IE 306 - Homework 2

Mehdi Saffar - 2016400411
Burak Berk Ozer - 2016400015
Mehmet Umut Oksuz - 2016400096

May 20, 2020

1 Introduction

In this project, we are given the observed interarrival times of customers to a system and we are tasked to find the best-fitting random distribution. To do that we will compare the empirical data with two of possible distributions: uniform distribution and exponential distribution. We will use various statistical tests to find the most fitting one.

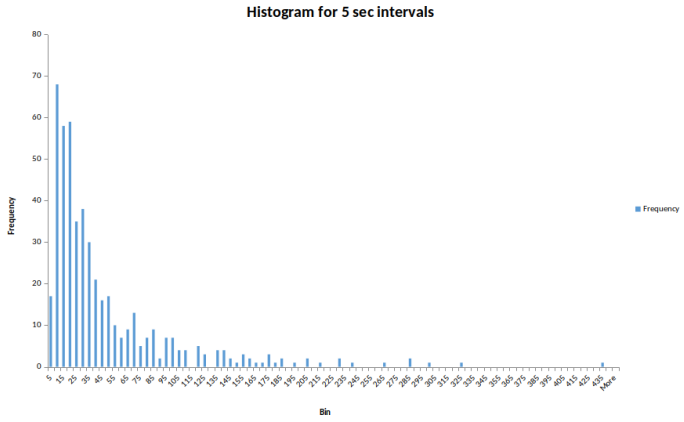
2 Kolmogorov-Smirnov test

For days 1 and 2 we have applied the Kolmogorov-Smirnov test to see if it is possible for that data to be uniformly distributed between 0 and 400. To scale the data, first we have divided all the data with 400. And then we calculated values for $R(i)$, i/N, $i/N - R(i)$ and $R(i) - (i-1)/N$. After getting all these values, we get the maximum and determine our KS statistic. According to our KS statistic, using a 0.05 significance level we have concluded that data for both days cannot come from a uniform distribution. All the details of our Kolmogorov-Smirnov test can be found in our Excel file.

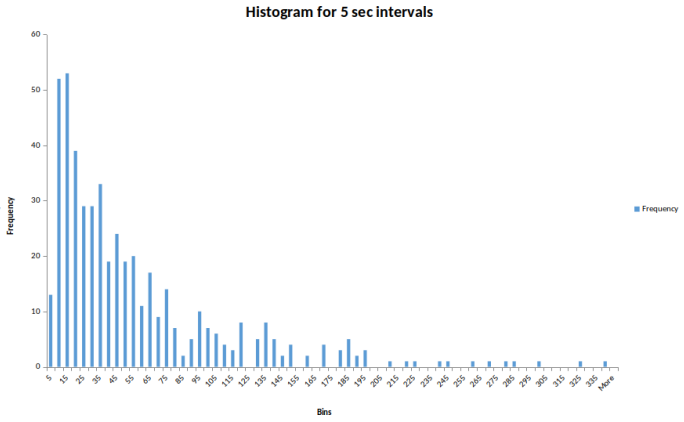
3 Dataset statistcs

	count	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
Day 1	488	44.752	52.804	1	13	26	54	434
Day 2	488	53.801	55.092	0	15	35	69.25	339

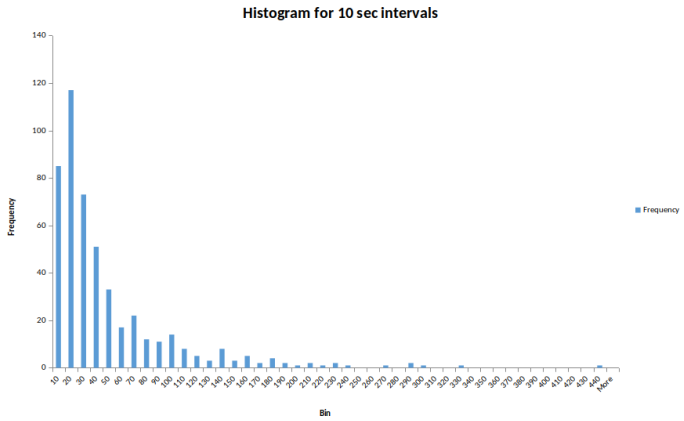
4 Frequency histograms



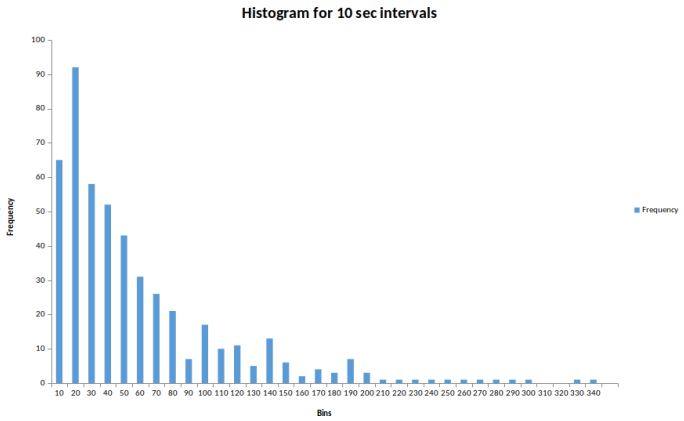
(a) Frequency histogram 5 seconds intervals of day 1



