Simulation-Based Construction of an M/M/1 Queue Model

Mustafa Abduljabbar
College of medicine, University of Wasit, 52001 Wasit, Iraq
mustafaajaddah@uowasit.edu.iq

Keywords: Poisson Distribution, Exponential Distribution, Birth and Death Process, Simulation of the M/ M/1 Model.

Abstract: The research paper presents a theoretical framework for constructing $(M\ M/1){:}\,(\infty/\infty/\ FCFS)$ queue models, focusing on derive The Poisson distribution with parameter ($\boldsymbol{\lambda t}$), discussed relationship between Poisson distribution and the exponential distribution, the Memoryless property of the Exponential distribution, the Assumptions of the M/ M/ 1 model, the construction of the M/ M/1 model, And which included transition states diagram for a Markov chain(diagram of birth and death process of M/M/1 model) containing a proposal to add the time interval (h), The purpose of adding (h) is to maintain the accuracy of information continuity. This procedure is supposed to apply to research papers that include Markov state transition diagrams [1], [2], [3], [4], [5] and This contributed to understanding the construction of the models. The research paper included an experimental aspect (Simulation), where the simulation results were consistent with the result of the theoretical model .The model was also discussed in terms of practical application and reference was made to a number of research papers that included the application of the M/M/1 model the aim of studying the M/M/1 model in detail that it is considered the basis for understanding more advanced models such as M/M/C, M/G/1 and G/M/1.

1 INTRODUCTION

Queuing theory is one of the methods of operations research that specializes in dealing with the problem of crowding in service centers, its use in optimizing patient flow in healthcare systems, managing customer queues in retail, or analyzing network traffic in telecommunications. The importance of the queuing theory lies in controlling the problem of long waiting times in queues and the costs that result from them. In these papers, we will focus on the simplest types of queuing models M/M/1, which represents the basis for understanding more advanced queuing models.

2 THE THEORETICAL ASPECT

In this section, the question of how the M/ M/1 model was built was answered. This is done by addressing a group of topics accurately and in sequence, starting with the Poisson distribution and ending with building the model.

2.1 Poisson Distribution

The Known as the distribution of rare events, it is a distribution derived from a binomial distribution that can be represented by the case of a coin being tossed an infinite number of times, and is considered a common distribution in queue theory.

To derive the Poisson distribution for a period of time of length (t) depends on modeling the number of rare events that occur during this period (t). According to the main steps of derivation [6], [7].

1) Poisson distribution assumptions:
   - Independence. Events within the time period are independent of each other.
   - Stationary. The probability of a given number of units occurring depends only on the length of the time period, not on their time location.
   - Orderliness (Rare). The probability of more than one unit occurring in a short period of time (h is very small and (h) is close to zero).
2) Mathematical Steps:
   - Dividing the period (t) into (n) of the small period (h):

$$h = \frac{t}{n}. \quad (1)$$

- Probability of one event occurring in the period (h):

$$\theta = \lambda h = \frac{\lambda t}{n}. \quad (2)$$

- Binomial Distribution Model :

$$p(k) = \binom{n}{k}\left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k}. \quad (3)$$

- When (n→ ∞):

$$\binom{n}{x} \approx \frac{n^x}{k!}, \left(1 - \frac{\lambda t}{n}\right)^{n-x} \approx e^{-\lambda t},$$

$$P(k) = \frac{(\lambda t)^k \; e^{-\lambda t}}{k!} \qquad k=0,1,2,\ \dots \quad (4)$$

Equation (4) represents the Poisson distribution where:

- k: The number of units within the time period (t);
- λ: Average number of units per unit of time;
- t: Studied time period .

