Social Security Office Queueing System

Course Project

OR7230: Probabilistic Operations Research

- Under Professor Xiaoning Jin

-By Prerit Samria (NU ID: 001380934)

Contents

Abstract

The United States Social Security Administration (SSA) is an independent agency of the U.S. federal government that administers Social Security, a social insurance program consisting of retirement, disability, and survivors benefits. These offices are visited daily by hundreds of visitors who wait for hours for service. The case study is based for the SSA Office at 10 Malcolm X Blvd, Roxbury, MA 02119, United States. To decrease the waiting time, SSA has introduced an online appointment system. Also, there are different types of queues depending on what you need help for. However, the average waiting time at any field office is still greater than 1 hour. The current queuing model being deployed at the SSA office is similar to k-M/M/1 models. The waiting time of visitors can be decreased by employing M/M/k model instead. The waiting time of queue in both models is estimated by running a simulation on Simulink Software.

Introduction

Problem Statement

The United States Social Security Administration (SSA) is an independent agency of the U.S. federal government that administers Social Security, a social insurance program consisting of retirement, disability, and survivors benefits. To qualify for most of these benefits, most workers pay Social Security taxes on their earnings. SSA is headquartered in Woodlawn, Maryland, just to the west of Baltimore, at what is known as Central Office. The agency includes 10 regional offices, 8 processing centers, approximately 1300 field offices, and 37 Tele-service Centers.

These offices are visited daily by hundreds of visitors who wait for hours for service. For every visitor, the waiting time is the total service time of all the visitors before her. To decrease the waiting time, SSA has introduced an online appointment system. Also, there are different types of queues depending on what you need help for. Still, the average waiting time at any field office is still greater than 1 hour. The increase in queue length frustrates both the employees and the visitors waiting for service. The service time for each visitor is generally the same. Thus, to decrease the waiting time, the current queueing model should be replaced with another more efficient model. This would not only decrease the waiting time of each customer but also increase the satisfaction of the visitors.

Objective

The Probabilistic Operations Research Method that can be used to solve the problem is a Queueing Model and simulation of such an SSA office. The current queuing model being deployed at SSA office is similar to k-M/M/1 models. The goal of the project is to decrease the waiting time of visitors by employing a different queuing model than the one currently being deployed at SSA office. The case study is based for the SSA Office at 10 Malcolm X Blvd, Roxbury, MA 02119, United States.

Data Sources

Data Collection

The data for the number of people who visit a Social Security Administration field office for the Fiscal Year 2016 (i.e., from 1st October 2015 to 30th September 2016) is available on the SSA website. The data for the Boston Field Office at 10 Malcolm X Blvd, Roxbury, MA 02119, United States is used for this project. The total number of visitors that visited this office in fiscal year 2016 is 19169. The total number of working days excluding the holidays during this period was 256. Thus the average number of visitors per day is approximately 75.

Table-1

Total number of visitors in 1 fiscal year	19169
Total number of working days excluding holidays	256
Average number of visitors per day	74.88

The working hours are from 9 am to 4 pm from Monday to Friday. Average number of visitors per day is 75 (see Table-1). On any working day, this office is busiest during 10 am to 2 pm. On average, 50 visitors arrive during this duration and wait in the system for about 60-75 minutes before getting out. This time duration is picked for observation since the rate of arrivals does not change much. Generally speaking, the more arrivals we include, the better. Imagine someone giving you a trick coin and asking you to determine the frequency with which it lands tails. How many times would you want to flip it before giving them a definitive answer? The obvious answer is that you would want to flip the coin a large number of times to get a good estimate of the probability of landing a tail.

Arrival Distribution

The arrival data has been acquired for March 10th, 2020 from 10 am to 2 pm. Please see table-2 for the arrival statistics during these 4 hours.

