Stochastic Analysis of a Bike-Sharing System

Ming Lyu

Project Advisor: Dr. John McSweeney

A DISSERTATION SUBMITTED TO THE DEPARTMENT OF MATHEMATICS OF ROSE-HULMAN INSTITUTE OF TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE

Department of Mathematics
Rose-Hulman Institute of Technology
August 23, 2020

1 Introduction

Bike-sharing systems contribute to the urban public transportation system, providing bikes with convenience and accessibility. In certain bike sharing system, people can borrow bikes from one dock in the system, and return it to another dock for free or with low charges. Bike-sharing systems provide a convenient way for short-distance travelling and an alternative and healthy method of transportation. In recent years, bike-sharing systems have grown rapidly in cities all over the world.

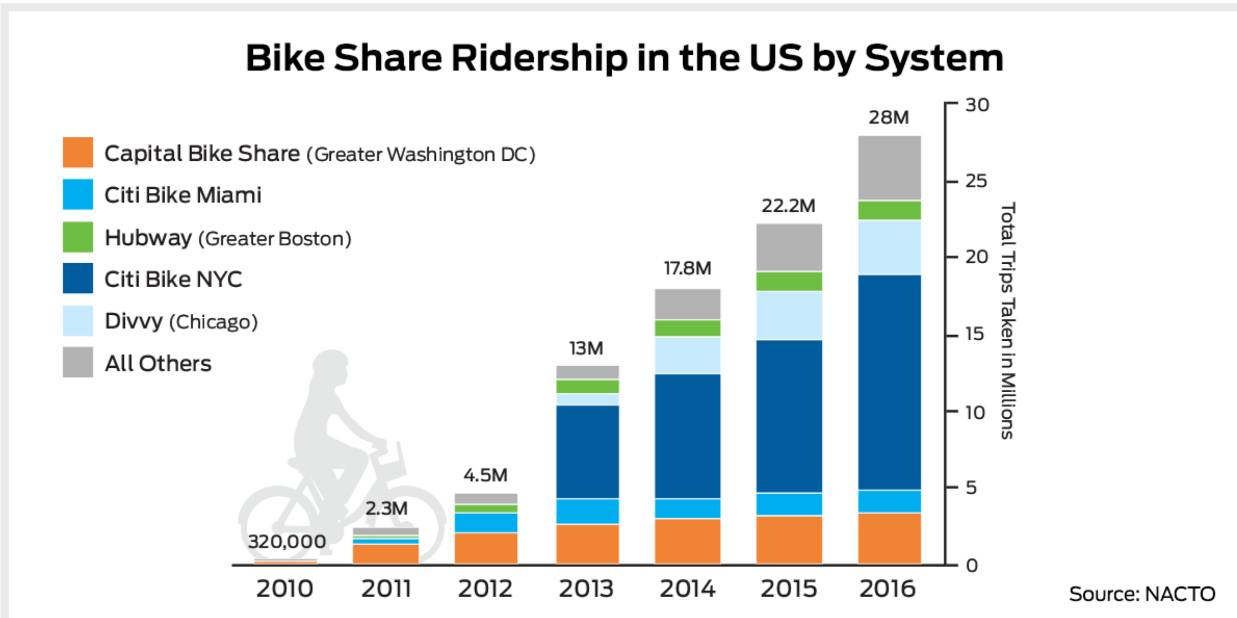

Figure 1: Bike Share Ridership in the US[1]

Data tracking the bike-sharing system is available online. For example, table 1 is a display of station data of a bike-sharing system in San Francisco Bay area. It records the name of the station, station's location (by latitude and longitude), number of docks and other information. There is also data tracking the information of bike rides in the bike-sharing system that we can use to analyze the bike-sharing system or to apply to certain models.

Table 1: Sample Station Data in Bike Sharing System

stationid	name	lat	long	dockcount	landmark	installation
2	San Jose Diridon Caltrain Station	37.329732	-121.901782	27	San Jose	8/6/2013
3	San Jose Civic Center	37.330698	-121.888979	15	San Jose	8/5/2013
4	Santa Clara at Almaden	37.333988	-121.894902	11	San Jose	8/6/2013
5	Adobe on Almaden	37.331415	-121.8932	19	San Jose	8/5/2013
6	San Pedro Square	37.336721	-121.894074	15	San Jose	8/7/2013
7	Paseo de San Antonio	37.333798	-121.886943	15	San Jose	8/7/2013
8	San Salvador at 1st	37.330165	-121.885831	15	San Jose	8/5/2013
9	Japantown	37.348742	-121.894715	15	San Jose	8/5/2013
10	San Jose City Hall	37.337391	-121.886995	15	San Jose	8/6/2013

The interest is to simulate the bike-share system using mathematical models to explore some properties of the system using Markov Process. One of the goal is to minimize the total waiting time of customers in the system by choosing appropriate parameters such as arrival rate and bike return rate. Furthermore, testing the model using real-world data is also an essential part of this project.

2 Mathematics

A continuous-time and discrete-space Markov Process is an appropriate modeling choice. There are finite number of states, which is constrained by the number of bikes in the system. The continuous-time part indicates that the time variable is continuous. Since state changes does not happen exactly at discrete time values such as 1 or 2, continuous-time is appropriate for modeling option.

2.1 Continuous-Time and Discrete-Space Markov Process

Consider a stochastic process $X(t)$ defined for $t \in [0, \infty)$, and for each $t, X(t) \in S \subseteq \mathbf{Z}$, where the state space S is a set of integers. We can therefore describe this process as being a continuous-time and discrete-space process. If for any $s, t > 0$, and $i, j, i_u, 0 \leq u < s$, it is the case that

$$P(X(t+s) = j | X(s) = i, X(u) = i_u, \text{ for } 0 \leq u < s) = P(X(t+s) = j | X(s) = i),$$

Then $X(t)$ is a continuous-time, discrete-space Markov Process.

