A Study Of Math Department

Comp312 project individual report

Group 2
Author: Zhancheng Gan
Student ID: 300378961

Introduction

The system being studied is the math dependent, it is the place for students collect their Assignments. In this system, it only have a single server, where arrivals are determined by a Poisson process and job service times have an exponential distribution. In our team what we did in this project is observe the queueing in a system that provides service to customers who arrive randomly; To practice measuring, analyzing, and modeling a real queueing system; To analyze observations and extract any useful information. customer arrival process and the service channel discipline and service times; To construct both a queue model and a simulation model of the system for observation; To evaluate the performance measures of the system; also to compare the model's performance measures with those of the real system.

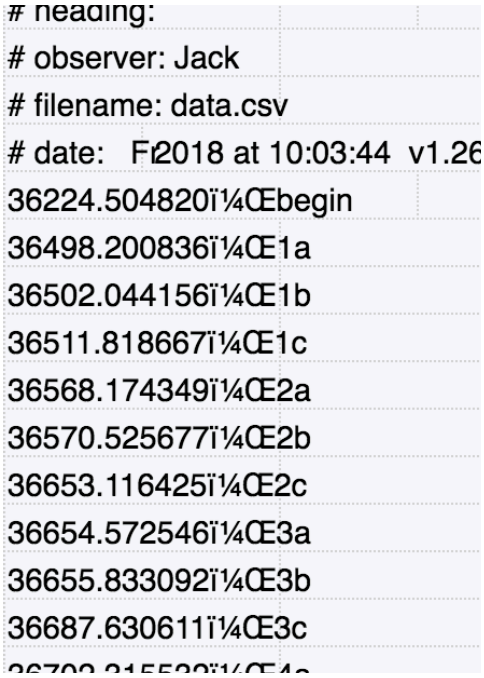
Description

In this team project I have activities in making observations; Organising data and Modelling the performance of the service unit.

Data collection

In data collection, the system we observed is open 4 hours per day on average. Our group has been collected 5 days data, and each people is responsible to collect 4hours data. At last, our team has been collected 20 hours data, there are 346 observations. During the data collection, we used the code it provides from the class. Recorded the data format as (1a 1b 2a 1c 2b 2c…) and so on. Where the numbers represent the first customer, second customer..the alphabet a, b, c represents the time when a customer arrives, the time when customer start service and the time when customer finish service. But we find out if there are so many people wait in the queue, it is hard to remember all the numbers for each people. Also, it is easy to miss record the leave time or arrive time for some of the customers, and some of the customers just leave when he thinks he has to for a long time. Therefore one of our members Nick is build a new data collected tool, in this program, it will automatically generate a queue. It can record the time when a customer arrives the system, and the queue time and service time are calculated in the program. It allows removing the people in the queue as well.

Organising Data



heading:
observer: Jack
filename: data.csv
date: Fr2018 at 10:03:44 v1.26
36224.504820ï¼Œbegin
36498.200836ï¼Œ1a
36502.044156ï¼Œ1b
36511.818667ï¼Œ1c
36568.174349ï¼Œ2a
36570.525677ï¼Œ2b
36653.116425ï¼Œ2c
36654.572546ï¼Œ3a
36655.833092ï¼Œ3b
36687.630611ï¼Œ3c

As we using 2 different data collecting program, It will lead to we got two different type of data format. So, I have to combine them together in excel. One of the format it already contains service time and queue time, so I have to do the transition in another format. Just did a bit of Subtraction we can get what we need. Using the corresponding C value minus B value will get the service time, B value minus A value will get the queue time and second arrive minus first arrive will get the inter arrive time. Such as (C1..N-B1..N, B1..N - A1..N and An+1…M - An…M). After we calculate those three values, we decide to clean up some abnormal value, like someone is waiting in the queue above 30min or someone takes 5min to service. Then what we left is the inter arrives values less than 1000 second and the service values less than 130 seconds. To calculate the parameter of Wq and Ws, it have to average all the waiting time in queue and all the waiting time in service. Add them up we got the total average waiting time in the system. Also, we can calculate the lamb and mu by 1/the expect inter-arrive time and 1/ the service time. Therefore we can use little law to calculate L Ls and Lq.