2.2 The Relationship Between Poisson Distribution and Exponential Distribution

Suppose that (T) is a random variable that represents the arrival time of a number of units, the probability distribution function (C. D. F.) is (F(t)), and the probability density function (p. d. f.) is (f(t)) where the cumulative distribution function is as follows [8]:

$$F(t) = P(T \leq t), \quad (5)$$

$$F(t) = 1 - P(T > t). \quad (6)$$

If $P(T \leq t)$ represents the appearance of at least one unit within the period (0, t), this probability can be obtained from (4), i.e.:

$$P(T \leq t) = 1 - e^{-\lambda t}. \quad (7)$$

$P(T > t)$ represents the probability that there will be no unit within the period (0, t), i.e.:

$$P(T > t) = e^{-\lambda t}. \quad (8)$$

2.3 Memoryless Property & the Assumption of the M/ M/ 1 Model

One of the important characteristics of exponential distribution is the property of Memoryless. It means that the probability of a future event occurring at the time (t+h) does not depend on the current time (t) of the occurrence of the event, but depends on the time (h), [9] i.e.:

$$p_r\ (T \geq t + h \mid T \geq t) = P(T > h) = e^{-\lambda h}. \quad (9)$$

$$So,\ p_r\ (T \leq t + h \mid T \geq t) = P(T \leq h) = 1 - e^{-\lambda h}. (10)$$

To clarify, If the number of units (n) in the queue at the present time (t) and in order to get the probability of an additional unit coming after one step of time (h), i.e. at time (t+h), it is formulated as follows:

$$p_{n+1}(t + h) = p_n(t) * (1 - e^{-\lambda h}). \quad (11)$$

The property of Memoryless represents the property of Markov.accordingly, the Assumption of the M/M/1 model was formulated as follows:

1) The time interval for the arrival of the units follows the exponential distribution (λ).
2) The time interval of the serviced units follows the exponential distribution (μ).
3) Arrival times are independent of service times.

2.3 Building a Model M/ M/ 1

Assuming a steady-state system, the M/M/1 model is constructed as follows [8], [10]:

1) We find the probability of one unit arrival occurring at the h-period :

From (5), let h present t → 0:

$$\begin{aligned} \text{P (one arrival in h)} &= \text{P}(\text{T} \leq \text{h}) = 1 - e^{-\lambda h} \\ &= 1 - \left(1 - \frac{\lambda h}{1!} + \frac{(\lambda h)^2}{2!} - \frac{(\lambda h)^3}{3!} + \cdots\right) \\ &= \lambda h - \frac{(\lambda h)^2}{2!} + \frac{(\lambda h)^3}{3!} - \cdots \\ &= \lambda h + \Delta(h). \end{aligned} \quad (12)$$

Where: $\Delta(h) = -\frac{(\lambda h)^2}{2!} + \frac{(\lambda h)^3}{3!} - \cdots,$

$\Delta(h)$: is very small, So $\Delta(h) \approx 0$.

P (one arrival in h) = $1 - e^{-\lambda h} \approx \lambda h$; for small.

2) With the same technique, we find the probability of one unit completing its service at h-period, So:

$$\text{P (one service completion in h)} = \text{P}(\text{T} \leq \text{h}) \approx \mu h;\ \text{for small h.} \quad (13)$$

3) According to the property of Memoryless, the process of units arrival and the process of units departure can be formulated in the system in the form of a Markov chain of the transitional states diagram called the birth and death process as in Figure 1.
4) The system maintains its stable state (balanced) when the probability of one unit arrival of a certain state is equal to the probability of one unit leaving the same state. The letter (h) in the diagram, which represents the interval (t→ 0)in which the transfer or non-transfer of one unit

occurs, was proposed by the researcher, From Figure 1 we derive the following equations:

5) Probability arrival of state (n):

$$p_n(t+h) = p_{n-1}(t) * \lambda h + p_{n+1}(t) * \mu h; \quad n \geq 0. \quad (14)$$

Where: $p_{-1}(t) = 0$;

6) Probability departure of state (n)

$$p_{n+(1\ or-1)}(t+h) = p_n(t) * \mu h + p_n(t) * \lambda h; \quad n \geq 0. \quad (15)$$

7) In order for the system to be balanced, the following must be satisfying:

$$p_n(t+h) = p_{n+(1\ or-1)}(t+h). \quad (16)$$

By substituting (14) and (15) into (16). After simplification and we obtain.