Observation No.	Time	Number of visitors arriving
1	1000 to 1005	0
2	1005 to 1010	1
3	1010 to 1015	0
4	1015 to 1020	0
5	1020 to 1025	0
6	1025 to 1030	1
7	1030 to 1035	0
8	1035 to 1040	1
9	1040 to 1045	0
10	1045 to 1050	1
11	1050 to 1055	1
12	1055 to 1100	0
13	1100 to 1105	2
14	1105 to 1110	1
15	1110 to 1115	1
16	1115 to 1120	2
17	1120 to 1125	1
18	1125 to 1130	2
19	1130 to 1135	0
20	1135 to 1140	2
21	1140 to 1145	2
22	1145 to 1150	1
23	1150 to 1155	1
24	1155 to 1200	0
25	1200 to 1205	3
26	1205 to 1210	1
27	1210 to 1215	0
28	1215 to 1220	2
29	1220 to 1225	1
30	1225 to 1230	2
31	1230 to 1235	1
32	1235 to 1240	3
33	1240 to 1245	0
34	1245 to 1250	2
35	1250 to 1255	1
36	1255 to 1300	2
37	1300 to 1305	1
38	1305 to 1310	2
39	1310 to 1315	2
40	1315 to 1320	1
41	1320 to 1325	2
42	1325 to 1330	1
43	1330 to 1335	2
44	1335 to 1340	2
45	1340 to 1345	1
46	1345 to 1350	2
47	1350 to 1355	1
48	1355 to 1400	0
	Total	55

Table-2: Frequency of arrivals from 10 am to 2 pm

Number of customers arriving in 5 mins	Frequency
0	12
1	14
2	9
>=3	2

Table-3: Frequency for each count of customers

To estimate the arrival distribution, Chi-Square Goodness of Fit Test is done *(see appendix - A)*. The arrival process is a poisson process with rate (λ_a) as 12.324 visitors per hour, or 0.2054 visitors per minute.

Service Distribution

The service time is assumed to be following a uniform distribution (15, 25) since the minimum and maximum service times observed were 15 mins and 25 mins respectively. The mean service time is 20 minutes.

Methodology

Model

The SSA office administers Social Security, which is a social insurance program consisting of retirement, disability, and survivors benefits. So there are different types of queues depending on what you need help for. The current queueing model that is being used is similar to k parallel M/G/1 models. In the SSA office under consideration, we are assuming that there are 5 servers (k = 5) servicing the needs of the visitors. This implies that there are 5 different type of queues. The queues that are used in this project are for the following coverages:

1. Old-Age, Survivors, and Disability Insurance (OASDI) - *Type 1 visitor*
2. Health Insurance and Health Services - *Type 2 visitor*
3. Assistance Programs - *Type 3 visitor*
4. Government Employee Retirement Systems - *Type 4 visitor*
5. Applying for New Social Security Number - *Type 5 visitor*

The overall arrival rate (λ_a) is 12.324 per hour, or 0.2054 per minute. This is then divided into 5 queues for the 5 different types. Each visitor is classified as type-i with probability P_i independently, for i = 1, 2, 3, 4, 5. For the sake of simplicity, we are assuming that each type-i visitor is equally likely to visit the office. Thus, each type-i visitor is equiprobable, i.e., $P_i = 0.2, \forall i$. Since the arrival process is a poisson process with rate λ_a, each type-i is also a poisson process with rate ($\lambda_a \times P_i$) for i = 1, 2, 3, 4, 5 (see Table-4).

Table-4: Probability of each Type-i visitor

Type of visitor	Number of visitors	Probability (P_i)	$\lambda_i = \lambda_a \times P_i$
Type-1	11	0.2000	0.0411
Type-2	11	0.2000	0.0411
Type-3	11	0.2000	0.0411
Type-4	11	0.2000	0.0411
Type-5	11	0.2000	0.0411
Total	55	1.0000	

In the office, there is no limit to the queue. To make sure that the queue does not explode, we will thus keep the waiting capacity to 60. Since we do not know the total number of visitors for each type-i, we will set the waiting capacity to 60 for all 5 queues. Thus, the queueing model that resembles the SSA office is *5 identical parallel M/G/1/60 queues*.

The performance of the current queueing model is compared with that of the proposed model, i.e., *M/G/5/60 model*. The proposed model has only 1 queue and the 5 servers are working in parallel. The arrival process follows a poisson distribution with rate λ_a, which is 0.2054 per minute. The service distribution is assumed to be uniform (15, 25) minutes.

The quantity of interest is the average amount of time a visitors spends in the queue (W_Q). This value is plotted for both models using simulation in Simulink software.

