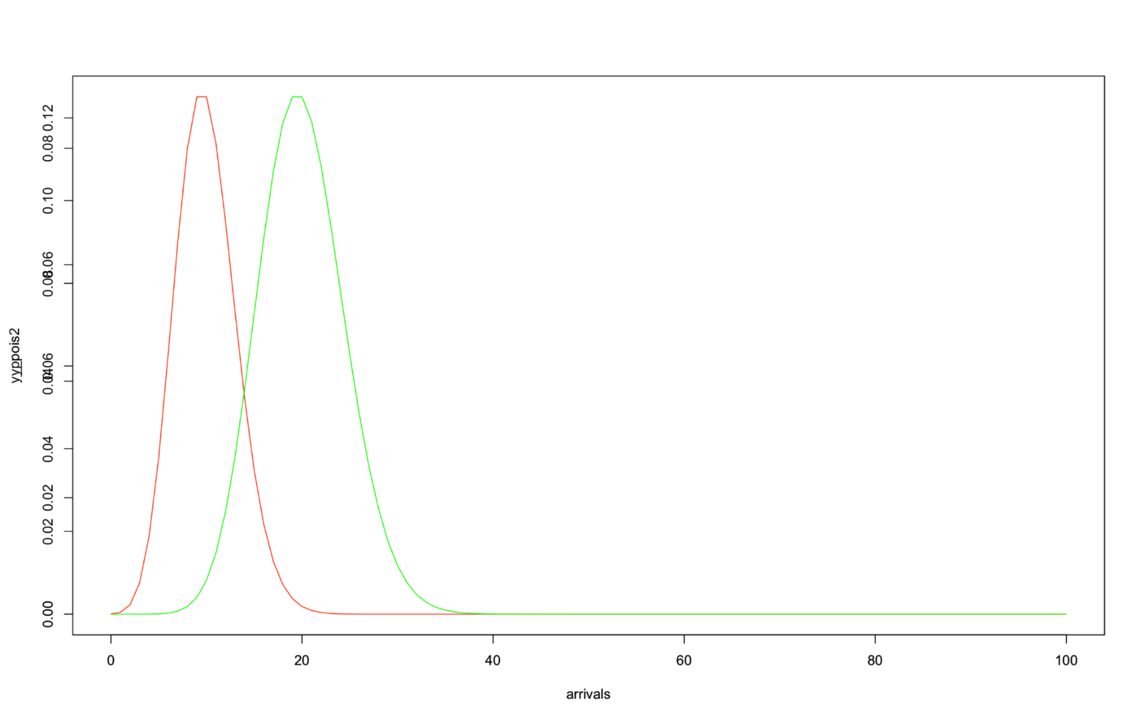
y_pois = dpois(arrivals,10)

plot(arrivals,y_pois,type = "l",col = 'red")

Ques2b):

The graph for lambda = 20 as above is the follows:





Comparison:
The graph of the poisson process for the above 2 processes only varies as Lambda varies from 10 and 20.
The lambda is the average visitors on a website in an hour. Hence the 2nd graph peaks at 20 arrivals as the avh arrivals in an hour is 20.
Ques2c):

For Combined poisson process :
N1(t) ~ P(Lambda1, t)
N2(t) ~ P(Lambda2, t)

And N1 is independent of N2
N(t) = N1(t) + N2(t)

N(t) ~ Possession (lambda,t)

