

DATA304 Project Group 4: A study of the LAB cafe at Victoria University

Kevin Ye, Vivian Dong, Patrick Quito, Tama Hoare

May 2022

1 Introduction

1.1 Description of LAB Cafe

Outline system cafe what they sell, who their customers are, their day to day tasks for servers and number of servers. Kevin

1.2 Customer interaction with the system - Vivian

Customers need to choose one of the three counters to order and pay, and they will wait if there is a queue. After ordering they will stay in the café till the staff call their name means the order is ready.

1.3 Process of data collection

4 people collecting data different days approx 2 hours there collecting. Tama

1.4 Process of data cleaning

missing values getting the read data file to work. Patrick

2 Data analysis

2.1 Fitting best fit distributions (Vivian)

We tried to approximate "inter-arrival time" and "service time" using the following 12 Distributions: Weibull Minimum Extreme Value distribution, Normal distribution, Weibull Maximum Extreme Value distribution, Beta distribution, Inverse Gaussian distribution, Uniform distribution, Gamma distribution, Exponential distribution, Log-normal distribution, Pearson Type III distribution, Triangular distribution, Erlang distribution. After fitting different distributions, we checked the Goodness of fit based on Chi-square Statistics. The outputs for "inter-arrival time" sorted in order of Goodness of fit looks like this:

Table 1: Distributions listed by Betterment of fit

| Distribution | chi square |
|--|-------------|
| Pearson Type III distribution | 9.155252 |
| Weibull Minimum Extreme Value distribution | 13.245287 |
| Beta distribution | 21.708357 |
| Log Normal distribution | 25.596288 |
| Inverse Gaussian distribution | 29.389634 |
| Exponential distribution | 29.515278 |
| Gamma distribution | 48.359331 |
| Triangular distribution | 209.930441 |
| Normal distribution | 332.531278 |
| Uniform distribution | 510.690318 |
| Erlang distribution | 672.400334 |
| Weibull Maximum Extreme Value distribution | 1137.915014 |

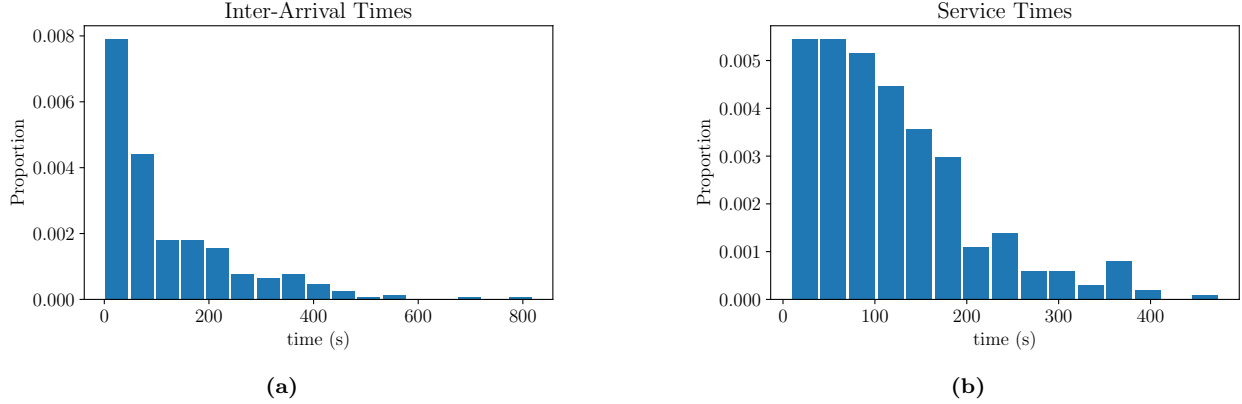
The outputs for "service time" sorted in order of Goodness of fit looks like this:

The Chi-square statistics suggest that the Pearson Type III distribution best approximates 'inter-arrival time'. We can also see that Beta distribution is the best fit for 'service time'. The python code using the Scipy Library to fit the distribution is from here: [Distribution Fitting Code](#) Suppose we had more time to do this part. In that case, we will add more distributions to fit our data and find a better fit distribution of the interarrival/service times. Furthermore, we can also use the Anderson-Darling test or other goodness-of-fit tests to compare whether we will get the same results.

Table 2: Distributions listed by Betterment of fit

| Distribution | chi square |
|--|-------------|
| Beta distribution | 1.231338 |
| Weibull Minimum Extreme Value distribution | 2.831316 |
| Pearson Type III distribution | 4.130412 |
| Gamma distribution | 4.131762 |
| Erlang distribution | 4.132443 |
| Inverse Gaussian distribution | 10.560874 |
| Log Normal distribution | 11.688749 |
| Exponential distribution | 29.775131 |
| Triangular distribution | 39.441479 |
| Normal distribution | 140.194689 |
| Uniform distribution | 305.594183 |
| Weibull Maximum Extreme Value distribution | 1080.829277 |

2.2 Histogram plots for visual evaluation (Patrick)

**Figure 1:** Histograms of inter-arrival times and service times

Here, we have divided the inter-arrival time and the service time data into a number of phases, such that the time occupied by each phase has a negative exponential distribution. This is used by Erlang regarding service to be completed in k phases which has the same negative-exponential distribution, thus obtaining what is called a k -Erlang distribution. The inter-arrival is the interval of time between each arrival. Assuming that the arrivals are independent, their distribution is exponential. The service times defines as the time required to serve a customer. These assumptions are further confirmed above by the histograms of the observations. We have observed the arrival of customers for a few hours.

