Vol. 33, No. 1 (2023) | DOI: 10.37190/ord230102

OPEN ACCESS

Operations Research and Decisions

www.ord.pwr.edu.pl

OPERATIONS RESEARCH AND DECISIONS QUARTERLY

Frequentist inference on traffic intensity of M/M/1 queuing system

Kaustav Dutta[1*] Amit Choudhury[1]

[1]*Department of Statistics, Gauhati University, Guwahati, Assam, India*

[*]*Corresponding author: kaustav0708@gmail.com*

Abstract

When we study any queuing system, the performance measures reflect different features of the system. In the classical M/M/1 queuing system, traffic intensity is perhaps the most important performance measure. We propose a fresh and simple estimator for the same and show that it has nice properties. Our approach is frequentist. This approach has the dual advantage of practical usability and familiarity. Our proposed estimator is attractive as it possesses desirable properties. We have shown how our estimator lends itself to testing of hypothesis. Confidence intervals are constructed. Sample size determination is also discussed. A comparison with a few similar estimators is also performed.

Keywords: *confidence interval, estimation, hypothesis testing, performance measures, single server Markovian model, traffic intensity*

1. Introduction

Over the last few decades, there has been growing interest in the statistical inference of parameters, performance measures and related parametric functions in the context of queuing systems. This interest has particularly grown in the last three decades or so. Even though the queuing theory has been in vogue for about one hundred years or so, the emphasis so far has largely been on the development of probability models to stochastically describe the myriad variety of practical queuing systems existing in real life and design their performance measures. However, when it comes to the actual implementation of these theoretical models to analyse practical queuing situations occurring in day-to-day life, practising queuing specialists have been facing a dilemma. In the absence of scientifically designed estimators of different queuing parameters and performance measures, they have had to resort to ad hoc methods. Decision-making has suffered in the sense that proper methods of statistical inference are still in the developmental phase in the context of queuing theory.

Received 11 May 2022, accepted 8 February 2023, published online 16 April 2023
ISSN 2391-6060 (Online)/

The costs of publishing this issue have been cofinansed by the program *Development of Academic Journals* of the Polish Ministry of Education and Science under agreement No. RCN/SP/0241/2021/1

Whenever any queuing analyst is required to analyse any live queuing situation, numerical determination of the performance measures becomes paramount. Medhi [26] stated that analysis of performance measures is very important as it helps in the identification of various queuing issues.'The main objective of any queuing study is to assess some well-defined parameters through which the nature of the quality of service can be studied. These parameters are known as performance measures. Performance measures are important as issues or problems caused by queuing situations are often related to customers' dissatisfaction with service or maybe the root cause of economic losses in a business. Analysis of the relevant performance measures of queuing models allows the cause of queuing issues to be identified and the impact of proposed changes to be assessed' (Choudhury and Medhi[13]).

Therefore, the need to apply theoretical queuing models to analyse real-life situations has led to demands for appropriate methods and techniques for the construction of estimators and its related aspects as well as matters relating to testing of hypothesis – all of which are constituents of statistical inference. This is the research gap that this paper shall address.

In general, statistical inference problems can be divided into two types namely parameter estimation and distribution selection. In the first type, a pre-specified stochastic model is selected and then its parameters are estimated. In the latter type, the collected data is examined to determine a particular probabilistic model. In the context of queuing theory, we are usually interested in statistical inferential problems of the first type more specifically we will be interested to know about the methods of construction of estimators of the performance measures.

In this paper, we propose a method for estimating the traffic intensity of the M/M/1 queuing model. The most powerful critical region for testing the traffic intensity is given. The confidence interval for the estimator is evaluated. An approach is shown for the determination of the sample size. The proposed estimator is compared with similar frequentist estimators. In the paper, Section 2 contains the literature review. Section 3 describes the M/M/1 queue. Section 4 contains the procedure for estimating the traffic intensity. Section 5 contains hypothesis testing. Section 6 contains the confidence interval construction. The technique of sample size determination is given in section 7. Section 8 contains the discussion and the comparison of estimators with other frequentist estimators. Section 9 concludes the paper.