Modeling

To modeling the performance of the service unit, Here, we are using 3 different models to simulate the performance of our System. The first model we use is an M/M/1 model, this is because our system only has one service. The second model we use is the best fit distribution model, use the inter-arrival times best fit (with exponential distribution) and the service times best fit (with gamma distribution). The third, M3 model uses the empirical distribution of the inter-arrival time and the service time to simulate the performance of our System.

For M1 and M2, I did write a Simpy code to simulate those two models. Where lamb and mu I set is 1/ E(inter-arrive time) and 1/E(service time). The time unit I use is second because this is easy for me to clarify the difference between two models. The number of customers and maximum time I set is a very larger number like 10000, it should be given enough sample to do the simulation.

On the basis of M1 model I build up the M3 model, I was added load data functions and empirical distribution functions. Changed the distribution to empirical distribution of the inter-arrival time and the service time. To build empirical distribution function I have to get a list of independent, identically distributed real random variables(X) and the common cumulative distribution function (F), what I did is remove the duplicate values in queue time list and service time list, Which will become a new Service-timeX list and inter-arrive timeX without the duplicates. Such as (X1…Xn). Then the common cumulative

$$\widehat{F}_n(t) = \frac{\text{number of elements in the sample} \leq t}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i \leq t},$$

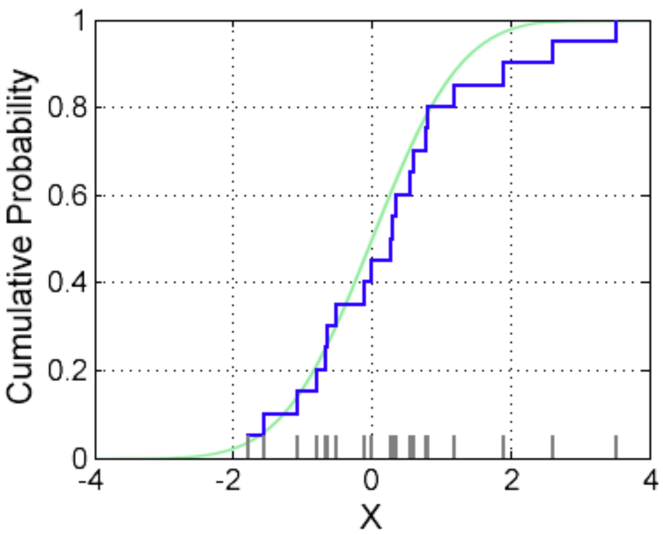
distribution function F(x) defines a number of elements in the sample < x / total number of elements. Once we got the X variables and F variables we can use the formula W[k-1]+((r-F[k-1])/(F[k]-F[k-1]))*(W[k]-W[k-1]) to get the simulate service time or queue time for each customer. Where r is a random number between 0 to 1. And the k value is the index of F list that larger than r.

	M1	M2	M3	**observation system**
W	52.370062977183	47.832741165061	65.012917049309	67.86803475
Wq	16.971467155669	12.396526048179	29.676469406179	32.73492416
WS	35.398595821513	35.436215116882	29.676469406179	35.13311059
L	0.4794858980176	0.4384528303535	0.5886484886771	0.6213550757
Lq	0.1554012774302	0.1136429753208	0.2687349171198	0.2996994292
LS	0.3240846205874	0.3248098550327	0.3199135715573	0.3216556465

Compare with M1and M2 model, there are quite similar. The only difference is the service times best fit gamma distribution, the inter-arrival distributions are the same. So, their output is quite similar as well. The waiting time in the queue between two models is just a few seconds. But to compare with the W in the system(67 seconds), which has a larger gap between.

I think this is because during the simulation we are using the exponential distributions such as assume all the customer arrive continuously and independently at a constant average rate. On the other hand, our observation system will have a rush time, normally it occurs during the break time between the lectures. In this time will have a lot of students come to collect their homework, that leads to suddenly increase the waiting time in the queue. The evidence is shown in the data, there have 25 people wait more than two min, it all occurs during the rush time. It also shows from the LQ, which is our observation system has the more average number of people in the queue.

I think we can not avoid this rush time problem, But we can improve by collect data in different time, especially when the service is not that busy. Like, we can collect the data from the system start server unlit the system end of the server, for all business hours. Not just the collecting the data during the middle period.