The second pair of histograms above are fitted with the best fit distributions. We have fitted the inter-arrival and service time histograms with a list of distributions listed by Betterment of fit. Based on the Chi-square Statistics, it suggests that Pearson3 is the best in approximating 'inter-arrival time' Data. For our 'service times' data, it suggests that beta distribution is the best fit for the data.

3 Simulation models

3.1 Performance Measures of collected data (Tama)

Assessing the performance measures of the collected data is key modelling the performance of the servicing unit. As shown in the tables above there were seven key performance measures assessed from the collected data.

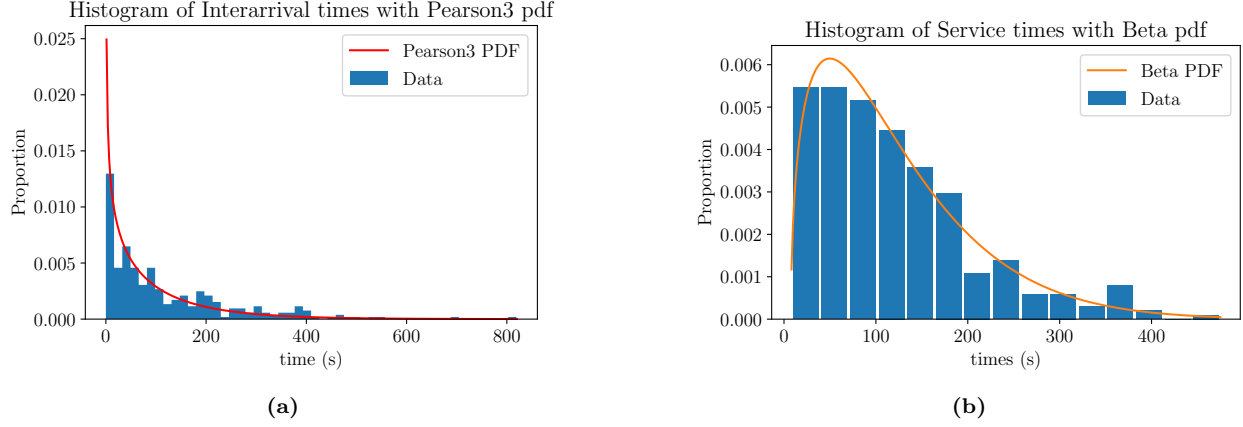


Figure 2: Histograms with best fit distribution pdf overlayed

Table 3: This is the caption that goes at the top of the table

| Performance Measures | Values calculated from data |
|---|-----------------------------|
| Average time in system (seconds), W | 140.07 |
| Average number of customers in the system, L | 1.1819 |
| Proportion of time servers are busy, B | 0.61148 |
| Effective arrival rate (per second), λ_{eff} | 0.0084381 |

Firstly, we can see that the average time that a customer spent at the café service unit, meaning the total time of queuing plus service, was 140,07 seconds. The average number of customers in the system at any given point was 1.18. We can see from the data that the server at the café was busy serving customers 61.15

The performance measure of effective arrival rate corresponds to the rate at which customers enter the service unit. We can see that 0.0084 customers entered the café per second. The correlates to 30.38 customers coming to the café per hour.

Table 4: This is the caption that goes at the top of the table

| Other parameters | Values calculated from data |
|--|-----------------------------|
| Average Inter-arrival time $\frac{1}{\lambda}$ (seconds) | 120.329 |
| Average Service time, W_s (seconds) | 120.77 |
| Average Queue Time, W_q (seconds) | 19.295 |

The average inter-arrival time is the average time difference between arrival of one customer and then the next customer. It is a time elapse between the arrival of the person and one following it in the queue. The value calculated from this in the data was 120.33 seconds.

The time it takes for one customer to be served is referred to as the service time. This is the time between a single customer reaching the server (no longer in the queue) and leaving the service unit. Hence it took on average 120.77s for a customer to be served at the café once they had reached the server.

The average queue time is the time between a customer arriving at the service system and getting to the point at which they start to get served. They spend that time in a queue. At the café it took on average 19.295 seconds for a customer to get to the server once they had arrived at the café. Hence they spent of average 19.295 seconds in a queue at the café.