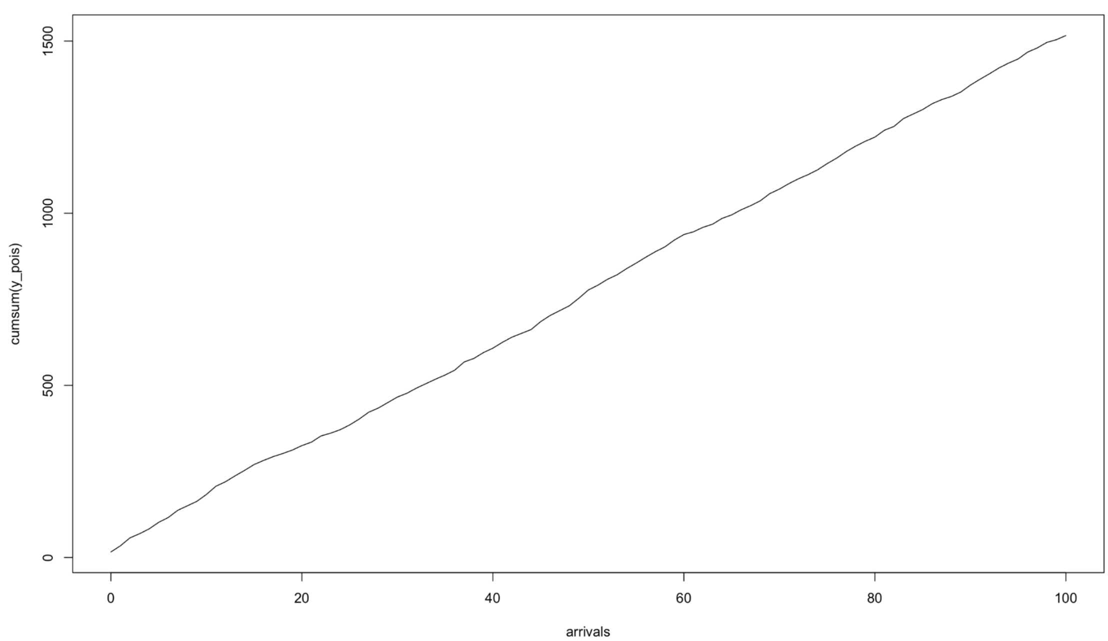
Therefore by MGF :
Mn(t) = E[e^N(t)s] = E[e^(N1(t) + N2(t))]
= e(lambda1 + lambda2)t(e^s-1)

Therefore lambda for N(t) is Lambda1 + Lambda2

Hence lambda = 10+5 = 15

For the total number of visitors we cumulative sum the random average visitors for trials .
The function rpois gives the random visitors on a website in an hour.

The graph of total visitors is given by:(X-axis -> time and Y axis -> No. of visitors)



Code On R:
lambda <- 5+10

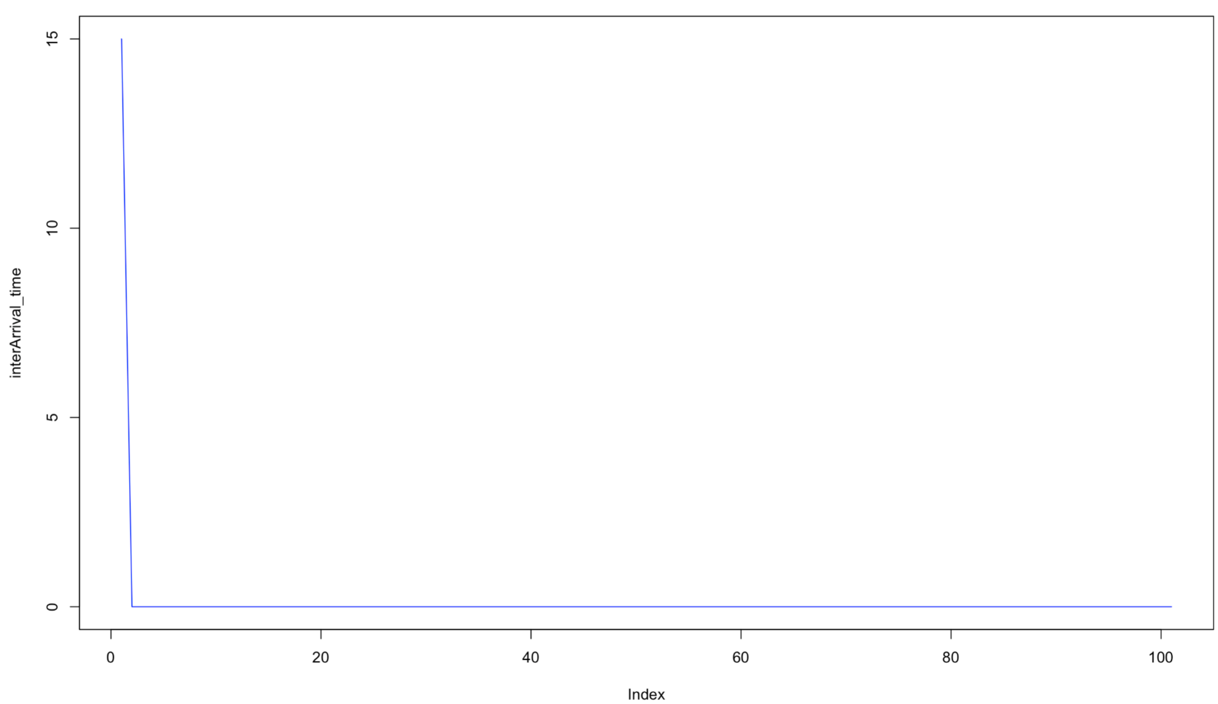
arrivals <- 0:100
y_pois =rpois(arrivals,lambda)
plot(arrivals,cumsum(y_pois),type = "l")