Discussion

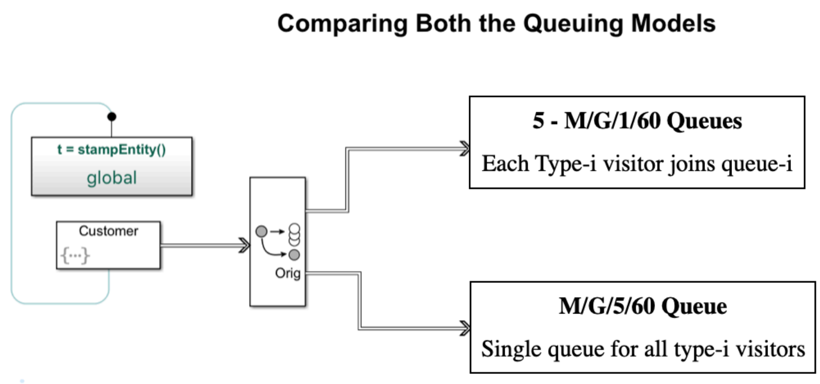


Figure-1: The arrival process is shown together for both the models.

The generation time and the exit time of entities (i.e., visitors) is recorded so that we can compute the waiting time. Each visitor is cloned after generation so that the two different line configurations can be run identically, i.e., so that we get same visitor entering the two different models at the same time. Figure-2 and figure-3 represents the current queueing model (i.e., 5 parallel - M/G/1/60 queues) and the proposed new model (i.e., M/G/5/60 model) respectively.

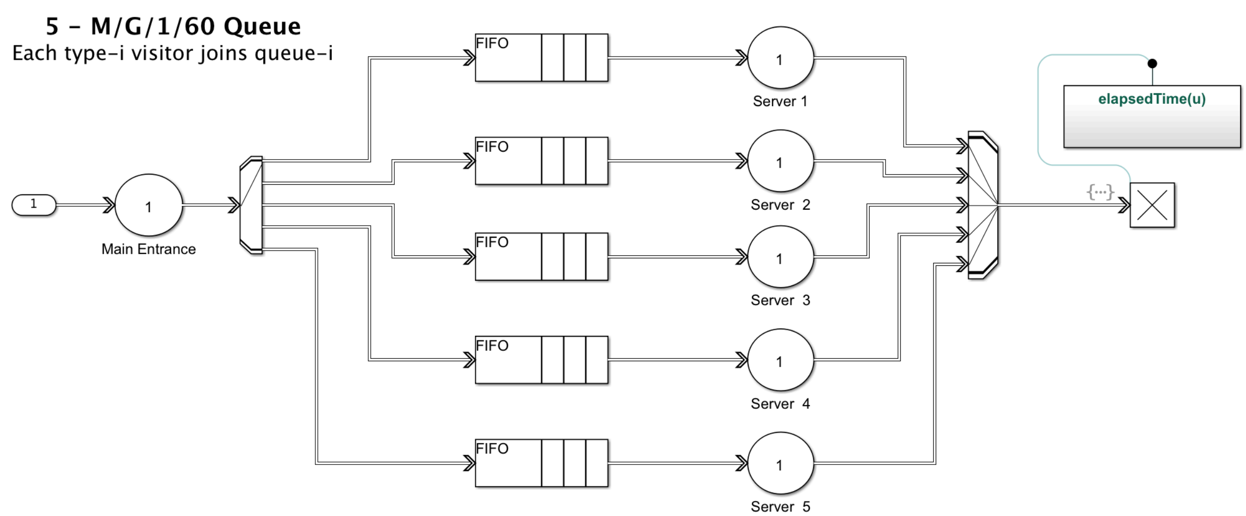


Figure-2: Queuing model setup for current model (5-M/G/1/60 Queues)

In figure-2, the 'Entity Output Switch' block after the 'main entrance' sends visitors randomly based on the probability distribution (P_i) to queue-i. Each server-i has service time that is uniformly distributed between 15 and 25 minutes, i = 1, 2, 3, 4, 5. Since, each type-i visitor is equiprobable, this model can be treated as 5 identical parallel M/G/1/60 queues. Thus, the total waiting time in queue for the system is equal to the waiting time in any queue-i, $\forall i = 1, 2, 3, 4,$ 5.