2. Literature review

The literature of statistical inference has developed in two branches viz. frequentist and Bayesian. In frequentist framework, the pioneering work is due to Clarke [17] who first obtained the maximum likelihood estimates of the arrival rate and service rate of the M/M/1 queuing model. Lilliefors [25] estimated the confidence intervals for the M/M/1 traffic intensity from the maximum likelihood estimates given by Clarke [17]. Bhat and Rao [9] proposed a technique based on statistical quality control to control traffic intensity in M/G/1 and GI/M/1 queues. Dave and Shah [20] derived estimators for arrival and service rates in an M/M/2 queuing model with heterogeneous servers using the maximum likelihood principle. Basawa and Prabhu [7] studied the moment estimates and maximum likelihood estimates for non-parametric and parametric models for a single server queue. They also investigated the limiting distribution of these estimators. Schruben and Kulkarni [30] considered stationary M/M/1 queuing model and showed that estimates of arrival rates and service rates result in a notable discrepancy between the estimated parameters for the model and actual parameters of the system. They have shown that the mean for the model does not exist even when the traffic intensity is restricted to be strictly less than one and also found out that the expected value of the estimator of popular measure is infinite regardless of the value of the actual traffic intensity. Basawa and Prabhu [8] discussed the problems of large sample estimation and tests for parameters in a single server queue and presented asymptotic properties for performance measures. Basawa et al. [6] developed estimates of inter-arrival and service time distribution in GI/G/1 model by developing estimating function involving waiting time data. They also compared the estimates with those obtained by maximum likelihood estimates. Zheng and Seila [38] extended the work by Schruben and Kulkarni [30] and have shown mathematically the nonexistence of standard errors and expectations of the common estimators of performance measures in an M/M/1 model. In turn, they proposed a method to construct estimators for the performance measures of the M/M/1 model which have desirable properties.

Srinivas et al. [33] discussed uniform minimum variance unbiased estimators (UMVUE) and maximum likelihood estimators of the various characteristics of M/M/1 queue. They also compared these estimates using the asymptotic expected deficiency (AED) criterion where they suggested the use of uniform variance unbiased estimators over maximum likelihood estimators for some measures. Srinivas and Udupa [34] considered M/M/1 queue and developed the best-unbiased estimation and CAN property for performance measures. They derived various UMVU estimators of performance measures using the Lehmann–Scheffe theorem. They also derived the probability distributions of UMVU estimators. Methods to construct asymptotic confidence intervals for traffic intensity and other performance measures are also suggested by them. Srinivas and Kale [32] studied the M/D/1 queuing system where they developed maximum likelihood estimation and UMVU estimation of traffic intensity and other performance measures. CAN property of maximum likelihood estimators is established. They also compared the maximum likelihood estimators with the UMVU estimators. Suyama et al. [35] considered M/M/s queuing model and derived MLE of ρ. They also showed that the estimate is equivalent to the moment estimator. Choudhury and Basak [11] derived an ML estimator of traffic intensity by exploiting the relationship between M/M/1 process and Bernoulli process. They also discussed the determination of sample size for the estimation of traffic intensity using a randomized testing procedure. Dutta and Choudhury [23] considered the M/M/1 queuing model where they extended the work of Zheng and Seila [38] to give a guiding principle about the use of alternative estimators for M/M/1 queuing model.

The Bayesian approach in queuing theory largely developed due to the work by Armero and Bayarri [2–4]. Choudhury and Borthakur [12] derived Bayesian estimates and credibility intervals of traffic intensity. They also presented the predictive distribution of system size at the departure epoch. Chowdhury and Mukherjee[14] discussed the estimation of waiting time in the M/M/1 queuing model in the form of its right tail area called exceedance probability. They evaluated the MLE of rate parameters λ and μ and also the exceedance probability. Moreover, their large sample properties were also studied. Chowdhury and Mukherjee [15] considered M/M/1 queuing model and constructed the Maximum Likelihood and Bayes estimator of traffic intensity. They also compared their frequentist and Bayesian estimators of traffic intensity using simulation. Jose and Manoharan [24] considered M/M/1 queuing model and obtained the arrival rate, service rate, and traffic intensity by using a bivariate prior distribution of arrival rate and service rate. They also obtained posterior distribution and the credible region for the traffic intensity.

Predictive distribution for system size was also obtained. Chowdhury and Mukherjee [16] considered the M/M/1 queuing model and obtained the Bayes estimator of traffic intensity and various other performance measures under the squared error loss function (SELF) and precautionary loss function (PLF) with beta-Stacy distribution as prior. They also performed a simulation study and observed that beta-Stacy as prior yields stable estimates of the queuing performance metrics (QPMs).