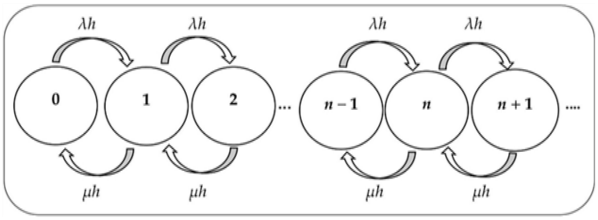

Figure 1: Diagram of birth and death process of (M/M/1): $(\infty, \infty, \text{FCFS})$ model.

$$p_{n-1}(t) * \lambda h + p_{n+1}(t) * \mu h = p_n(t) * h\,(\mu + \lambda) \quad (17)$$

And by substituting the cases $\{0,1,2,\dots,n,n+1\}$ in to (17) to obtain:

$$\therefore p_1(t) = \frac{\lambda}{\mu} p_0,\ p_2(t) = \frac{\lambda^2}{\mu^2} p_0,\ p_3(t) = \frac{\lambda^3}{\mu^3} p_0,$$

$$\therefore p_n(t) = \frac{\lambda^n}{\mu^n} p_0 \qquad n = 1,2,3,\dots \quad (18)$$

And since the number of units is independent of the time period t, therefore ($p_n(t) = p_n$)

$$p_n = p_0 (\frac{\lambda}{\mu})^n;\ n = 1,2,3,\dots \quad (19)$$

p_0: denotes the probability of the system being empty (with zero units).

$$\because \sum_{n=1}^{\infty} p_n = 1,$$

$$\therefore p_0 \sum_{n=1}^{\infty} (\frac{\lambda}{\mu})^n = 1,$$

$$p_0 = \frac{1}{\sum_{n=1}^{\infty} (\frac{\lambda}{\mu})^n} = \frac{1}{1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 + \left(\frac{\lambda}{\mu}\right)^3 + \cdots}.$$

The denominator is a geometric series whose ratio is $(\frac{\lambda}{\mu})$.

$$\therefore p_0 = 1 - \frac{\lambda}{\mu} \qquad ;\lambda < \mu. \quad (20)$$

By substituting (20) in to (19):

$$p_n = (\frac{\lambda}{\mu})^n (1 - \frac{\lambda}{\mu}) \quad ;\ n = 0,1,2,3,\dots \quad (21)$$

Equation (21) represents the probabilistic model $((\text{M/ M/1}): (\infty, \infty, \text{FCFS}))$.

Using (21), the performance efficiency measures listed in Table 1 are computed, which are applicable to many systems.

Table 1: Represents important performance measures.

No.	Performance Measure	Mathematical Expression
1	The average waiting time in the system (W_s)	$W_S = \frac{1}{(\mu-\lambda)}$
2	The average waiting time in the queue (W_q)	$W_q = \frac{\lambda}{\mu(\mu-\lambda)}$
3	The average number of customers in the system (L_s)	$L_s = \lambda\, W_S$
4	The average number of customers in the queue (L_q)	$L_q = \lambda\, W_q$
5	The utilization factor (ρ)	$\rho = \frac{\lambda}{\mu}$

3 THE EXPERIMENTAL ASPECT

The simulation aims to validate the theoretical M/M/1 model by comparing its performance measures with those of the experimental M/M/1 model. The simulation model was implemented in MATLAB R2021b.

The program writing was included in Appendix [11] – [12].

3.1 Methodological Steps for Building the Simulation Model

The steps to write the simulation are as follows :

1) Entering the program parameter values: lambda, mu, warm-up period, number of replicates, total customers.
2) Generating exponentially distributed data using the exprnd command with parameter (1/lambda) representing interarrival times.
3) We compute the cumulative sum of interarrival times using the cumsum command to obtain arrival times.
4) Generating service times: Exponentially distributed data is generated using the expend command with parameter (1/mu).
5) Calculating departure times:
   - The departure time of the first arrival = Arrival time of the first arrival + Service time of the first arrival.

- Departure times of subsequent arrivals = Maximum(the current arrival's time, the previous departure time) + the current service time.

