MTH371 - Stochastic Processes and Applications
Assignment 1 Report

Aditya Chetan (2016217)

October 11, 2018

1 Question 1

In this question, we were supposed to perform the following two tasks:

- **Task 1:** Simulate a Bernoulli process over 30 time steps. The Bernoulli process has parameter $p = 0.8$.
- **Task 2:** Simulate the first inter-arrival time of the Bernoulli process mentioned above and plot it's cdf.

1.1 Task 1

We know that a Bernoulli process can be generated from sampling a Binomial Distribution. This is because a Bernoulli process is essentially a collection of many i.i.d. Bernoulli RVs, same as a Binomial RV.

Hence, this task can be achieved by sampling an array of size 30 from a Binomial RV with $p = 0.8$, using the function `rbinom()`

The output obtained for Task 1 is shown below:

```
[1] "The simulated Bernoulli process gives the following observations:"
[1] 1 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0
[26] 1 1 0 1 0
[1] "Here 1 denotes that the page was printed and 0 that it was not"
```

1.2 Task 2

For this task, we know that the inter-arrival times of a Bernoulli process are i.i.d. Geometric RVs. Hence, the first inter-arrival time would be $Geometric(p = 0.8)$. Using the function, `pgeom` in `R`, we can easily plot it's cdf for 30 trials. The graph for the same that we obtained is given below in Figure 1.

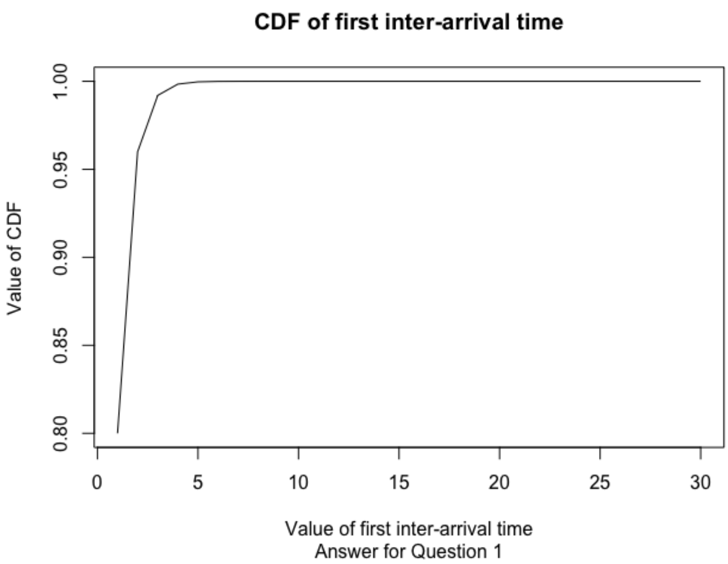


Figure 1: Q1 Task 2: CDF of first inter-arrival time

2 Question 2

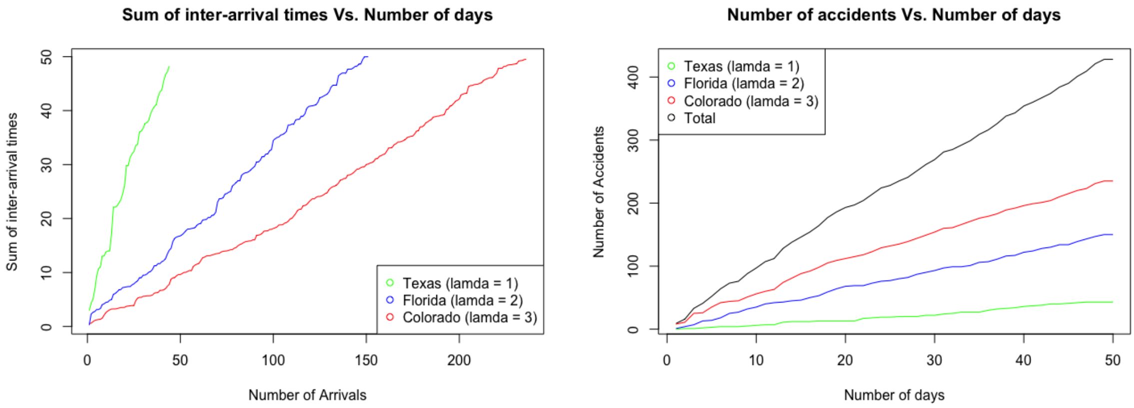
In this question, again we had to perform 2 tasks:

- **Task 1:** Given the information about the poisson processes, we have to plot the sum of inter-arrival times of the accidents on each of the highways.
- **Task 2:** Plot the total number of accidents on each of the highways and together.

To do this, I followed the following steps:

- We know that the inter-arrival times in a poisson process are i.i.d. Exponential RVs. Hence, I sampled random numbers from the exponential distribution $Exponential(\lambda)$ using the `rexp` function in `R`. Here, λ is the rate parameter for the highway under consideration.
- I continued generating samples for each highway till the sum of inter-arrival times crossed the total time, i.e., 50 days.
- Using this array of inter-arrival times, I calculated the number of accidents happening each day on each highway and the sum of inter-arrival times.

The plots generated are given in Figure 2.



(a) Task 1: Sum of inter-arrival times Vs. Days

(b) Task 2: Days Vs. Number of Accidents

Figure 2: Plots generated for Q2.