2.2 Stationary Distribution

A stationary distribution of discrete Markov Chain with transition matrix P is a probability distribution π that remains constant in the Markov chain as the time progresses. We denote this as,

$$\pi = \pi P,$$

If such π exists, then $\pi P^2 = (\pi P)P = \pi P = \pi$. Similar proof using induction can be applied to the case $\pi P^n = \pi$. If the Markov chain reaches the stationary distribution π at any point t_0, it remains constant for all $t \geq t_0$.
If we apply the concept of stationary distribution to a continuous-time Markov Chain, we first define q_{ij} as the transition rate, the rate a continuous time Markov chain moves between states.

$$q_{ij} = \lim_{h \to 0} \frac{1}{h} P(X(t+h) = j | X(t) = i),$$

Then for a stationary distribution π_i, we have,

$$\pi_i \sum_{j \neq i} q_{ij} = \pi_j \sum_{i \neq j} q_{ji},$$

The flow in rate equals to the flow out rate at each state so that the distribution stays constant.

3 Table of Symbols/Nomenclature

Symbol	Name
i	# of bikes in use
$S(i, j)$	i = # bikes in use, j = # bikes at station 1
π_k	Stationary distribution at state k, k represents of bikes in use
$\pi_{(k,c)}$	Stationary distribution at state (k, c), k represents of bikes in use and c represents the number of people waiting in line to pick up bikes
C_{cost}	Total cost
$C_{i,capacity}$	Capacity of station i
C_{wait}	Maximum of the waiting list to pick up bikes
T_{wait}	Total waiting time of all customers in Bike Share system
b	# of bikes in the Bike Share system
p	# of stations in the Bike Share system
λ_i, λ	Possion arrival distribution with rate λ at station i or λ means there is only one station in the system
μ_i,μ	Trip length exponential distribution with mean rate μ at station i or μ means there is only one station in the system

4 Model

In the section, check section 3 for symbol reference.

4.1 Derivation From M/M/1 Queuing Problem Model

M/M/1 queuing problem model is that there is a single service counter, which has an exponential service time distribution with a mean service rate of μ customers per minute. The customers arrives with a Possion arrival distribution with rate λ. Since we just have one station in our bike sharing system, we apply the same idea to model the arrival/leave process in the bike sharing system. We describe the model as the following:

$$i \to i + 1, \lambda$$

$$i \to i - 1, \mu$$

The stationary distribution exists iff $\lambda < \mu$ with,

$$\pi_n = \frac{(\lambda/\mu)^n}{1 + 1/(1 - \lambda/\mu)}, n = 0, 1, 2, ...$$

If the queue has finite capacity N, then we have a stationary distribution regardless of μ and λ and π_n depends on the capacity N,

$$\pi_n = \frac{(\lambda/\mu)^n}{1 + 1/(1 - \lambda/\mu)}, n = 0, 1, 2, ..., N$$

4.2 Derivation From Birth and Death Process Model

The death and birth process model assumes that the birth rate and death rate are proportional to the population instead of being constant numbers all the time. In the bike sharing system, we assume that the rate of people arriving at the bike station is constant but the rate of people returning bikes to the station is proportional to the number of people using the bikes. It is logical to assume that the more people using the bikes, the more likely the number of people returning the bikes will increase. We describe the model as the following:

4.2.1 One Station

We have the transitions as the following,

$$i \to i+1, \lambda$$

$$i \to i-1, \mu i$$

We have stationary distribution with equation,

$$\mu\pi_1 - \lambda\pi_0 = 2\mu\pi_2 - \lambda\pi_1 = 3\mu\pi_3 - \lambda\pi_2 = ... = 0.$$

Then we have,

$$\mu\pi_1 - \lambda\pi_0 = 0,$$
$$\pi_1 = \pi_0\frac{\lambda}{\mu},$$
$$2\mu\pi_2 - \lambda\pi_1 = 0,$$
$$\pi_2 = \frac{\lambda\pi_1}{2\mu},$$
$$\pi_2 = \pi_0\frac{\lambda^2}{\mu^2 1 \times 2},$$
$$\vdots$$
$$\pi_n = \pi_0\frac{\lambda^n}{\mu^n n!}.$$

We can rewrite the equation as,

$$\sum_{i=0}^{\infty}\pi_i = \pi_0\sum_{n=0}^{\infty}\frac{\lambda^n\pi_0}{\mu^n n!} = 1$$

By applying Taylor Series,we know that,

$$\sum_{i=0}^{\infty}\frac{\lambda^n}{n!} = \mathbf{e}^{\lambda},$$

$$\sum_{i=0}^{\infty}\frac{\lambda^n\pi_0}{\mu^n n!} = \pi_0\sum_{i=0}^{\infty}\frac{(\lambda/\mu)^n}{n!} = \pi_0\mathbf{e}^{\lambda/\mu}$$

Therefore, we have

$$\pi_0(\pi_0 \mathbf{e}^{\lambda/\mu}) = 1$$
$$\pi_0^2 = \mathbf{e}^{-\lambda/\mu}$$
$$\pi_0 = \sqrt{\mathbf{e}^{-\lambda/\mu}}$$

.