6) Calculating system waiting times= departure times − arrival times.
7) Calculating queue waiting times = system waiting times–service times.
8) Excluding warm-up period data from both queue waiting times and system waiting times:
   - Initial observations are discarded (warm-up period).
   - Without warm-up period, you would get biased metrics (e.g., unrealistically low waiting times).
9) Metrics calculation:
   - averages for waiting times in system;
   - averages for waiting times time in queue;
   - averages for number of customers in system (lambda * waiting times in system);
   - averages for number of customers in queue (lambda * waiting times in queue).

10) Simulation replication. To obtain more reliable results than relying on a single simulation run.
11) Calculating the average metrics. Representing the experimental performance measures.
12) Results display. Comparison of the experimental performance measures with the theoretical performance measures.

Table 2: Result of simulation.

λ =6, μ =9, ρ = 0.78		
Metrics	Theoretical model	Simulation
W_q	0.5	0.5
W_s	0.39	0.39
L_s	3.50	3.50
L_q	2.72	2.72
λ = 0.8, μ=1, ρ =0.80		
Metrics	Theoretical model	Simulation
W_q	5	5
W_s	4	4
L_s	4	4
L_q	3.20	3.20
λ =3, μ=7, ρ = 0.43		
Metrics	Theoretical model	Simulation
W_q	0.25	0.25
W_s	0.11	0.11
L_s	0.75	0.75
L_q	0.32	0.32

3.2 Discussing the Simulation Results

The simulation outputs are presented in Table 2.

Table 2 presents the results, showing general agreement between the theoretical model M/M/1 and simulation performance measures. Thus, the theoretical M/M/1 model was validated and can be reliably.

4 APLICATION

To apply the M/M/1 model to real data, It requires two sets of data:

1) Arrival time data. It can be represented in one of two cases:
   - The arrival intervals of units follow the exponential distribution (based on the assumptions of the M/M/1 model).
   - The number of units received follows the Poisson distribution (based on the pure birth model [10]), and this case has been adopted in the applied aspect of the research paper.
2) Service time data. Represents service intervals (based on M/M/1 model assumptions).

In practice, the data are in the form of classified data (frequency table) and the steps of practical application are as follows:

1) Finding the estimation of the parameters λ, μ.
2) Adopting the chi-square test for goodness of fit[1] to verified whether the data arrival times and service times track the exponential distribution or not.
3) Verify the condition λ<μ.
4) Calculation of performance measures.

The studies that included the application of a model are the application of what has been mentioned [13], [14].

5 CONCLUSIONS

Based on the simulation methodology and results:

1) The results of the theoretical model's performance measures generally match those calculated through simulation, indicating that the model represents a queuing system with a single service center.
2) This is achieved when the following conditions are met:
   - Interarrival times follow an exponential distribution with parameter (λ), or its

[1]Five goodness of fit tests, https://www.cimt.org.uk/projects/mepres/alevel/fstats_ch5.pdf.

equivalent, the arrival of units at time (t) follows the Poisson distribution with the parameter (λt).

- Service times follow an exponential distribution with parameter (μ).
- The sample size should be large and there should be a warm-up period represented by excluding a part of the initial data (delete first 10% of data system waiting times or queue waiting times) to treatment the bias resulting from the initial data.

3) The difficulty in implementing it is due to the strictness of its conditions (The difficulty in implementing it is due to the strictness of its conditions (the difficulty in obtaining data that meets its conditions).

ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Research,Transfer and Start-Up Center (Forschungs, Transferund Grunderzentrum), the Anhalt University of Applied Sciences and the state of Saxony-Anhalt for our research.

REFERENCES

[1] F. Hirbod, M. Eshghali, M. Sheikhasadi, F. Jolai, and A. Aghsami, “A state-dependent M/M/1 queueing location-allocation model for vaccine distribution using metaheuristic algorithms,” Journal of Computational Design and Engineering, vol. 10, no. 4, pp. 1507–1530, 2023.