Cruz et al. [19] considered Bayesian estimation in the M/M/S model by using beta distribution as prior for ρ. They presented a closed-form expression for the predictive distribution of the number of customers in the system at the departure epoch. Almeida and Cruz [1] studied M/M/1 model and considered Jeffreys before obtaining the posterior distributions of some parameters of interest. Cruz et al. [18] studied M/M/1/k queuing model where they used Bayesian inference and Monte Carlo simulation techniques to evaluate estimators under finite samples. Deepthi and Jose [21] considered M/M/R queuing model and derived the conditional posterior densities of mean arrival rate and mean service rates by assuming multivariate gamma distribution as prior for service rates and gamma distribution as prior for arrival rates. They also used the Markov chain Monte Carlo method to obtain the Bayes estimate and credible interval for M/M/3 queuing model as a particular case of M/M/R model under various loss functions. Deepthi and Jose [22] considered M/Ek /1 queuing model and described the Bayesian estimation of queue parameters and various queue performance measures by using Mckay's bivariate gamma distribution as prior under entropy loss function and squared error loss function. They also obtained the closed expressions for the Bayes estimators. They performed a simulation study to compute Bootstrap Bayes estimate and credible regions of various queue characteristics and compared them with those obtained using Markov Chain Monte Carlo method. Basak and Choudhury [5] considered the estimation of traffic intensity in a M/M/1 queuing model. They derived the Bayes estimator of ρ under squared error loss function assuming two forms of prior information on ρ. They also compared their proposed Bayes estimators with the estimators based on the maximum likelihood principle.

3. The M/M/1 queuing system

The single-server Markovian queuing model, also known as the M/M/1 model in Kendall's notation, *is the simplest non-trivial queue* [36], yet is possibly the most widely used. In this model, it is assumed that customers arrive into the system at the rate of λ, and the time between successive arrival of customers follows an exponential distribution. There is just one server and the rate at which the server offers service is μ. The time required to serve each customer (by the server) is also random and follows an exponential distribution. Service times of customers are independent of each other. There is no restriction on the waiting space, and customers are served first come first served principle. The calling population is assumed to be infinite.

One of the important parameters of this model is traffic intensity ρ, which is defined as the ratio of arrival rate λ to service rate μ. To ensure the equilibrium and stability of the model, this parameter must be less than one. Assuming equilibrium is very frequent in queuing theory [4]. The restriction that traffic intensity should be less than one implies that $\lambda < \mu$ which is both a necessary and sufficient condition. The length of the queue would otherwise "explode" as the number of customers would go on increasingly indefinitely. Even if case this assumption is not initially met, operational manager(s) usually tweak the system and see to it that the queue size does not go on increasing indefinitely. This in turn ensures that the restriction on traffic intensity is met.

When analysing any real-life queuing phenomenon, various features are generally of interest so as to know about the effectiveness of the queuing system. These are captured by what are called performance measures. Generally, there are three types of such measures of interest viz. (1) some measure of the waiting time that a typical customer might endure, (2) some measure of the number of customers that may accumulate in the queue or system, and (3) a measure of the idle time of the servers. Since most queuing systems have stochastic elements, these measures are often random variables, so their probability distributions – or at least their expected values – are sought [31]. For the M/M/1 system, the widely used performance measures are:

- mean system size $L_s = \dfrac{\lambda}{\mu - \lambda}$,
- mean queue size $L_q = \dfrac{\lambda^2}{\mu(\mu - \lambda)}$,
- average waiting time in the system $W_s = \dfrac{1}{\mu - \lambda}$,
- average waiting time in a queue $W_q = \dfrac{\lambda}{\mu(\mu - \lambda)}$,
- traffic intensity $\rho = \dfrac{\lambda}{\mu}$.

4. Estimation of traffic intensity

We consider two data streams, one each from the arrival process and the service mechanism. Let $x_1, x_2, \ldots, x_{n_1}$ be a random sample of inter-arrival times drawn from the queuing model. Because of the assumptions of the model (outlined in Section 3), these inter-arrival times would be independent and identically distributed samples drawn from $\exp \lambda$. Similarly, let $y_1, y_2, \ldots, y_{n_2}$ be a random sample of service times of the server of the model. Again because of the assumptions of the model, this random sample will be independently and identically distributed from exp μ. Since inter-arrival and service times are independent of each other, it is also assumed that $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$ are mutually independent.

