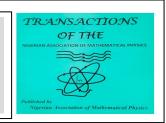


Transactions of The Nigerian Association of Mathematical Physics



Journal homepage: <u>https://nampjournals.org.ng</u>

MATHEMATICAL ANALYSIS OF PERFORMANCE MEASURES OF M/M/1 QUEUE MODELS.

¹S.A Ogumeyo and ²C.E.O. Omole

¹Department of Mathematics. Delta State University of Science and Technology, Ozoro Nigeria. ²Department of Mathematics, College of Education, Warri, Nigeria.

ARTICLE INFO

Article history: Received xxxxx Revised xxxxx Accepted xxxxx Available online xxxxx

Keywords: Queue, Customers, Arrival, Distributions, Waiting times.

ABSTRACT

The goal of this research is to expand existing single server queue models by deriving their performance measures through the application of probability laws. In this paper, we present a server queue model which consists of balking customers whose behaviours are characterised by discouragement and impatience during long queues. The averages number of such customers and response times of the queue system are being used to derive the average time customers have to wait and the number of customers who wait for service. Bayes law of total probability and Laplace-Stieltjes transforms are being applied to derive these parameters in single-server queue systems presented in this research. A numerical example is also presented to validate the model parameters. It is observed that, in order to evaluate the distribution of the response and waiting time, the distribution at the instant a customer joins it must be known. It is also observed that, the model distribution's parameters from both the theoretical and numerical illustrations of the single server (M/M/1) with balking customers presented in this paper conform to the Little's theorem on queues.

1.0 INTRODUCTION

The authors in [1], remark that the number of customers who patronize a service facility could be finite or infinite. According to Ogumeyo and Nwamara [2] and Wagner [3], customers' arrival and departure in queuing system follow Poisson distribution law with arrival rate λ and service rate μ which are assumed to be independent and exponentially distributed. The busy periods are periods during which the server is busy servicing customers while idle periods are periods during which the server is not servicing customers as defined in [3]. The author in [4], stated that the role of a queue system involves identifying the probability distribution of customers' rate of arrival and departure including the service pattern. Basic queue structures and queue channels are discussed in Subagyo [5] and Kakiay [6]. Hillier and Lieberman [1] stated that queues emerge as a result of service rate being lower than the rate of demand for service, liken to situations commonly experienced in super – markets, hospitals, banks, filling stations etc. The three states of a queuing system is reported in [7], which includes transient, steady and explosive states. A queue is said to be in a transient state if its behavior varies with time, if its behavior does not change with time, it is said to be in steady state condition while an explosive state is when a queue system builds up to infinity.

*E-mail address:*simonogumeyo64@gmail.com

https://www.doi.org/10.60787/tnamp-19-125-134

^{*}Corresponding author: Ogumeyo S.A.

^{1118-4752 © 2024} TNAMP. All rights reserved

As contained in Ogumeyo and Nwamara [2], several queuing models in literature have addressed different aspects of queuing systems. For example, a queue model to decongest highways traffic is reported in Nugraha [8]. A queuing model to minimize patient waiting time in a healthcare facility is presented in Kembe*et al.* [9]. Commonly used mathematical symbols by Kendall classifying queues are stated in [1], [3] and [4].

Bhat [10], presented single server queue models which centred on utility of servers, including their busy and idle periods' probability distribution functions with the aim of ascertaining their levels of usefulness. Weber [12] and Takagi [13] remark that performance determines the steady-state of a queue which according to Kobayash [11] is attained when the expected number of arrivals is equal to the mean service time.

The behaviour of customers is very important in the management of a queue system. According to Weber [12], balked customers are customers who refused enter the queue because the queue is too lengthen while reneged customers are customers who join the queue but after sometime they get discouraged and exit it. Jockeying customers are the ones that move from one queue to another in search of a quicker or faster service when there are alternative queues in the system. A queue is said to be in transient state if its behaviour varies with time while a queue is said to be in a steady state if its behaviour in a long interval is independent of time. The measures of performance of a queue system hinges on the ability to manage its service facility with the aim of striking an optimum balance between the waiting cost of time and cost of keeping the system idle time, Ogumeyo and Nwamara [2]. Waiting cost can be direct or indirect. Direct waiting cost consists of unutilized manpower and equipment which has to be paid for by business owners. This increases the cost of production of goods and services. Indirect cost centres on loss of customer which leads to decreased sales and less profits. The purpose of this research is to expand the queue model in [2] with the aim of analysing its performance measures.