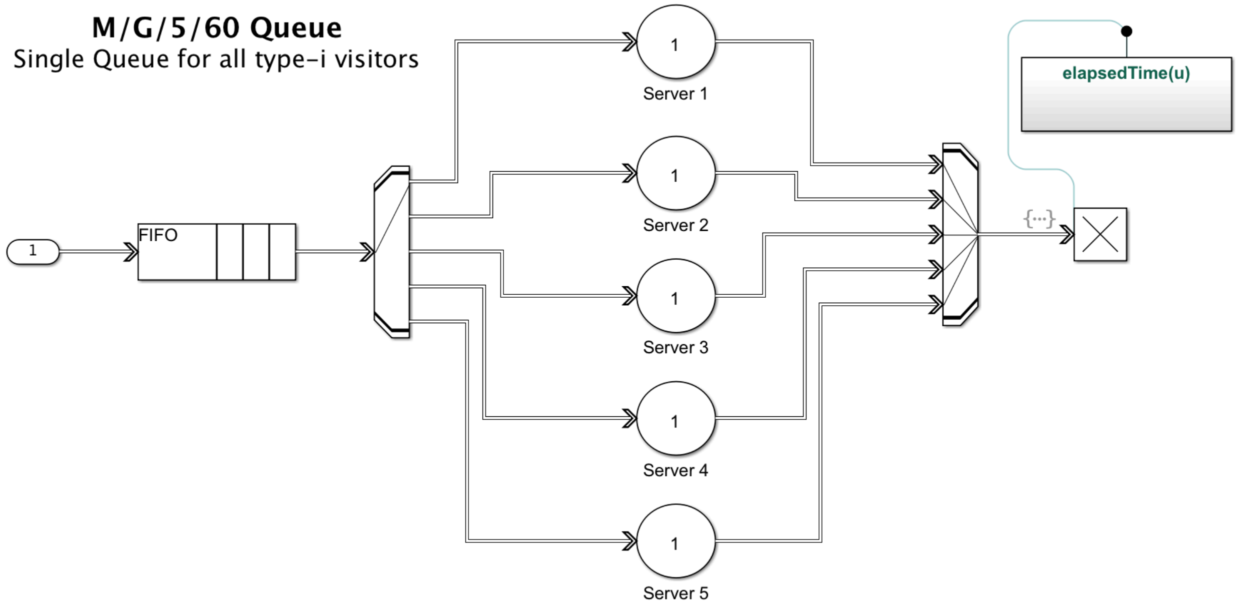


Figure-3: Queuing model setup for M/G/5/60 Queue

In figure-3, the 'Next in Line Selector' block after the FIFO queue sends visitors to the queue that is empty.

The simulation is run for 100000 time units. The graph for average waiting time in queue (W_q) is plotted for both the models. See figure-4 and figure-5.