We now define $z_1 = \sum_{i=1}^{n_1} x_i \sim \gamma(n_1, \lambda)$ and $z_2 = \sum_{j=1}^{n_2} y_j \sim \gamma(n_2, \mu)$. Then, the distributions of z_1 and z_2 are given by [28]

$$f(z_1) = \frac{\lambda^{n_1}}{\Gamma(n_1)} \mathrm{e}^{-\lambda z_1} {z_1}^{n_1 - 1}, \quad z_1 > 0$$

$$f(z_2) = \frac{\mu^{n_2}}{\Gamma(n_2)} \mathrm{e}^{-\mu z_2} {z_2}^{n_2 - 1}, \quad z_2 > 0$$

The joint distribution of z_1 and z_2 is

$$f(z_1, z_2) = \frac{\lambda^{n_1}}{\Gamma(n_1)} \frac{\mu^{n_2}}{\Gamma(n_2)} \mathrm{e}^{-\lambda z_1 - \mu z_2}\, {z_1}^{n_1 - 1} {z_2}^{n_2 - 1}, \quad z_1 > 0,\ z_2 > 0 \tag{1}$$

We now make the following transformation

$$u = \frac{z_2}{z_1} \quad \text{and} \quad v = z_1 \tag{2}$$

We have from (1)

$$f(u, v) = \frac{\lambda^{n_1}}{\Gamma(n_1)} \frac{\mu^{n_2}}{\Gamma(n_2)} e^{-v(\lambda+\mu u)} v^{n_1+n_2-1} u^{n_2-1}, \quad u > 0,\ v > 0$$

After some elementary algebra, it can be shown that the distribution of u is

$$f(u) = \frac{\lambda^{n_1}}{\Gamma(n_1)} \frac{\mu^{n_2}}{\Gamma(n_2)} u^{n_2-1} \frac{\Gamma(n_1+n_2)}{(\lambda+\mu u)^{n_1+n_2}}, \quad u > 0 \tag{3}$$

Therefore,

$$E(u) = \frac{n_2}{n_1 - 1} \frac{\lambda}{\mu}$$

Thus, the mean of this distribution is $\frac{n_2}{n_1-1}\frac{\lambda}{\mu}$ and it is now possible to construct an unbiased estimator of traffic intensity ρ using the result

$$E\left(\frac{n_1 - 1}{n_2} u\right) = \frac{\lambda}{\mu} \tag{4}$$

Hence, $\frac{n_1-1}{n_2}u$ is an unbiased estimator of ρ.

Again, after some algebra, it can be shown that,

$$E\left(\frac{n_1 - 1}{n_2} u\right)^2 = \frac{(n_2+1)(n_1-1)}{n_2(n_1-2)} \left(\frac{\lambda}{\mu}\right)^2 \tag{5}$$

Relations (4) and (5) now give us

$$V\left(\frac{n_1 - 1}{n_2} u\right) = \frac{(n_1+n_2-1)}{n_2(n_1-2)} \left(\frac{\lambda}{\mu}\right)^2$$

$$\therefore V\left(\frac{n_1 - 1}{n_2} u\right) \to 0 \quad \text{as} \quad n_1 \to \infty, n_2 \to \infty$$

Hence, $\frac{n_1-1}{n_2}u$ is a consistent estimator of ρ.

Again from (3),

$$\begin{aligned} f(u) &= \frac{\lambda^{n_1}}{\Gamma(n_1)} \frac{\mu^{n_2}}{\Gamma(n_2)} u^{n_2-1} \frac{\Gamma(n_1+n_2)}{(\lambda+\mu u)^{n_1+n_2}}, \quad u > 0 \\ &= \left(\frac{1}{\rho}\right)^{n_2} \frac{1}{\left(1+\frac{u}{\rho}\right)^{n_1+n_2}} \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\Gamma(n_2)} u^{n_2-1}, \quad u > 0 \end{aligned} \tag{6}$$

$$= g(t, \rho)\, h(u) \ \text{ where } t = u$$

Hence, by using Neyman factorisation theorem, u is a sufficient estimator of ρ, i.e., $\dfrac{n_1-1}{n_2}u$ is a sufficient estimator of ρ. Our estimator $\dfrac{n_1-1}{n_2}u$ is promising. It is not only unbiased but is also consistent and sufficient for ρ.