2.0 Methodology and Materials

Mathematical Notations

The following are the mathematical symbols associated with the proposed models:

 λ = Arrival rate of customers, $\frac{1}{\lambda}$ = average time the server is idle, S = customer time expenditure in the system. μ = rate of departure from the system, \overline{Q} = Average population of customers in the queue, \overline{T} = average time of response to a customer, \overline{N} = Average population of customers in the system.

Derivation of a Steady-State by using Baye's Rule

The methodology adopted in this research is Baye's rule of total probability which was also applied by the author in [12]. The steady state distribution for arrival and departure processes are obtained by using Baye's rule of total probability as stated in the following equations:

$$P_{i} = \frac{\lambda_{0} \dots \lambda_{i-1}}{\mu_{0} \dots \mu_{i}} = P_{0}, \quad i = 1, 2, \dots, \quad P_{0}^{-1} = 1 + \sum_{i=1}^{\infty} \frac{\lambda_{0} \dots \lambda_{i-1}}{\mu_{0} \dots \mu_{i}}....(1)$$

Let X_a, X_b represent the state of instant arrival and departure of customers from the system, respectively and let $II_k = P(X_a = k), D_k = P(X_b = k), k = 0, 1, 2, ...$ represent their distributions. If we apply Baye's rule to the above expressions,

$$II_{k} = \lim_{h \to 0} \frac{(\lambda_{k}h + 0(h))P_{k}}{\sum_{j=0}^{\infty} (\lambda_{j}h + 0(h))P_{j}} = \frac{\lambda_{k}P_{k}}{\sum_{j=0}^{\infty} \lambda_{j}P_{j}}.$$
(2)

Similarly

Equations (1)-(4) state that at steady-state in a queuing system, the average arrival rate and the rate of service are similar but cannot have the same value compared to steady – state distribution at a random point. Equation (5) clearly proves this.

If the arrivals of customers are uniformly distributed over a multi-server m system, then the average population of customers who arrive at the service facility during time T is $\lambda T/m$. Hence the utilization of the server can be expressed as stated in equation (6).

 $U_{\rm s} = \frac{\lambda}{m\mu}.....(6)$

If we represent U and T as the times of waiting and response by the *ith* customer respectively by defining the waiting time to be the customer's time spent in the queue while waiting for services, and response time to be the customer's time spent in the system, then

 $\mathbf{T}_{\mathbf{i}} = U_{\mathbf{i}} + V_{\mathbf{i}}....(7)$

In equation (7) V_irepresent service time of the queue system where U_i and T_j are assumed to be random variables. Thus, their mean values represented by \overline{U}_i and \overline{T}_j respectively can be used to measure the efficiency of the queue system. If we proceed to represent N(t) = 0 as the event that at time T the server is idle. Then the server utility during time T can mathematically be expressed as

Where N(t) > 0 and T is a protracted time interval. As T approaches infinity (∞), equation (9) is established with probability 1.

$$U_{s} = \lim_{T \to \infty} \frac{1}{T} \int_{0}^{T} x(N(t) \neq 0) dt = 1 - P_{0} = \frac{E\delta}{E\delta + Ei}.....(9)$$

If we represent P_0 as the steady state probability that the server is idle while $E\delta$ and Ei represent the average busy period, average idle period of the server respectively then, by applying ergodic property of Markov chain contained Feller [14] which states that if X(t) is an ergodic Markov chain and A is a subset of its state space equation (10) holds with probability 1.

Where m(A) represents the mean time of the chain in subset A and $m(\overline{A})$ is the mean time of \overline{A} during a cycle, respectively. While P_i represents the ergodic (stationary, steady state) distribution of X(t).

By Burke's Theorem stated in Feller [14], at each period *t*, the population of customers N(t) present in the queue is not determined by the order of their departure times prior to t in *single server* queue systems have steady states with arrival rate λ .