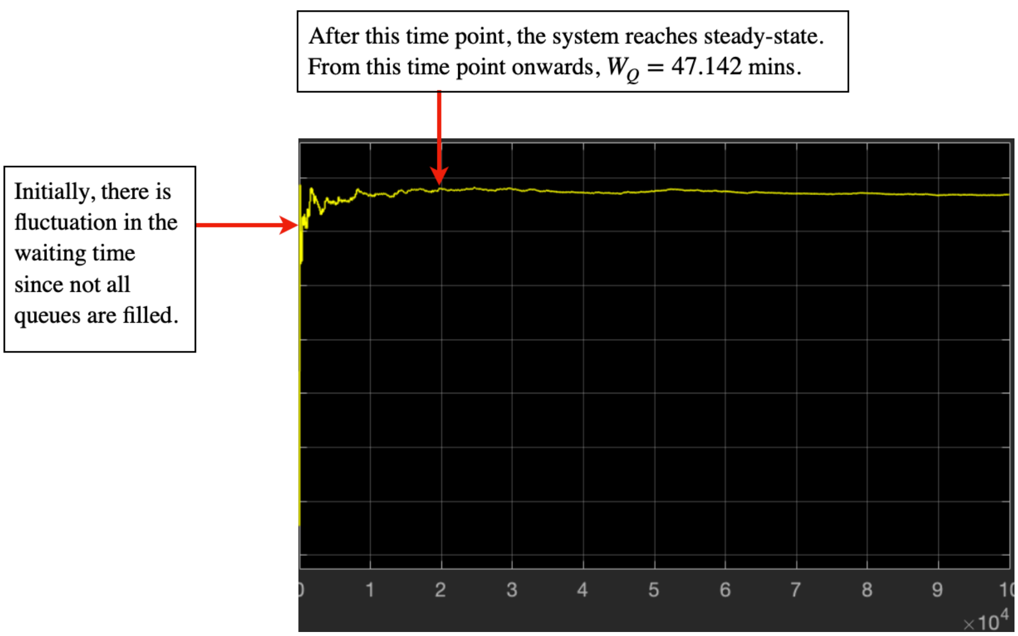


Figure-4: Plot of W_Q vs time for current model (5 parallel - M/G/1/60 queues)

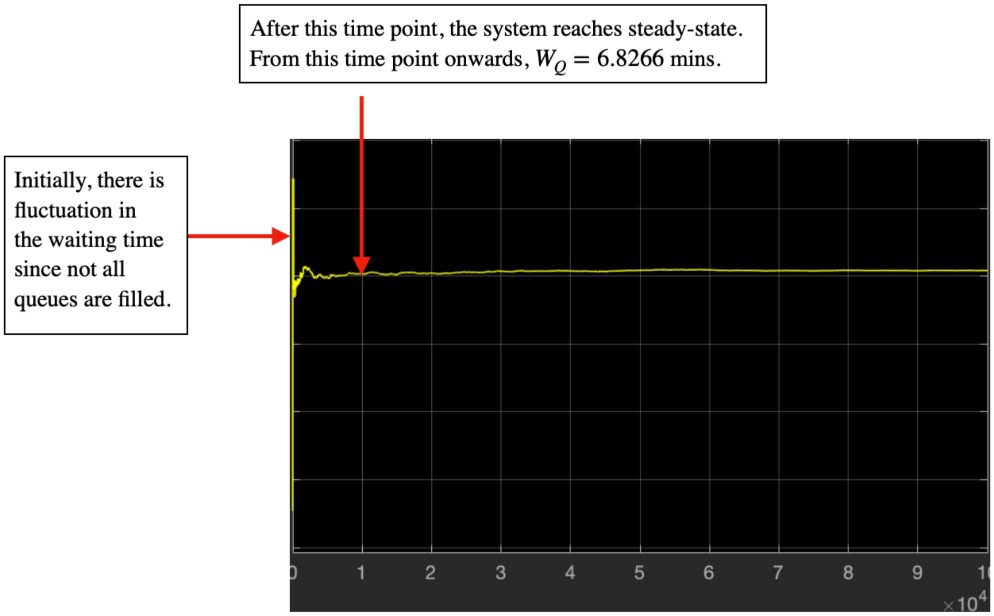


Figure-5: Plot of W_Q vs time for the proposed model (M/G/5/60 model)

From figure-4 and figure-5, we observe that the average waiting time in queue for current model is much more than that for the proposed model. The current model reaches steady-state in 2000 time units, and the proposed model reaches steady-state in 1000 time units.

Table-5: Comparison of average waiting time in queue (W_Q) and system (W_S) for both models

Model	W_Q **(in minutes)**	W_S **(in minutes)**
Current Model	47.1420	67.1420
Proposed Model	6.8266	26.8266

As we can see from table-5, for the current model, W_S is in the range of the initial guess in which we hypothesized that on average, the visitor spends 60 to 75 mins in the system. W_Q decreases considerably from 47.142 mins in the current model to 6.8266 mins in the proposed model. Similarly, W_S also decreases considerably from 67.142 mins in the current model to 26.8266 mins in the proposed model *(see appendix-B for calculation of W_Q and W_S)*.