The above is from the perspective of how many bikes at the station with infinite patience. We optimize the model by limit the number of people that can wait to pick up the bikes. We use $\pi_{(k,c)}$ here to describe the state. k represents the number of bikes in use and c represents the number of people waiting to pick up bikes. Notice that when $k < b$, c is always 0 as there are bikes at the station. $c > 0$ can happen when $k = b$.

$$\mu\pi_{(1,0)} - \lambda\pi_{(0,0)} = 2\mu\pi_{(2,0)} - \lambda\pi_{(1,0)} = ... = b\mu\pi_{(b,0)} - \lambda\pi_{(b-1,0)} =$$

$$b\mu\pi_{(b,1)} - \lambda\pi_{(b,0)} = ... = b\mu\pi_{(b,C_{wait})} - \lambda\pi_{(b,C_{wait}-1)} = 0$$

$$\pi_{(k,0)} = \frac{\lambda\pi_{(k-1,0)}}{k\mu}, k = 0, 1, 2, 3, ..., b,$$

$$\pi_{(b,z)} = \frac{\lambda\pi_{(b,z-1)}}{b\mu}, z = 0, 1, 2, 3, ..., C_{wait}$$

By solving this recurrence relationship, we have,

$$\pi_{(1,0)} = \frac{\lambda}{\mu}\pi_{(0,0)},$$

$$\pi_{(2,0)} = \frac{\lambda^2}{\mu^2 2!}\pi_{(0,0)},$$

$$\vdots$$

$$\pi_{(b,0)} = \frac{\lambda^b}{\mu^b b!}\pi_{(0,0)},$$

$$\pi_{(b,1)} = \frac{\lambda^{b+1}}{\mu^b b!(b\mu)}\pi_{(0,0)},$$

$$\vdots$$

$$\pi_{(b,C_{wait})} = \frac{\lambda^{b+C_{wait}}}{\mu^b b!(b\mu)^{C_{wait}}}\pi_{(0,0)} = \frac{\lambda^{b+C_{wait}}}{\mu^{b+C_{wait}} b! b^{C_{wait}}}\pi_{(0,0)}$$

Since they sum up to 1, we have,

$$\sum_{i=0}^{b} \frac{\lambda^i}{\mu^i i!}\pi_{(0,0)} + \sum_{j=1}^{C_{wait}} \frac{\lambda^{b+j}}{\mu^{b+j} b! b^j}\pi_{(0,0)} = 1$$

With this stationary distribution, we can set up goal to minimize,

$$\sum_{j=1}^{C_{wait}} j\left(\frac{\lambda^{b+j}}{\mu^{b+j} b! b^j}\pi_{(0,0)}\right) + \sum_{i=0}^{b-C_{wait}-1} (b - C_{wait} - i)\frac{\lambda^i}{\mu^i i!}\pi_{(0,0)},$$

which is the sum of expected number of person at waiting list of picking up bikes and return bikes.

4.2.2 Two Stations

Let the U represent the total number of bikes in the bike sharing system.
(i, j) = state of i bikes in use and j bikes at station 1.
Since we just have two bike stations in the system, this 2-dimensional state space is sufficient. With fixed number of bikes in the system and known number of bikes in use and bikes at station 1, we can calculate the number of bikes at station 2. Thus, the number of bikes at station 2 can be represented by i and j, which is $U - i - j$.
Suppose that the arrival rates at both stations are the same and both stations have the same capacity C.
Transition Rate:
General Cases:

$$(i, j) \to (i + 1, j), \lambda, \ \forall i, j, i + j < U$$

User picks up a bike from station #2

$$(i, j) \to (i + 1, j - 1), \lambda, \ \forall i, j, i + j < U, j \neq 0$$

User picks up a bike from station #1

$$(i, j) \to (i - 1, j + 1), \mu i, \ \forall i, j, i + j \leq U, i \neq 0$$

User returns a bike to station #1

$$(i, j) \to (i - 1, j), \mu i, \ \forall i, j, i + j \leq U, i \neq 0$$

Boundary Cases:

$$(0, 0) \to (1, 0), \lambda$$

$$(1, 0) \to (0, 0), \mu$$

$$(0, C) \to (1, C), \lambda$$

$$(0, C) \to (1, C - 1), \lambda$$

$$(1, C) \to (0, C), \mu$$

$$(1, C - 1) \to (0, C), \mu$$

$$(U, 0) \to (U - 1, 0), \mu U$$

$$(U, 0) \to (U - 1, 1), \mu U$$

$$(U - 1, 0) \to (U, 0), \lambda$$

$$(U - 1, 0) \to (U, 1), \lambda$$

5 Simulation

5.1 Derivation From M/M/1 Queuing Problem Model with Finite Patience

We extend the model from 4.1. If the users come to the station to pick up a bike and there is no bikes at the station, they will wait until bikes show up with infinite patience. The same works for returning bikes. We keep track of the number of pick-up waiting list and return waiting list as well as number of bikes in use and number of bikes at the station at each time step. For this model, we divide the simulation into three situations, where $\lambda = \mu, \lambda > \mu$ and $\lambda < \mu$. The following sections present the results of simulation and brief analysis of the results. There is only one bike station in the bike sharing system. For this particular simulation, there are total of 30 bikes in the bike sharing system. The capacity of the station is 20. The simulation starts with 20 bikes at the station and 10 bikes in use.