3.0 Analysis of Performance Measure of M / M / 1 Queue with Balking Customers

Balking customers are customers who enter a queue system and exit without receiving service. A modification of a single server queue (M / M / 1) system in which customers are impatient for service as a result of too many arrivals is being considered in this paper. Let b_j represent the probability of a customer joining the system where k is the number of customers already in the queue system at the time he arrived. This can be mathematical expressed as

From equation (11), it obvious that b_j represent various category of customers. The objective of this research is to find such probabilities formulas for the main performance measures without complication. Hence, we consider the following equations (12) and (13).

By applying the normalization condition to equation (13) we obtain equation (14).

 $P_j = \frac{p^j}{j!} e^{-p}, \ j = 0, 1, \dots$ (14)

Stability condition is usually *p* less than infinity ($p < \infty$), since condition p < 1 is not needed in single server queues. Since the customers' arrival satisfies Poisson distribution law, equation (15) can be used to obtain the performance measure of the system as follows:

Performance Measures

$$\bar{X} = \bar{N} - U_s = p(1 - x^{-p}) = p + x^{-p} - 1$$

$$E(X^2) = \sum_{k=1}^{\infty} (k - 1)^2 P_k = \sum_{k=1}^{\infty} k^2 P_k - 2 \sum_{k=1}^{\infty} k P_k + \sum_{k=1}^{\infty} P_k$$

$$= E(X^2) - 2\bar{N} + U_s = p + p^2 - 2p + U_s = p^2 - p + 1 - x^{-p}$$

Thus

$$Var(X) = E(X^{2}) - (E(X))^{2} = p^{2} - p + 1 - x^{-p} - (p + x^{-p} - 1)^{2}$$

= $p^{2} - p + 1 - x^{-p} - p^{2} - x^{-2p} - 1 - 2px^{-p} + 2p + 2x^{-p}$
 $p - x^{-2p} + x^{-p} - 2px^{-p}$
= $p - x^{-p}(x^{-p} + 2p - 1).....(18)$

Bayes rule in equation (2) can be used to determine the probability of customer joining the system at his arrival as shown in equation (19) and (20).

$$P_{J} = \lim_{h \to 0} \frac{\sum_{j=0}^{\infty} (\lambda_{j}h + o(h))P_{j}}{\sum_{j=0}^{\infty} (\lambda_{h} + o(h))P_{k}} = \frac{\sum_{j=0}^{\infty} \lambda_{j}P_{j}}{\sum_{j=0}^{\infty} \lambda_{P}} = \frac{\mu(1 - x^{-p})}{\lambda} = \frac{1 - x^{-p}}{p}....(19)$$

In order to determine the response and waiting distribution times, we need to know the distribution of the queue prior to customers' arrival to joins the system. This also requires an application of Bayes rule. Hence, we obtain.

We observe that in equation (19) and (20), $II_k \neq P_k$.

In this case, we shall first of all determine the mean response time \overline{T} and then the waiting time \overline{W} . By applying the law of total expectation, we obtain equations (21) and (22)

$$\bar{T} = \sum_{k=0}^{\infty} \frac{k+1}{\mu} \Pi_k = \frac{1}{\mu} \sum_{k=0}^{\infty} \frac{(k+1)P_{k+1}}{1-x^{-p}} = \frac{1}{\mu(1-x^{-p})} \quad N = \frac{p}{\mu(1-x^{-p})}...(21)$$

$$\bar{W} = \bar{T} - \frac{1}{\mu} - \frac{1}{\mu} \left(\frac{\rho - x^{\rho} - 1}{1-x^{-\rho}} \right)....(22)$$

We recall from equation (5), that

Equation (24) conforms to Little's formula for(M/M/1) queue system with Balking Customers. T and W distributions can be determined by applying the same procedure as stated earlier. That is

$$f_T(x) = \sum_{j=0}^{\infty} f_T(x|j) \quad \text{II}_j = \sum_{j=0}^{\infty} \frac{\mu(\mu x)^j x^{-\mu x}}{j!} \quad \frac{p^{j+1}}{(j+1)!} \frac{x^{-p}}{1-x^{-p}}$$