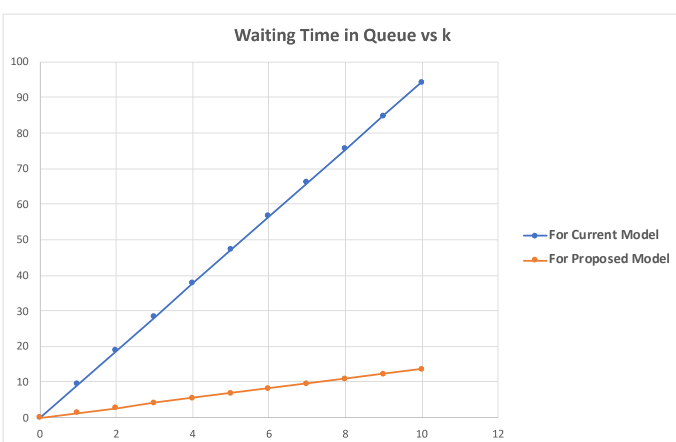


Figure-6: Plot of W_Q vs k for both models

From figure-6, we observe that as the number of different types of visitors (or equivalently number of servers for those types) increases, the difference in waiting times in queue between the models keeps on increasing. The waiting times in queue increases linearly with the number of types of visitors for both the models. Also, the increase in W_Q for current model is much more than that of the proposed model.

Conclusion

The average waiting time a visitor spends in the queue, W_Q decreases considerably from 47.142 mins in the current model to 6.8266 mins in the proposed model. Similarly, the average waiting time a visitor spends in the system, W_S also decreases considerably from 67.142 mins in the current model to 26.8266 mins in the proposed model. Thus, the waiting time can be decreased substantially if the proposed model is used. Also, the customer satisfaction would increase due to the decrease in waiting time. For the proposed model to be implemented, all the employees would have go under training to gain expertise in dealing with all types of visitors to the SSA office. The benefits of this proposed model far outweigh the training costs in the sense that the SSA office will save more money by serving the visitors faster and improving their overall experience with the SSA.

Appendix - A

Number of customers arriving in 5 mins, xi	Observed Frequency, Oi
0	12
1	14
2	9
>=3	2
Total	37

Table-5: Frequency for each count of customers

$f(x_i, \lambda) = \mathrm{e}^{-\lambda}\dfrac{\lambda^{x_i}}{x_i!}$

$x_1, x_2, \ldots, x_{12}, x_{13}, \ldots, x_{26}, x_{27}, \ldots, x_{35}, x_{36}, x_{37}$

Values of $x_i = 0$, for i = 1, 2, …, 12

Values of $x_i = 1$, for i = 12, …, 26

Values of $x_i = 2$, for i = 27, …, 35

Values of $x_i = 3$, for i = 36, 37

Using Maximum Likelihood Estimator to estimate the arrival rate λ_a.

$$L(\lambda) = \prod_{i=1}^{37} f(x_i, \lambda) = (\mathrm{e}^{-\lambda}\frac{\lambda^0}{0!})^{12} \times (\mathrm{e}^{-\lambda}\frac{\lambda^1}{1!})^{14} \times (\mathrm{e}^{-\lambda}\frac{\lambda^2}{2!})^{9} \times (\mathrm{e}^{-\lambda}\frac{\lambda^3}{3!})^{2}$$

$$\Rightarrow L(\lambda) = \mathrm{e}^{-37\lambda}\frac{\lambda^{38}}{2^9 \times 6^2}$$

$$\Rightarrow L(\lambda) = \mathrm{e}^{-37\lambda}\frac{\lambda^{38}}{a}, \text{ where } a = 2^9 \times 6^2$$

Taking log on both sides,

$$\Rightarrow l(\lambda) = log(L(\lambda)) = -37\lambda + 38log(\lambda) - log(a)$$

Differentiating both sides with respect to λ and equating it to 0,

$$\frac{dl}{d\lambda} = -37 + \frac{38}{\hat{\lambda}} = 0$$

$$\Rightarrow \hat{\lambda} = \frac{38}{37} = 1.027 \text{ visitors per 5 mins}$$

Chi-Square Goodness of Fit Test (χ^2- GOF Test)

H_0: Data fits Poisson Distribution with $\lambda = 1.027$ (Claim)

H_1: Data does not fit Poisson Distribution with $\lambda = 1.027$ (Claim)

$\alpha = 0.01$ (i.e., 99% confidence), n = 37

$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, with $\lambda = 1.027$, for x = 0, 1, …