5.1.1 $\lambda = 2, \mu = 2$

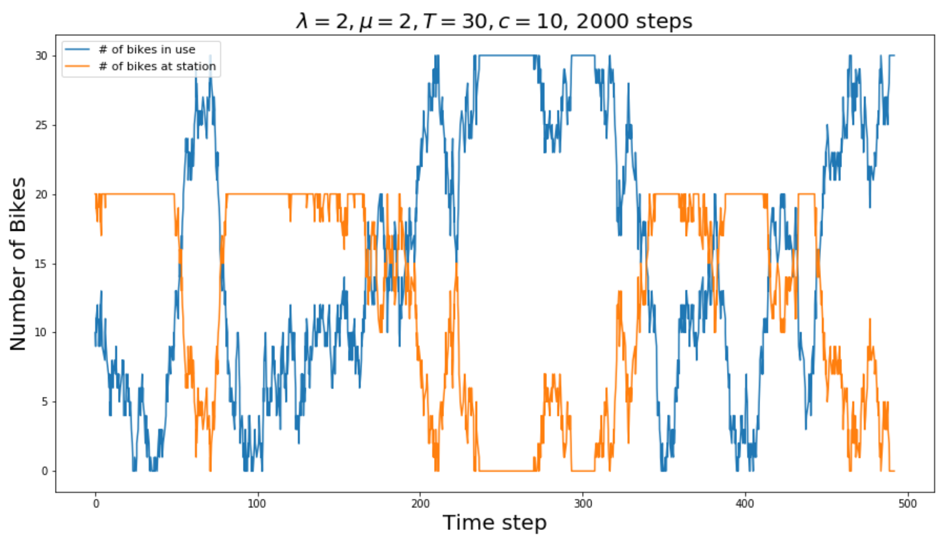

Figure 2: Number of Bikes At Station and Number of bikes in Use vs Time Step

Since λ and μ have the same value in this case, we see random behavior in this graph. The rate of people picking up bikes and returning the bikes are the same and hence we do not see stable behavior of number of bikes at station. Also, we can see a reflection of number of bikes at station. It only happens when there is only one station in the system as number of bikes in use and number of bikes in the station sums up to a constant. In this case, they sum up to 30.

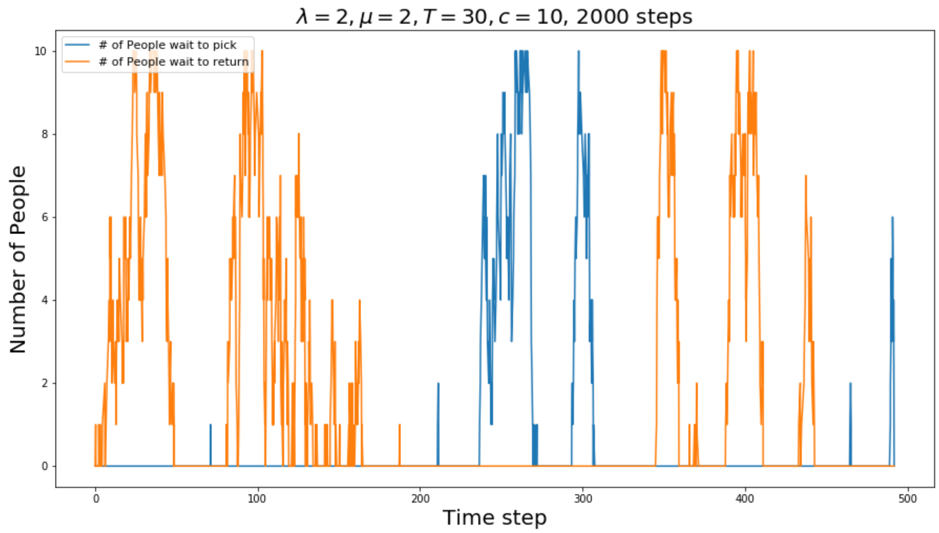

Figure 3: Number of People Wait to Pick Up and Return a Bike vs Time Step

Since λ and μ have the same value in this case, we see random behavior in this graph. Besides, when the number of people waiting to return bikes to the station is greater than 0, the number of people waiting to pick up bikes at station should be 0 and vice versa.

5.1.2 $\lambda = 2.5, \mu = 2$

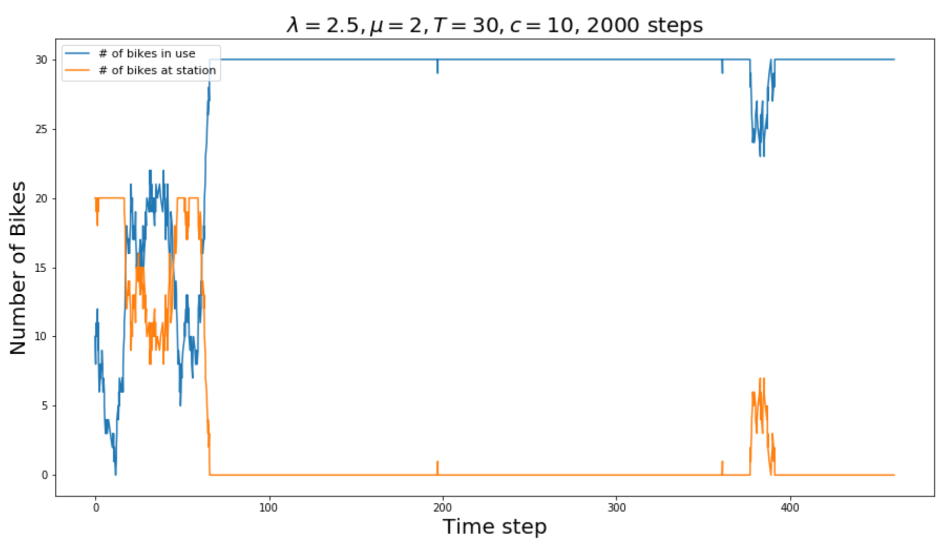

Figure 4: Number of Bikes At Station and Number of bikes in Use vs Time Step

Since λ is greater than μ in this case, we see that the number of bikes in use increase rapidly to 30. The rate of people picking up bikes is greater than the rate of returning the bikes and we see such pattern. After it reaches 30, which is the total number of bikes in the system, it reaches the stationary distribution afterwards.