3.2 M1 model (Patrick)

The M1 model is a task where we investigate and use either M/M/1 or M/M/C depending on our structure. An M/M/1 queue represents the queue length in a system having a single server, where arrivals are determined by a Poisson process and service times gave an exponential distribution. In the case of our project, we have an M/M/3 queue as there were 3 servers from where we gathered our data form.

$$\pi_0 = \frac{1}{\sum_{k=0}^{s-1} \frac{\rho^k}{k!} + \frac{\rho^s}{s!} \frac{1}{1-\frac{\rho}{s}}}$$

$$\pi_0 = \frac{1}{\frac{\rho^0}{0!} + \frac{\rho^1}{1!} + \frac{\rho^2}{2!} + \frac{\rho^3}{3!} \frac{1}{1-\frac{\rho}{3}}}$$

$$\pi_0 = \frac{1}{1 + \rho + \frac{\rho^2}{2} + \frac{\rho^3}{6} \frac{1}{1-\frac{\rho}{3}}}$$

$$\pi_0 = 0.3690202951$$

$$B = 1 - \pi_0 = 0.6309797049$$

$$L = \pi_0 \frac{\frac{\rho^{s+1}}{s!s}}{(1 - \frac{\rho}{s})^2} + \rho$$

$$L = \pi_0 \frac{\frac{\rho^4}{3!3}}{(1 - \frac{\rho}{3})^2} + \rho$$

$$L = 1.033745189$$

$$W = \frac{L}{\lambda} = 124.3899904$$

Upon generating all the values (check the table below), we have used the formula to compare to the original data we've gathered. This is used to confirm that the values from the M1 model are correct after doing the math.

Table 5: Comparing performance measures of Collected data and M1 model

| Performance Measures | Collected Data | M1 model |
|------------------------|----------------|-----------|
| W | 140.07 | 124.33 |
| L | 1.1819 | 1.0435 |
| B | 0.61148 | 0.63064 |
| λ_{eff} | 0.0084381 | 0.0083952 |

In creating the M1 model, we took inspiration from the assignments and labs to be able to produce these outcomes. We have passed our data on arrival and service rate through the model and got the values above. We had a mean arrival rate of 0.008310517468785045 per second, and a mean service rate of 0.008279591609419837 per second. The LAB Cafe is open from 7:30 am to 3:00 pm a total of 7.5 hours per day which is equivalent to 27000 seconds which would be the max time of each simulation. W is denoted as the average waiting of customers, L is the average number of customers in the system, B is the proportion of time that servers are busy, and lambdaEff is the effective arrival rate.

3.3 M2 model (Vivian)

The M2 model is a simulation model developed using SimPy to model the LAB cafe customer waiting and serving system. M2 model uses the interarrival times best fit distribution (Pearson3) and the service times best fit distribution (Beta) to simulate the performance. Below is the comparison between the original data performance measure estimates and the performance measures estimates produced by M2 model.

Table 6: Comparing performance measures of Collected data and M2 model

| Performance Measures | Collected Data | M2 model |
|------------------------|----------------|----------|
| W | 140.07 | 143.28 |
| L | 1.1819 | 1.6283 |
| B | 0.61148 | 0.71400 |
| λ_{eff} | 0.0084381 | 0.011347 |

From the table above, we can see that the estimated W from the M2 model has a difference of approximately 3.21 from the estimate provided by the original data collected. The L difference between the two estimations by the collected data and the M2 model is about 0.45. The difference in the proportion of time servers are busy (B) was 0.1 between the two estimates of the collected data and the M2 model. The effective arrival rate λ_{eff} has a difference of 0.0029 approximately between the original data and the M2 estimate. We can see that the M2 model is a decent fit for the actual data as the differences are relatively small.

3.4 M3 model (Kevin)

Table 7: Comparing performance measures of Collected data and M3 model

| Performance Measures | Collected Data | M3 model |
|------------------------|----------------|-----------|
| W | 140.07 | 127.14 |
| L | 1.1819 | 1.0853 |
| B | 0.61148 | 0.62465 |
| λ_{eff} | 0.0084381 | 0.0085284 |

The M3 model is a simulation model developed using SimPy to model the LAB cafe customer waiting and serving system. The distribution of interarrival and service times are modelled after the empirical distributions of the interarrival times and services times recorded from the original data. From the M3 model produced some performance measures estimates in the table above which we can compare to the original data performance measure estimates to gauge how well of a fit this M3 model is at simulating the nature of the real life system.

From the table we can see that estimated W from the M3 model has a difference of approximately 13 to the estimate provided by the original data collected. The L difference between the two estimations by the collected data and the M3 model is about 0.1. The difference in the B , proportion of time servers are busy was 0.01 between the two estimates of the collected data and the M3 model. The effective arrival rate λ_{eff} has a difference of 0.0001 approximately between the original data and the M3 estimate. We can see that The M3 model is a decent fit for the original data as the differences are around about 10% of the original data estimates.

4 Conclusion

4.1 Results Summary

4.2 Improvements

4.3

4.4