Since the evaluation of $f_T(x)$ including $f_w(x)$ is difficult, we apply Laplace transform to evaluate them. Since the Laplace transform $L_T(s)$ and $L_w(s)$ can be obtained, it means we can derive their higher moments. Thus,

Mean Response Time

The mean response time \overline{T} can be determined by using equation (26) to verify the formula as follows:

$$L'_{T}(s) = \frac{x^{-p}}{1 - x^{-p}} \cdot x^{\frac{\mu p}{\mu + 1}} (-\mu p (\mu + s)^{-2})$$

$$L'_{T}(0) = \frac{x^{-p}}{1 - x^{-p}} x^{p} \cdot \frac{p}{\mu} = -\frac{p}{\mu (1 - x^{-p})}$$

Hence

$$\bar{T} = \frac{p}{1 - \mu p} \cdot \mu p = -\frac{p}{\mu (1 - x^{-p})} \cdot \mu p = -\frac{p}{\mu (1 - x^{-p})}$$
(27)

 $\mu(\overline{1-x^{-p}})$ (27)

Equation (27) is the average time response of the queue. Similarly, the average waiting time can be obtained by using equation (26). The Laplace – transform method can be used to obtain Var(T) and (W) of the queue system as stated below:

Equation (29) is the variance of the response time. Moreover, since U and T are considered as a random sum, it follows that:

Var (U) =
$$\left(\frac{1}{\mu}\right)^2 \left(\frac{1}{1-x^{-p}}(p+x^{-p}-1) + \frac{p-x^{-p}(p^2+p)}{(1-x^{-p})^2}\right)$$

= $\frac{1}{(\mu(1-x^{-p}))^2}((p+x^{-p}-1)(1-x^{-p}) + p - x^{-p}(p^2+p))$

Thus

Var(T) = Var(U) +
$$\frac{1}{\mu^2}$$

Var(T) = $\left(\frac{1}{\mu(1-x^{-p})}\right)^2 (p+x^{-p}-1)(1-x^{-p}) + p - x^{-p}(p^2+p) + (1-x^{-p})^2$

$$=\frac{(1-x^{-p})(p+x^{-p}-1+1-x^{-p})+p-x^{-p}(p^{2}+p)}{(\mu(1-x^{-p}))^{2}}$$
$$=\frac{2p-2px^{-p}-p^{2}x^{-p}}{\mu(1-x^{-p})^{2}}\dots\dots\dots(34)$$

which is the variance of the response time.

Note: If λ and μ represent measurements of demand for service and the capacity of performance of the service facility respectively, then $\mu - \lambda$ represents the excess capacity of the queue system to satisfy the demand. By following similar procedure, other performance measures of queue system of M/M/I are obtained as follows:

(a) Waiting and Service time on the queue is

$$W_s = \frac{\lambda}{\mu(\mu - \lambda)}....(35)$$

(b) Probability that there is no customer in the queue:

$$P(x>1) = \left(\frac{\lambda}{\mu}\right)^2.$$
(36)

4.0 **Numerical Example**

A road transport company has two clerks working in its office. The first clerk only handles V.P. passengers' registration/issuance of receipts while the second clerk handles the business class passengers. Assuming the service time for both clerks is exponentially distributed with mean service time 2 minutes per passengers. The arrival rate of V.P. corresponds to Poisson distribution law with a mean arrival rate of 16 per hour. While the business class passengers arrival rate is 14 per hour. Determine the following performance measures of the queue system:

(a) The average waiting time of V.P. passengers and business class passengers if each clerk could handle both the V.P passenger class and the business class passengers.

(b) What would be the effect if we increase the service time 7 minutes.

Solution

р

There are two independent queuing systems: V.P passenger class and business passenger class with the arrival following Poisson distribution and the service time being exponentially distributed.