2.1 Observations

2.1.1 Task 1

From the plot obtained in Task 1, it can be observed that the highway with the highest rate has the lowest sum of inter-arrival time for the same number of arrivals. Alternatively, the highway with the highest rate gets to the same sum of inter-arrival times in the highest number of arrivals.

Thus, Colorado, which has the highest rate ($\lambda = 3$ accidents/day) has the lowest sum of inter-arrival times for the same number of arrivals. Colorado is followed by Florida and Texas ($\lambda = 2$ accidents/day and $\lambda = 1$ accident/day respectively).

This is an expected observation. This is because sum of inter-arrival times can also be interpreted as the time required to get to a particular number of arrivals. So for a highway with a higher rate of accidents, the time required to get to the same number of accidents would be lesser as compared to other highways. Hence, the plot gives an expected output.

2.1.2 Task 2

From the plot for Task 2, it can be observed that the total number of arrivals/accidents is higher for a highway with higher rate of accidents for the same number of days.

Thus, Colorado, which has the highest rate ($\lambda = 3$ accidents/day) has the highest number of total arrivals/accidents till date on any given day. Colorado is followed by Florida and Texas ($\lambda = 2$ accidents/day and $\lambda = 1$ accident/day respectively).

Again, this seems intuitive as a highway with a higher rate for accidents, is more probable to have a higher total number of accidents till date. This intuition is observed to be confirmed from the plot (b) in Figure 2.

3 Question 3

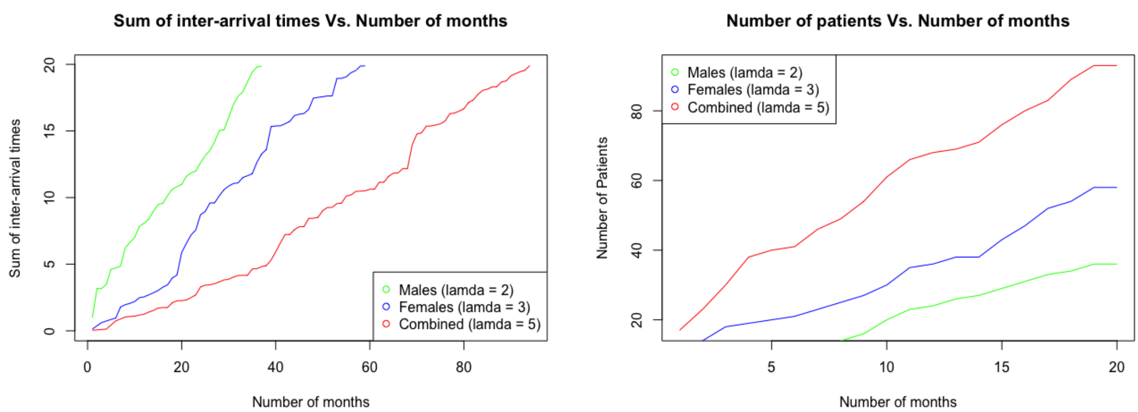
In this question we were supposed to perform the following tasks:

- **Task 1:** Given the rates for Male patient and Female patient arrivals, simulate the poisson processes and perform the same analyses as done in Q2 for both of them.
- **Task 2:** Consider the poisson process formed by combining the Male and Female processes. Now simulate it and perform the same analyses as in Q2.

To achieve these tasks I did the following:

- For simulating the separate poisson process, I performed the same steps as mentioned in Q2.
- For the combined process, I used the fact that a process that is a combination of two poisson processes, $Poisson(\lambda_1)$ and $Poisson(\lambda_2)$ can be given by, $Poisson(\lambda_1 + \lambda_2)$. Thus, the combined process would be a poisson with rate, $\lambda = \lambda_1 + \lambda_2 = 2 + 3 = 5$ people/month. Then I performed the same steps as in Q2 to simulate it and obtain the desired plots.

The desired plots are shown in Figure 3.



(a) Task 1: Sum of inter-arrival times Vs. Months

(b) Task 2: Months Vs. Number of Patients

Figure 3: Plots generated for Q3.

3.1 Observations

3.1.1 Task 1

From the plot obtained in Task 2, it can be observed that the category of patients with the highest rate has the lowest sum of inter-arrival time for the same number of arrivals. Alternatively, the category of patients with the highest rate gets to the same sum of inter-arrival times in the highest number of arrivals.

Thus, the combined arrival of male and female patients has the highest sum of inter-arrival times in the highest number of arrivals, followed by females and males respectively.

This is an expected observation. This is because sum of inter-arrival times can also be interpreted as the time required to get to a particular number of arrivals. So for a patient category with a higher rate of arrivals, the time required to get to the same number of arrivals would be lesser as compared to other categories of patients. Hence, the plot gives an expected output.

3.1.2 Task 2

From the plot for Task 2, it can be observed that the total number of patient arrivals is higher for a patient category with higher rate of arrivals for the same number of months.

Thus, the combined category of males and females, which has the highest rate ($\lambda = 5$ arrivals/month) has the highest number of total arrivals till date on any given number of months. The combined category is followed by Females and Males ($\lambda = 3$ arrivals/month and $\lambda = 2$ arrivals/month respectively).

Again, this seems intuitive as a category with a higher rate of arrivals, is more probable to have a higher total number of arrivals till date. This intuition is observed to be confirmed from the plot (b) in Figure 3.