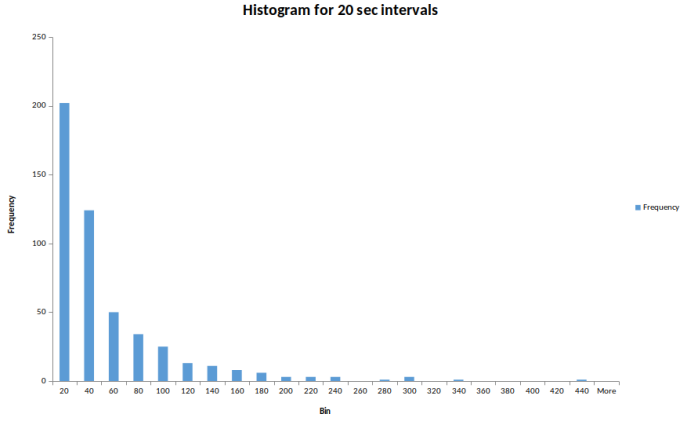
(b) Frequency histogram 5 seconds intervals of day 2



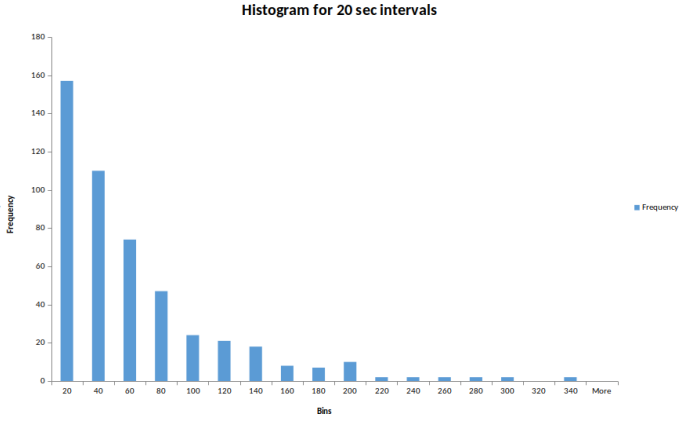
(a) Frequency histogram 10 seconds intervals of day 1



(b) Frequency histogram 10 seconds intervals of day 2



(a) Frequency histogram 20 seconds intervals of day 1

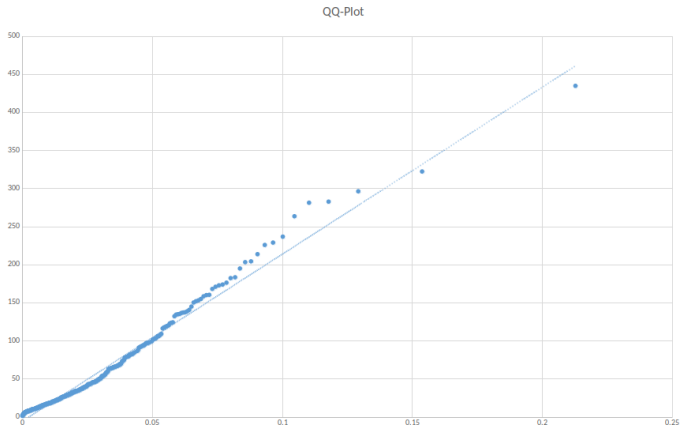


(b) Frequency histogram 20 seconds intervals of day 2

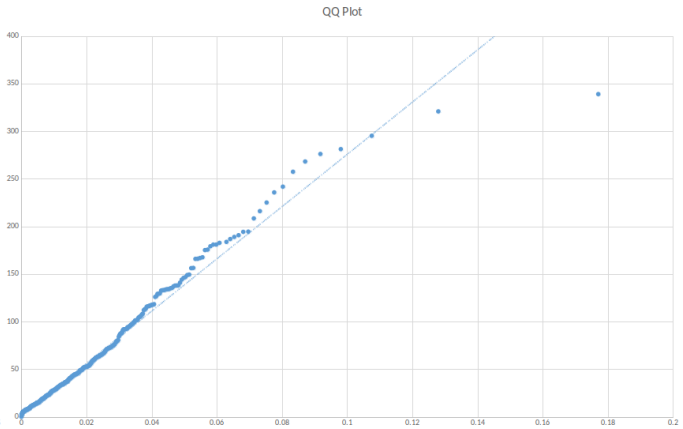
5 Chi-Squared test

We have conducted a Chi-Squared test for both days separately. We used 10 second intervals as expected and determined our bins accordingly. Frequencies and expected frequencies of all bins are calculated and according to those values Chi-Squared statistic is determined. For both days, we see that our Chi-Squared statistc is smaller than the table value. So, we cannot reject the hypothesis that this data are coming from an exponential distribution with means specified in section 3 of the report.