V.P. class Passenger: Given that $\lambda = 14/h$ and For

$$\mu = 2/\text{minutes}$$
 (i.e. 30/hour)

By applying our expected waiting time formula we have
$$\frac{\lambda}{14}$$
 $\frac{14}{7}$ $\frac{7}{7}$

$$W_{q} = \frac{1}{\mu(\mu - \lambda)} = \frac{1}{30(30 - 14)} = \frac{1}{15x16} = \frac{1}{240} hour$$

$$\frac{7x60}{240} = 1.75 \text{ min } utes$$

$$\approx 2 \text{ min } utes$$

For the Business Passenger Class

Expected waiting time in Queue:

 $W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{16}{30(30 - 16)} = \frac{8}{15x14} = \frac{4}{105} hour$ $= \frac{4x60 \min \ utes}{105} = 2.29 \min \ utes$

Treating the given problem as a single queue with two servers (Clerks), we have the following parameter values:

 $\lambda = 14 + 16 = 30 / \text{h}, \ \mu = 30/\text{h}, \ \text{number of servers/clerks s} = 2, \ P = \frac{\lambda}{s\mu} = \frac{30}{2x30} = \frac{1}{2}$

Now
$$P_{o} = \left[\sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^{n} + \frac{1}{s!} \left(\frac{s\mu}{s\mu - \lambda}\right)\right]^{-1}$$

$$= \left[\sum_{n=0}^{1} \frac{1}{n!} \left(\frac{1}{2}\right)^{n} + \frac{1}{2!} \left(\frac{1}{2}\right)^{s} \left(\frac{60}{60 - 30}\right)\right]^{-1}$$

$$= \left[1 + \frac{1}{2} + \frac{1}{2} \left(\frac{1}{2}\right) \cdot 2\right]^{-1}$$
Po $= \left[1 + \frac{1}{2} + \frac{1}{2} \left(\frac{1}{2}\right)\right]^{-1} = 2^{-1} = \frac{1}{2}$

Average Waiting time of arrival in the queue:

$$W_{q} = \frac{L_{q}}{\lambda} \left[\frac{1}{(S-1)!} \left(\frac{\lambda}{\mu} \right)^{S} \frac{\mu}{(S\mu - \lambda)^{2}} \right] P_{o}$$
$$= \left(\frac{1}{2} \right)^{2} \frac{30}{(60 - 30)^{2}} X \frac{1}{2} = \frac{1}{4} X \frac{1}{30} X \frac{1}{2} = \frac{1}{240} hour$$
$$= \frac{60}{240} \min \ utes = \frac{1}{4} = 0.25 \min \ ute$$

To obtain the combined waiting time with 4.5 increase service time, when $\lambda = 30/h$ we will have $\mu = \frac{60}{4.5} = \frac{60}{90} = 60 \text{ x} \frac{2}{9} = \frac{20x^2}{2} = \frac{40}{2}/hour$.

$$4.5 \quad 90 \quad 9 \quad 3 \quad 3$$

Note: $\frac{\lambda}{\mu} = \frac{30}{40} x^3 = \frac{9}{4}$
Hence $P_0 = \left[\sum_{n=0}^{1} \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n + \frac{1}{s!} \left(\frac{\lambda}{\mu}\right)^s \frac{s\mu}{s\mu - \lambda}\right]^{-1}$(4)

By substituting the above values into (4) we have

$$P_{o} = \left[\sum_{n=0}^{1} \frac{1}{n!} \left(\frac{9}{4}\right)^{n} + \frac{1}{2!} \left(\frac{9}{4}\right)^{2} \frac{2\left(\frac{40}{3}\right)}{2\left(\frac{40}{3}\right) - 30}\right]^{-1}$$
$$= \left[1 + \frac{9}{4} + \frac{81x2x\frac{40}{3}}{\left[2x^{2}6x2x\frac{40}{3}\right]} - \frac{1}{30}\right]^{-1}$$

$$= \left[1 + \frac{9}{4} + \frac{81}{16} - \frac{1}{30}\right]^{-1} = \frac{16}{133} \text{ min } utes$$

Expected waiting time of arrivals in the system is

$$W_{q} = \frac{1}{(S-1)!} \left(\frac{\lambda}{\mu}\right)^{5} \frac{\mu}{(S\mu - \lambda)^{2}} X P_{o}$$
$$= \left(\frac{9}{4}\right)^{2} \frac{\frac{40}{3}}{\left(\frac{40}{3} - 30\right)^{2}} X \frac{16}{133} = \frac{243}{10} \min utes = 24.3 \min utes$$

5.0 Analysis of Results

From the results presented in the numerical example of section 4.0, we observed that the time a VIP passenger has to wait before being attended to by the clerk is 2 minutes while a business passenger waited for 2.29 minutes before receiving service. If the queue problem is treated as a single queue with the two clerks as servers, then the average time to wait for a customer to arrive at the queue system is 0.25minute. While the combined average waiting time for a VIP passenger and a business passenger to arrive at the system with 4.5 increase service time 2.7 minutes. With increase in service time to 4.5, the expected time to wait for a customer to arrive at the system is 24.3 minutes.