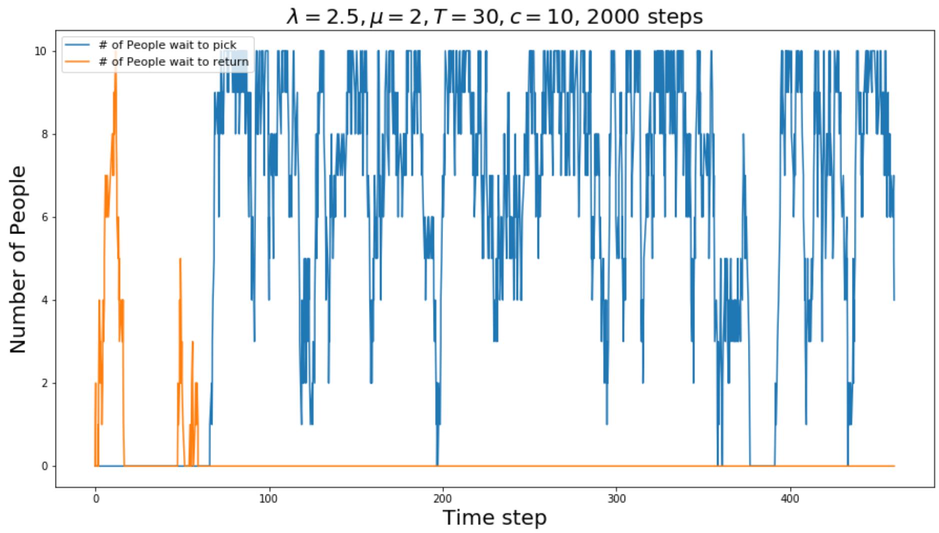

Figure 5: Number of People Wait to Pick Up and Return a Bike vs Time Step

Since λ is greater than μ in this case, the system is popular and people are more likely to pick up bikes than to return bikes. It causes that most of the time, people will be waiting to pick up bikes. In the graph, we can see that the number of people waiting to pick up bikes is greater than 0 and it indicates that people are waiting in line to pick up bikes.

5.2 Derivation From Birth and Death Process Model with Finite Patience

The simulations of this section is based on the model from section 4.2.1, where the return rate was proportional to the number of bikes in use. Since there are many combinations of λ and μ that we can choose, we only explore one of them to validate our solutions of stationary distribution. For the set up, we have,

$$C_{wait} = 10, b = 30, p = 1, C_{1,capacity} = 20$$

5.2.1 $\lambda = 1, \mu = 0.05$

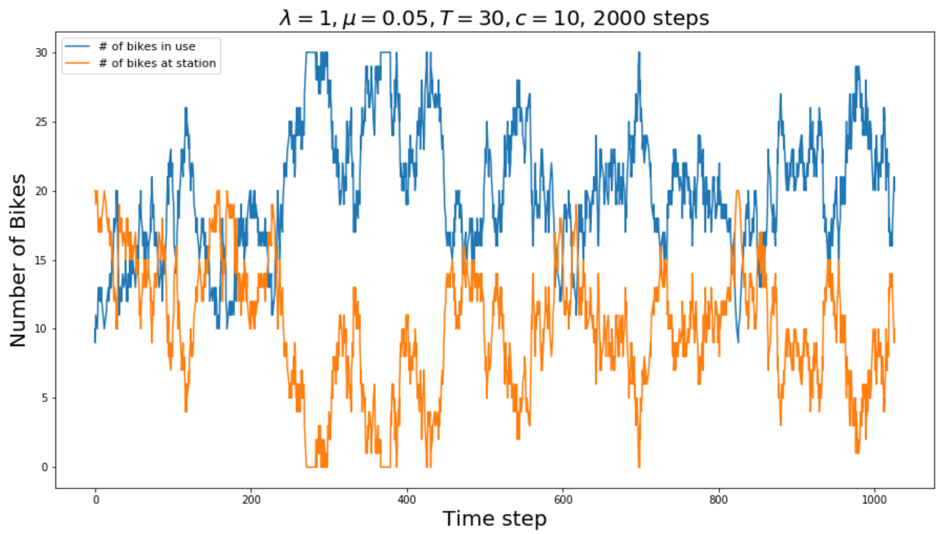

Figure 6: Number of Bikes At Station and Number of bikes in Use vs Time Step

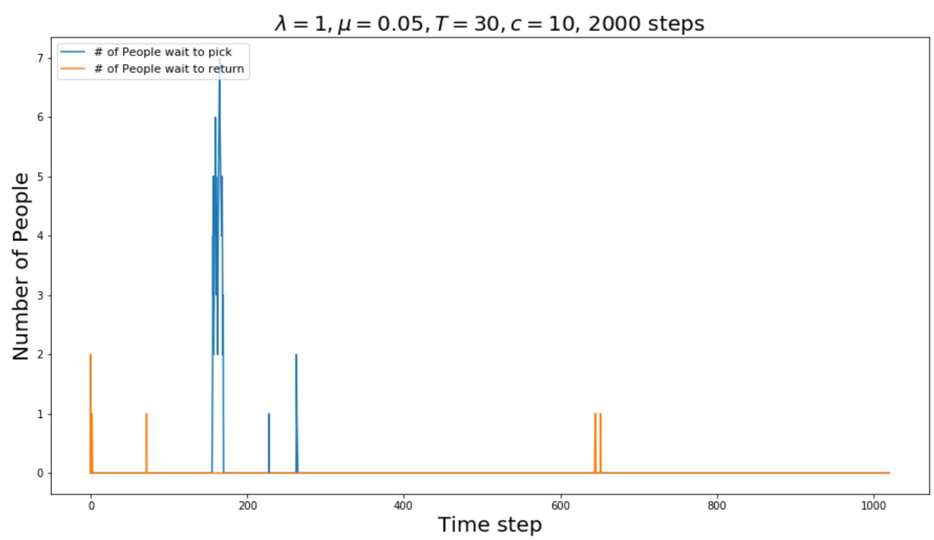

Figure 7: Number of People Wait to Pick Up and Return a Bike vs Time Step

We choose a good combination and μ and λ so that the almost no body have to wait to pick up bikes or return bikes according to Figure 7. There is a symmetric pattern in Figure 6. Notice that since there is only one station in the system, number of bikes in use and number of bikes at the station always sum up to b. That causes the pattern to be symmetrical. Next, we will explore the stationary distribution of the system. To better view the distributions, we set