6 QQ-Plot

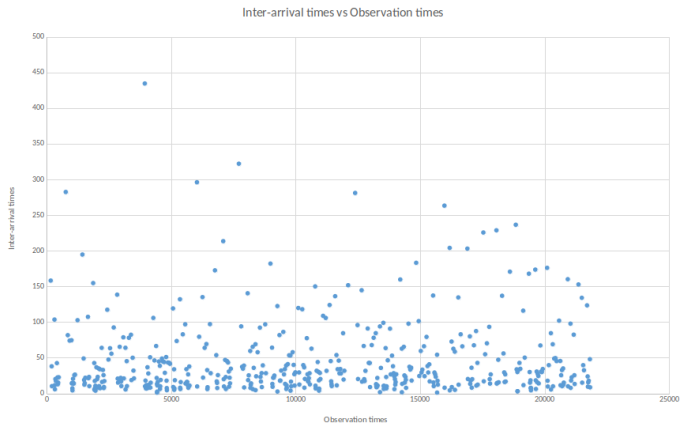


(a) QQ-Plot of day 1

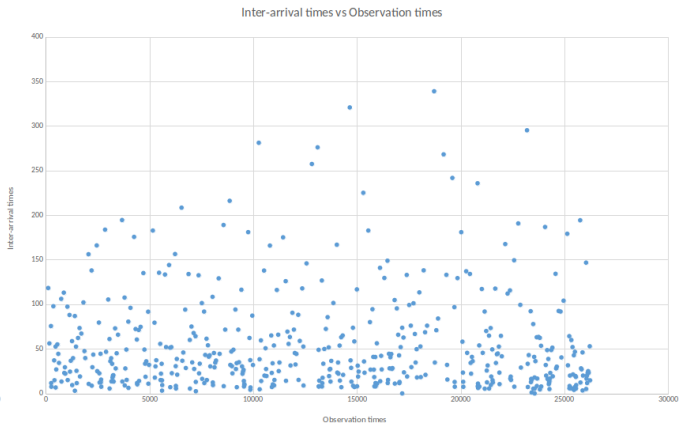


(b) QQ-Plot of day 2

7 Interarrival time plot



(a) Interarrival time vs Observer times for day 1



(b) Interarrival time vs Observer times for day 2

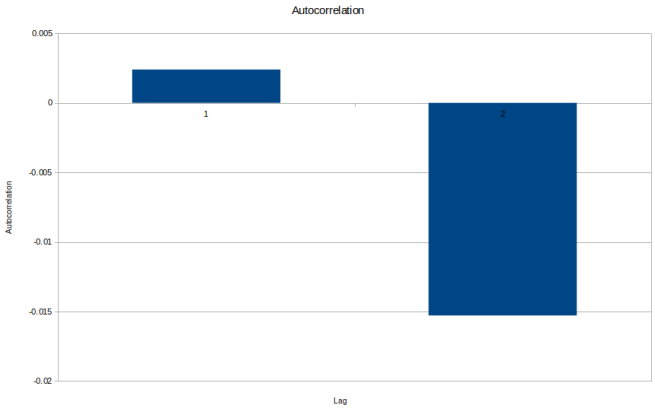
8 Autocorrelation test

In order to do the autocorrelation test we shift data forward once for lag 1 and twice for lag 2, then calculate using this formula:

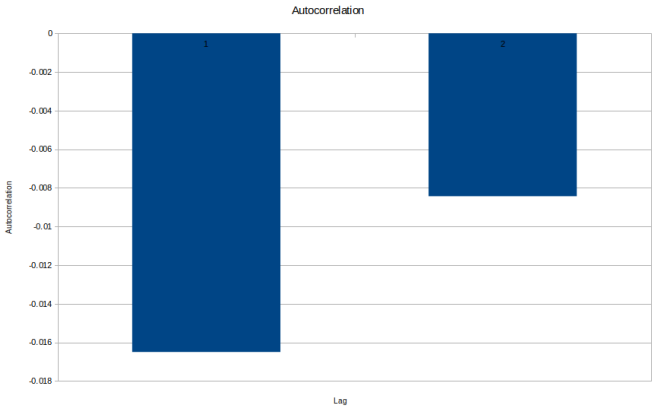
$$r_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{1}$$

where x is original data, y is the lagged version. Calculations are done inside the spreadsheet.

After calculating autocorrelation coefficient for lag 1 and lag 2 for both days we obtain the following plots. We find that there is low autocorrelation < 0.1 therefore the underlying random number generator is good.



(a) Autocorrelation for day 1



(b) Autocorrelation for day 2