[2] V. K. Vijayashree and K. Ambika, “An M/M/1 queue subject to differentiated vacation with partial interruption and customer impatience,” Quality Technology & Quantitative Management, vol. 18, no. 6, pp. 657–682, 2021.

[3] S. Ammar, “Behavior analysis of an M/M/1 vacation queue in random environment,” Quality Technology & Quantitative Management, vol. 18, no. 4, pp. 397–417, 2021.

[4] J. Wang, Y. Zhang, and Z. G. Zhang, “Strategic joining in an M/M/K queue with asynchronous and synchronous multiple vacations,” Journal of the Operational Research Society, vol. 72, no. 1, pp. 161–179, 2021.

[5] Y. Kerner and O. Shmuel-Bittner, “Strategic behavior and optimization in a hybrid M/M/1 queue with retrials,” Queueing Systems, vol. 96, no. 3–4, pp. 285–302, 2020.

[6] R. V. Hogg, E. A. Tanis, and D. L. Zimmerman, Probability and Statistical Inference, 9th ed. Boston, MA, USA: Pearson, 2013.

[7] J. F. Shortle, J. M. Thompson, D. Gross, and C. M. Harris, Fundamentals of Queueing Theory, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2018.

[8] L. Kleinrock, Queueing Systems, Volume 1: Theory. New York, NY, USA: John Wiley & Sons, 1975.

[9] G. Marsaglia and A. Tubilla, “A note on the ‘lack of memory’ property of the exponential distribution,” The Annals of Probability, vol. 3, no. 2, pp. 353–354, 1975.

[10] H. A. Taha, Operations Research: An Introduction, 8th ed. Hoboken, NJ, USA: Pearson Prentice Hall, 2007.

[11] O. V. Johnson and M. Ganiyu, “Simulation approach to life-data queue event modeling,” Computer Engineering and Intelligent Systems, vol. 8, 2019.

[12] X.-Z. Lv, B.-X. Fan, X.-W. Wang, and X.-H. Zhang, “A modeling and simulation method of queue systems based on Stateflow,” in Proc. 3rd Int. Conf. Systems and Informatics (ICSAI), 2016, pp. 541–545.

[13] A. O. Odior, “Application of queueing theory to petrol stations in Benin-City area of Edo State, Nigeria,” Nigerian Journal of Technology, 2013. [Online]. Available: https://doi.org/10.4314/njt.322.664

[14] J. O. Igbinoba, O. K. Sule, and M. Akinwunmi, “Application of queueing theory in optimization of service process: A case study of GT Plaza fast food,” Journal of Multidisciplinary Engineering Science and Technology, vol. 6, no. 1, pp. 9320–9327, 2019.

APPENDIX

This is snippet of the MATLAB code used for generating interarrival and service times[2].

```
% MATLAB code for M/M/1 simulation
clc
clear all
lambda = 6; % arrival rate
mu = 9; % service rate
num_customers = 1000; % total customers
warm_up = 100; % warm-up period
% Generate interarrival times (exponential distribution)
inter_arrivals = exprnd(1/lambda, 1, num_customers);
arrival_times = cumsum(inter_arrivals);
% Generate service times
service_times = exprnd(1/mu, 1, num_customers);
% Compute departure times
departure_times = zeros(1, num_customers);
departure_times(1) = arrival_times(1) + service_times(1);
for i = 2:num_customers
departure_times(i) = max(arrival_times(i),
```

[2]M/M/1 and M/M/K, 2025, https://www.mathworks.com/matlabcentral/fileexchange/66760-m-m-1-and-m-m-k.

```
departure_times(i-1)) +
service_times(i);
   end
   % Calculate waiting times
   system_waiting_times =
departure_times - arrival_times;
   queue_waiting_times =
system_waiting_times - service_times;
   % Exclude warm-up period
   system_waiting_times =
system_waiting_times(warm_up+1:end);
   queue_waiting_times =
queue_waiting_times(warm_up+1:end);
   % Compute metrics
   W_s = mean(system_waiting_times) %
average system waiting time
   W_q = mean(queue_waiting_times) %
average queue waiting time
```