6.0 Conclusion:

The measures of performance of a queue system hinges on the ability to manage its service facility with the aim of striking an optimum balance between the waiting cost of time and cost of keeping the system idle, Ogumeyo and Nwamara [2]. Waiting cost can be direct or indirect. Direct waiting cost consists of unutilized manpower and equipment which has to be paid for by business owners. This increases the cost of production of goods and services. Indirect cost centres on loss of customer which leads to decreased sales and less profits. Hence, there is a great need to develop mathematical queue model to enhance efficient management of queue in business organizations. The purpose of this research is to expand the queue model in [2] with the aim of analysing its performance measures which include: (a) number of customers in the queue plus customers currently receiving service (b) Queue Length (c) Length of time a customer is expected to wait before he receives service (i.e expected waiting time) (d) The fluctuation rate (variance) of the queue length. (e) The probability that there is no customer in the queue. (f) The probability that the queue system exceeds a given capacity K. In this paper, we have mathematically derived and analysed the performance measures of single server queue systems based on the distributions of waiting and response times. To get the probability distributions, we observed from the analysis that we have to first of all, know the distribution of the system at the time an arriving customer joins the queue. This was obtained by applying Bayes rule of total probability. The formulas for the means and variances were later transformed for easy computation by using generating functions and Laplace- transforms. It was also observed that, the formulas for the mean and variance of the distribution of the response and waiting times conform to the Little's theorem on queuing system. This was illustrated in the numerical example presented in section 4.0.

References:

- [1] Hiller, F.S, and Lierberman, G.J. (2010). Introduction to Operations Research. Irwin/McGraw-Hill
- [2] Ogumeyo S.A and Nwamara C.C. (2019).Derivation of A Finite Queue Model with Poisson Input and Exponential Service. Journal of the Nigerian Association of Mathematical Physics Volume 52 pp.53-58
- [3] Wagner, H.M. (2001). 'Principles of Operations Research' pp. 854-865. Prentice-Hall of India
- [4] Bronson R. and Naadimuthu (1997). Operations Research 2nd Edition, Schaum's outline series, McGraw-Hill, New York

- [5] Subaagyo, P, Marwan, A. (1992) Operations Research, Yogakarta: BPFE
- [6] Kakaiy, T.J. (2004). DasarTeoriAntrianUntukKehidupanNyata. Yogyakarta: "Queue Models with Balking and Reneging". Available online at https://doi.org/10.1051/ro/2019064.
- [7] Bohm, W. (2016) 'A Course on Queuing Models'. Chapman and Hall/CRC.
- [8] Nugraha, Dedi, (2013). Penentuan Model System AntreanKendaran di GerbangTolBanyumanik, Skripsi, FSM, Statistika, UniversitasDiponegoro.
- [9] Kembe, M.M., Onah, E.S., Lorkegh, S.A., (2012). A Study of Waiting and Service Costs of a Multi-Server Queuing Model in a Specialist Hospital, International Journal of Scientific and Technology Research, 5(2): 2277-8616.
- [10] Bhat, U.N. (2005) 'An introduction to Queuing Theory: Modeling and Analysis in Applications. Birkhauser.
- [11] Kobayashi, H (1978) 'Modelling and Analysis'. An Introduction to System Performance Everluation Methodology. Addison-Wesley, Reading, MA.
- [12] Weber, T. (2019) Solving Performance Models Based on Basic Queuing Theory Formulas
- [13] Takagi, H. (1993) Queuing Analysis. A Foundation of Performance Evaluation. Vol.2. Finite systems. North-Holand, Amsterdam
- [14] Feller, W. (1968) "An Introduction to Probability Theory and its Applications" Vol. 1, 3rd Edition. New York: Wiley Inc.