$$\lambda = 1, \mu = \frac{1}{35}$$

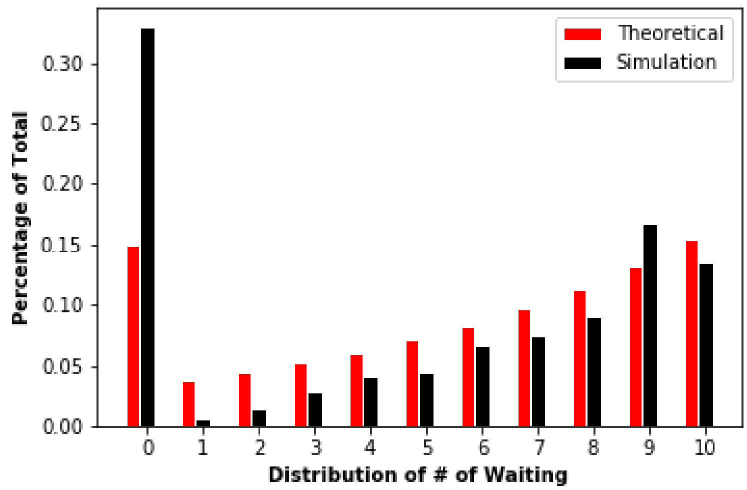

Figure 8: Distribution of Waiting Time with 500 Steps

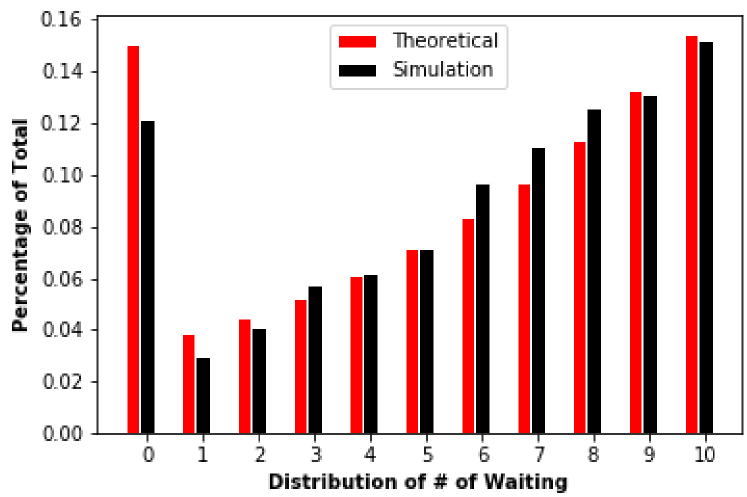

Figure 9: Distribution of Waiting Time with 2000 Steps

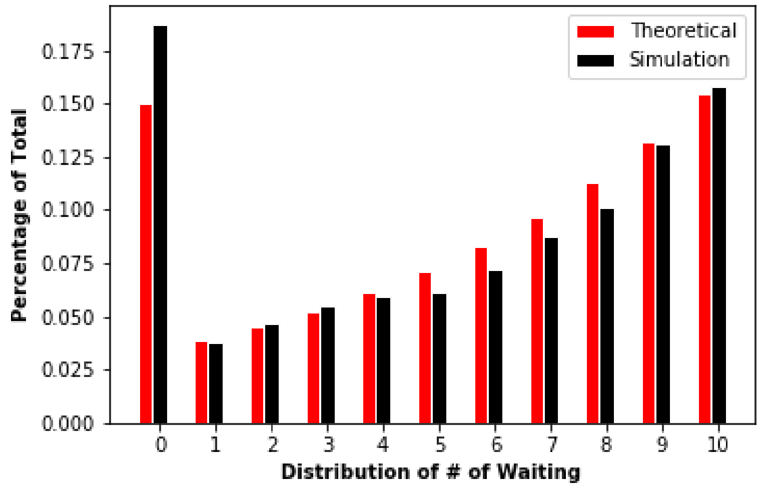

Figure 10: Distribution of Waiting Time with 8000 Steps

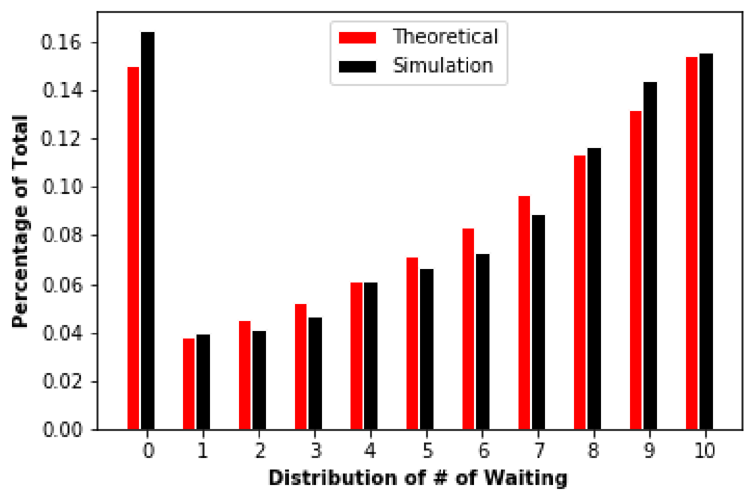

Figure 11: Distribution of Waiting Time with 20000 Steps

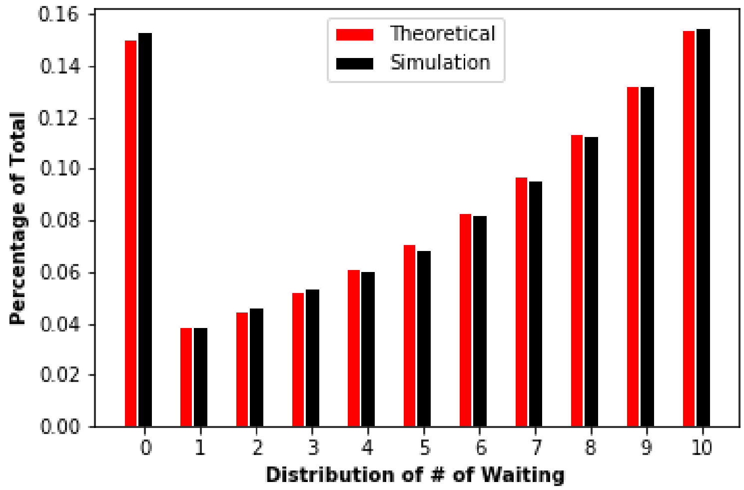

Figure 12: Distribution of Waiting Time with 100000 Steps

The red bar shows the theoretical distribution of the waiting time while the black bar shows the theoretical stationary distribution calculated form section 4.2.1. From Figure 8 to Figure 12, we can see that as we increase the number of steps, the empirical distribution is approximating the theoretical stationary distribution. That further proves that our solution is right. Next, we explore the combination of λ and μ that minimizes the total waiting time.

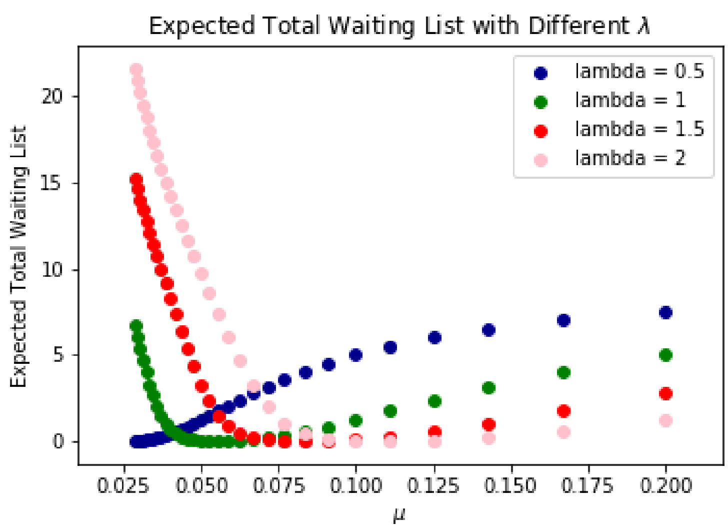

Figure 13: Total Waiting Time with $\lambda = 1$

Based on the graph, we can see that with around $\mu = 0.05$, we reaches the local minimum when $\lambda = 1$. It indicates that we have about 10 bikes in the station when reaching stationary distribution. We can think of it as $\lambda = k\mu$. For this case, we have $\lambda = 1$. Then, we have $k = \frac{1}{0.05} = 20$. Since there are 20 bikes in use, we have 10 bikes at the station. 