2nd part:

The 1st inter arrival time is simulated as the exponential for the lambda parameter :
The graph of the 1st interarrival time is given by the exponential function of r.

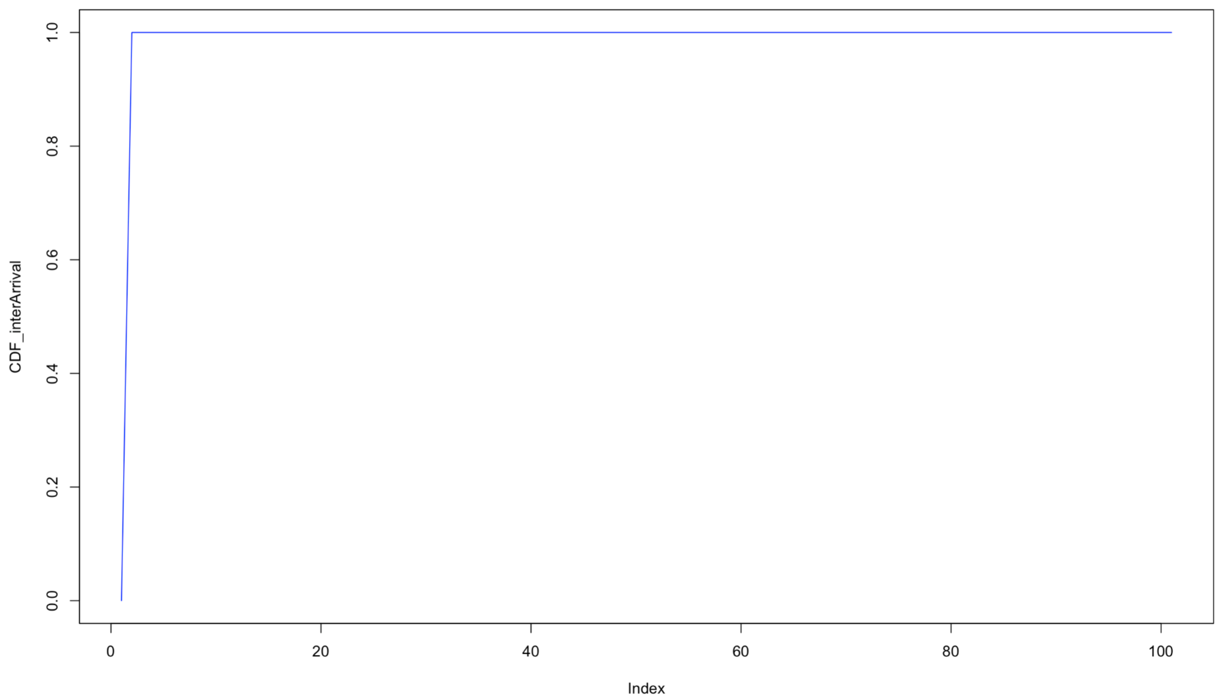
As X1 follows exponential function the graph is as follows:(X-axis as time and Y-axis as InterArrival Time)

The code on R is:
lambda <- 5+10
interArrival_time <- dexp(seq(0,t,1),lambda)
plot(interArrival_time,type = 'l',col = 'blue')



The graph for the cdf of the interarrival time is given by:(X-axis as time and Y-axis as CDF)

As it is a CDF it should peak towards 1.



Code on R :

lambda <- 5+10
CDF_interArrival <- pexp(seq(0,t,1),lambda)
plot(CDF_interArrival,type = 'l',col = 'blue')