Thus, $P_0 = f(0) = e^{-\lambda}\frac{\lambda^0}{0!} = 0.3581$, $P_1 = f(1) = e^{-\lambda}\frac{\lambda^1}{1!} = 0.3678$,

$P_2 = f(2) = e^{-\lambda}\frac{\lambda^2}{2!} = 0.1888$, $P_3 = f(3) = e^{-\lambda}\frac{\lambda^3}{3!} = 0.0646$

S. No.	x_i	Observed, O_i	Expected, $E_i = nP_i$
1	0	12	13.25
2	1	14	13.61
3	2	9	6.99
4	≥ 3	2	3.15
	Total	37	37

Since $E_4 < 5$, combine 3rd and 4th categories.

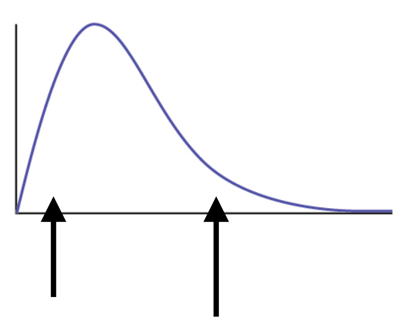
S. No.	x_i	Observed, O_i	Expected, $E_i = nP_i$
1	0	12	13.25
2	1	14	13.61
3	≥ 2	11	10.14
	Total	37	37

Using χ^2- GOF Test, number of categories, k = 3 and number of parameters, m = 1.

Thus, degree of freedom, df = k - m - 1 = 1

$$\chi_c^2 = \sum_{i=1}^{3}\frac{(O_i - E_i)^2}{E_i} = 0.1179 + 0.0112 + 0.0729 \Rightarrow \chi_c^2 = 0.202$$

$\chi^2_{\alpha,df} = \chi^2_{0.01,1} = 6.635$



χ_c^2 does not lie in the critical region

$\Rightarrow$ fail to reject H_0

$\chi_c^2 = 0.202$ $\chi^2_{\alpha,df} = 6.635$

Thus, we have enough evidence to support the claim that the data fits Poisson Distribution with $\lambda = 1.027$ visitors per 5 mins

$\Rightarrow \lambda = 0.2054$ visitors per min

Appendix-B

For Current Model: 5 parallel - M/G/1/60 queues

Service time, $S \sim U(15{,}25)$

E[S] = 20 mins, $\lambda_i = 0.0411$ per min

Thus, $\lambda_i \times E[S] = 0.822 < 1 \Rightarrow$ Stable system

So, we can assume that the capacity is infinity.

Thus, the average waiting time a visitor spends in the queue, $W_Q = W_{Q_i}$

$$\Rightarrow W_Q = \frac{\lambda_1 E[S^2]}{2(1-\lambda_1 E[S])}\text{ , where } E[S^2] = var(S) + (E[S])^2$$

$$\Rightarrow W_Q = \frac{16.7825}{0.356} = 47.142 \text{ mins}$$

So, the average waiting time a visitor spends in the system, $W_S = W_Q + E[S]$

$$\Rightarrow W_S = 47.142 + 20 = 67.142 \text{ mins}$$

This value is value in the range of the initial guess in which we hypothesized that on average, the visitor spends 60 to 75 mins in the system.

For Proposed Model: M/G/5/60 model

Service time, $S \sim U(15{,}25)$

E[S] = 20 mins, $\lambda = 0.2054$ per min

So, $var(S) = 8.3333$, and $E[S^2] = var(S) + (E[S])^2 = 408.3333$

The formula to calculate W_Q for a M/G/k system is approximated as:

$$W_Q = \frac{\lambda^k E[S^2](E[S])^{k-1}}{2(k-1)!(k-\lambda E[S])^2(\sum_{n=0}^{k-1}\frac{(\lambda E[S])^n}{n!} + \frac{(\lambda E[S])^k}{(k-1)!(k-\lambda E[S])})}$$

Here, k = 5

$$\text{Thus, } W_Q = \frac{23885.8472}{(38.192)(36.9662 + 54.6482)} = 6.8266 \text{ mins}$$

So, the average waiting time a visitor spends in the system, $W_S = W_Q + E[S]$

$$\Rightarrow W_S = 6.8266 + 20 = 26.8266 \text{ mins}$$