10 is half of the station capacity, which leaves both buffers for people who wants to pick up bikes and return bikes. Likewise, we can observe the same performance with different λ values. It reaches the local minimum when the combination of λ and μ results in half of the capacity of the station to be filled when reaching stationary distribution.

6 Data

The data set records the information of a bike share system in Bay area for two years. It includes the information of stations, usages of each bike, duration of each trip and so on. Next, we will explore this data set and extract useful information that we can use on our model.

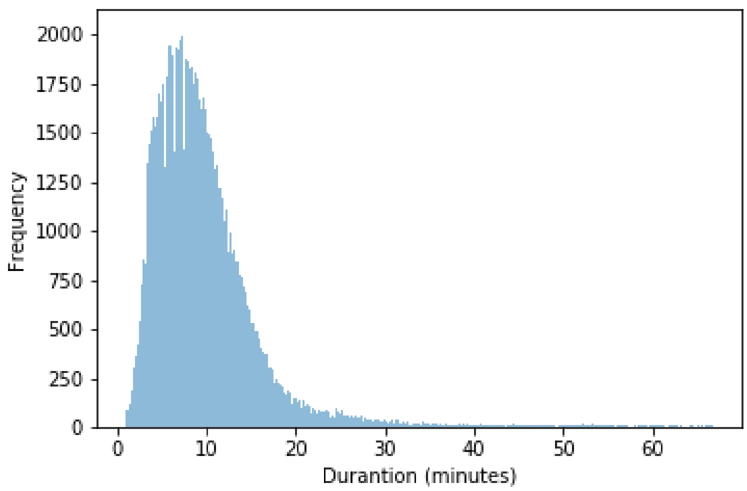

Figure 14: Time Duration of Bike Rides

The graph is the distribution of trip duration of the whole data set. In our model, we assume that the trip duration follows a exponential distribution. In this graph, we can see that trip duration has a bell-shaped distribution. Although we can do an approximation on the distribution in the graph using Gamma distribution, it then breaks the memoryless property of Markov process.

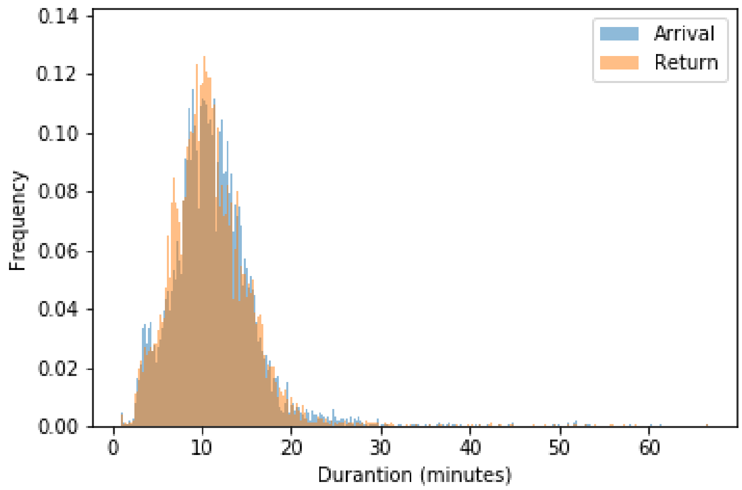

Figure 15: Time Duration Distribution of Arrival Trip Duration and Departure Trip Duration

This graph is the Time Duration Distribution of Arrival Trip Duration and Departure Trip Duration of station San Francisco Caltrain (Townsend at 4th). Although there are some difference, the shapes of the distributions are similar in terms of frequency percentage. It indicates that we can view departure trips the same as the arrival trips in our model.