Overall, We find out the M3 model have the best performance overall. That is because empirical distribution is to do the simulation base on the observation data. In the above image, the green line is empirical distribution function, the blue line is the corresponding sample. As long as we have collected enough data, and do the data cleaning properly. The more standard data will have, the empirical distribution function we get will more close to the corresponding sample. Therefore the result we get will be more accurate.

like/dislike during my work on the project

I find out one of the very interesting things is sometimes there has a brief of time our M/M/1 system will suddenly change to M/M/2 system (occur when one of the staff is free, come up to help the staff in the counter). Especially when the time is busy. This will lead to having a double service speed of the system and a half waiting time in the queue. During the data collection, what we deal with this problem is removing the data for who went to the second server. But, this will still cause shorter waiting time in queue of this observation. Therefore, in theory, the Wq(average waiting time in queue) in M/M/1, it should be longer than Wq (average waiting time in queue) in observation. In the result, all the WS(around 35.40 seconds per people) are almost the same as the observation(35.13 seconds per people), but the estimate of M/M/1 WQ: 16.61 (second per people) compare to the observation WP: 32.84 （second per people）, the observation WP have much short time. I think this is another way to validity the rush time problem, although we did remove the data for who went to the second server. We can not assume that is exactly a M/M/1 system.

Difficulties

During this project, one of the difficulties I meet is when I start analyzing the data I find out our data did not fit with the gamma distribution. That is because gamma did not work when data have value 0. This Occur in the inter arrive table, I think the reason cause this problem is the rounding mistake in Java program (we set the value type as an integer). If the inter arrive time between two arrives is less than one second. It will be auto-round to 0. To ensure the integrity of the data, we discussed changes all the 0 values to 0.001 which means two customers never come at the same time. It fixes the problem for using arrive data to fit the gamma distribution.

Another difficulty I meet is when I Modelling the performance of the service unit,
In M3 empirical distribution, the python code I wrote will occur an negative delay time. Then I use the print statement to debug where the problems are, I find out this because in the arrive data there have a very small value 0.001 if we are substitute data into the formula W[k-1]+((r-F[k-1])/(F[k]-F[k-1]))*(W[k]-W[k-1]) when r is larger than F[k] and W[k-1] are very small. But the real reason is I did not return 0 when k equals to 0. After I added this condition to the code, the bug is fixed.

Start the project anew

If I am going to do start this project new, I will be collect data more rigorously. Because the data is the most important things in this project. It will reflect everything we do after. In the original data collection, the data are saved in two formats, one is the ABC format(a=when customer arrive, b=when customer start service, c = when customer finish service) and another format is (time, queue time, service time). This because we use two different program tools to collect the data. This situation leads to we have to waste a lot of time to combine two data together. In this time we collected 346 in total remove the error data caused by human, and the abnormal data, what we left is 304 data. Most of the data being removed are the data with high values. The average of service time after removes the abnormal data are decreasing about 10 seconds. It is hard to comment on is or not remove too much data after consider the rush time problem. If we have more time to collect the data, we can get even more complete information about our system, the result will become more accurate.

Considered the problems has been occurred, If I am start the project anew I will consider about the M/M/2 simulation, not just remove the data pretend there is an M/M/1 System.If I am doing this, the data will be better to represent the real situation, thus to get more accurate simulation. But, is hard to simulate when the system changes between M/M/2 and M/M1, also we have to know about the probability swapping the systems. But this probability is not fully depends on the the length of waiting queue, it also consider about the second stuff, in what situation she/he will come to help.

Personal Observations

I think Excel is a very useful tool in this project, it helps me a lot for organizing the data. I did spend a day to learn how to use excel. If there have a link to Excel tutorial in this project it might very helpful. Another thing I want to talk about is: if this use on commercial purpose might be quite fun. For example, the observation system small coffee shop normally it only has one server(one staff). But in the rush time, the waiting line becomes too long then customer start leaving. Now the boss plan to hire a part-time staff, the question is how long should part-time staff work on per week, at what time, can make the most income for the coffee shop? This might be fun because when the staff works on different time you will get the different income. Of course, let the part-time staff work on rush time are the most cost-effective. But how to measure the accurate time? Also, the system is swapping between m/m/1 and m/m/2. But the problem is you can not get the income details for the coffee shop.