Using the data available, we find that the arrival rate $\lambda = 0.3$, return rate $\mu = 0.09$, number of bikes in the system $T = 444$ and capacity of the station $C_{capacity} = 20$.

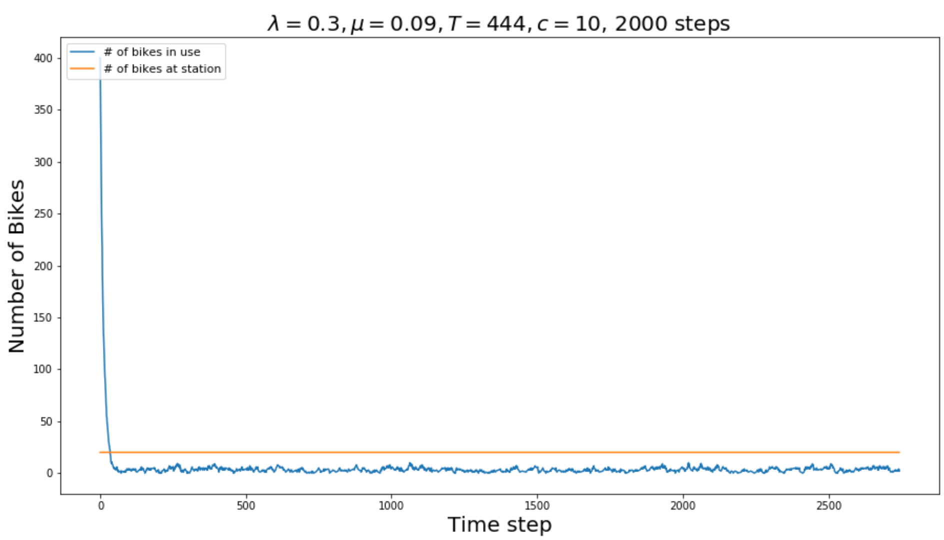

Figure 16: Number of Bikes At Station and Number of bikes in Use vs Time Step

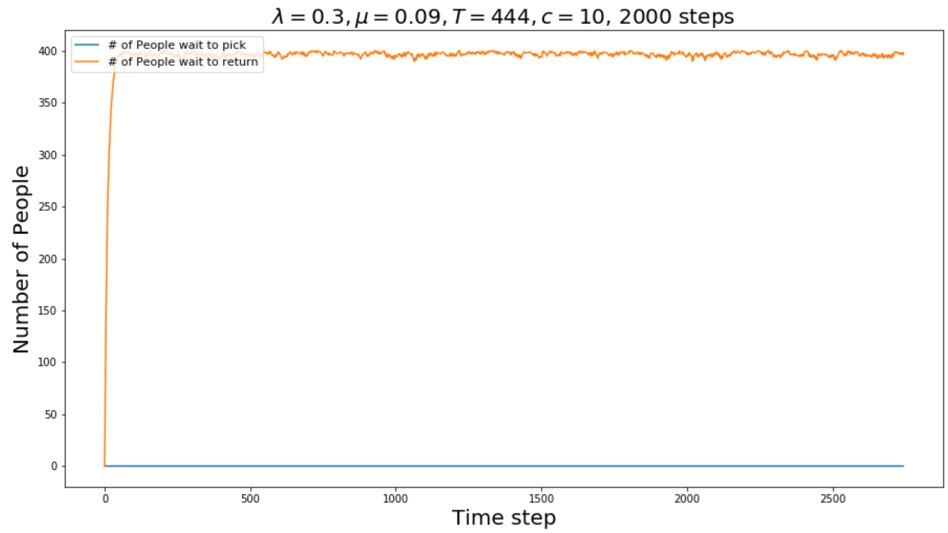

Figure 17: Number of People Wait to Pick Up and Return a Bike vs Time Step

The graph looks very extreme due to the larger number of bikes in the system. Since the return rate depends on the number of bikes in use, the return rate is much larger than the arrival rate. Besides, it is rather not logical to assume every bike is in the system. Some bikes might arrives at the station and leave and never come back while other bikes goes in and out of the station more frequently than others. The model might perform more practically if we add more stations to the model.

7 Code

If you visit `https://github.com/843098306/Senior-Project`, you will find programs where you can simulate various bike-sharing systems. There are two separate files. One is the function file, which includes the functions to simulate the bike-sharing system. The other file include the instruction on how to use these functions to simulate the system and how to plot various relevant quantities. Instructions are provided along with the files.

8 Conclusion and Future Work

The project simulates the bike-sharing system using continuous-time and discrete-space Markov Process. The model transforms from the the queuing theory model and birth and death model. Several quantities including stationary distribution and waiting time are studied using the model.

The focus is on the customer's point of view so far. By choosing appropriate parameters, we can minimize the total waiting time of the system. We can also analyze the system in the company's point of view. For example, we can add more parameters such as revenue and cost and see how the change of these parameters affect the profit brought by the system.

For the future work, more complexity can be added to the model. For example, the system can have more than 1 station. It might be more difficult to derive properties such as stationary distribution from the model but it is more interesting to study the behaviors of the system.

References

[1] S. Ink, "Bike share in the us: 2010-2016." [Online]. Available: https://nacto.org/bike-share-statistics-2016/