5. Testing the hypothesis

Having constructed an estimator, we outline a method to test the hypothesis on ρ. We have from (6),

$$f(u) = \left(\frac{1}{\rho}\right)^{n_2} \frac{1}{\left(1+\dfrac{u}{\rho}\right)^{n_1+n_2}} \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} u^{n_2-1}, \quad u>0$$

Let

$$\frac{n_1-1}{n_2}u = w$$

$$\therefore\ f(w) = \left(\frac{1}{\rho}\right)^{n_2} \frac{1}{\left(1+\dfrac{n_2 w}{(n_1-1)\rho}\right)^{n_1+n_2}} \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \left(\frac{n_2}{n_1-1}\right)^{n_2} w^{n_2-1}, \quad w>0 \tag{7}$$

Let us consider the problem of testing a simple hypothesis: $H_0 : \rho = \rho_0$ against a simple alternative hypothesis $H_1 : \rho = \rho_1$

According to the Neyman–Pearson lemma, the most powerful critical region is given by [28]

$$\frac{f(w,\rho_1)}{f(w,\rho_0)} > k \implies \left(\frac{\rho_1}{\rho_0}\right)^{n_1} \left(\frac{(n_1-1)\rho_0+n_2 w}{(n_1-1)\rho_1+n_2 w}\right)^{n_1+n_2} > k$$

Case 1. If $\rho_1 > \rho_0$

$$(n_1-1)\rho_0 + n_2 w > k_1((n_1-1)\rho_1 + n_2 w) \implies w > k_2$$

Therefore, the critical region is $W = (w :\ w > k_2)$.

Case 2. If $\rho_1 < \rho_0$

$$(n_1-1)\rho_0 + n_2 w < k_3((n_1-1)\rho_1 + n_2 w) \implies w < k_4(say)$$

Therefore, the critical region is $W_1 = (w :\ w < k_4)$.

The constants k_2 and k_4 are so chosen that the size of the critical region is α. Thus, k_2 is determined.

So that taking $\dfrac{n_2 w}{(n_1-1)\rho_0} = t$ in equations (8)–(10) we arrive at:

$$
\begin{aligned}
&P[w \in W|H_0] = \alpha \implies P_{H_0}(w > k_2) = \alpha \\
&\implies \int_{k_2}^{\infty} \left(\frac{1}{\rho_0}\right)^{n_2} \frac{1}{(1+t)^{n_1+n_2}} \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \left(\frac{n_2}{n_1-1}\right)^{n_2} w^{n_2-1} dw = \alpha \\
&\implies \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \int_{\frac{n_2 k_2}{(n_1-1)\rho_0}}^{\infty} \frac{t^{n_2-1}}{(1+t)^{n_1+n_2}} dt = \alpha
\end{aligned}
\tag{8}
$$

The integral in (8) is an incomplete beta integral which can be solved using standard techniques. k_4 can be determined from the incomplete beta integral given below

$$
\frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \int_{0}^{\frac{n_2 k_4}{(n_1-1)\rho_0}} \frac{t^{n_2-1}}{(1+t)^{n_1+n_2}} dt = \alpha
$$

6. Estimation of the confidence interval

Let U and L be the upper limit and lower limits, respectively, for the estimator given in (4), the distribution of which is given in (7). Then U and L can be evaluated using the following equation

$$
P(w > U) = \frac{\alpha}{2} \quad \text{and} \quad P(w < L) = \frac{\alpha}{2}
$$

where

$$
\begin{aligned}
P(w > U) = \frac{\alpha}{2} &\implies \int_{U}^{\infty} \left(\frac{1}{\rho_0}\right)^{n_2} \frac{1}{(1+t)^{n_1+n_2}} \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \left(\frac{n_2}{n_1-1}\right)^{n_2} w^{n_2-1} dw = \frac{\alpha}{2} \\
&\implies \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \int_{\frac{n_2 U}{(n_1-1)\rho_0}}^{\infty} \frac{t^{n_2-1}}{(1+t)^{n_1+n_2}} dt = \frac{\alpha}{2}
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
P(w < L) = \frac{\alpha}{2} &\implies \int_{0}^{L} \left(\frac{1}{\rho_0}\right)^{n_2} \frac{1}{(1+t)^{n_1+n_2}} \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \left(\frac{n_2}{n_1-1}\right)^{n_2} w^{n_2-1} dw = \frac{\alpha}{2} \\
&\implies \frac{\Gamma(n_1+n_2)}{\Gamma(n_1)\,\Gamma(n_2)} \int_{0}^{\frac{n_2 L}{(n_1-1)\rho_0}} \frac{t^{n_2-1}}{(1+t)^{n_1+n_2}} dt = \frac{\alpha}{2}
\end{aligned}
\tag{10}
$$

The integrals in (9) and (10) are incomplete beta integrals which can be solved using numerical techniques.

7. Determination of sample size

We proposed a frequentist estimator for traffic intensity (ρ) in (4). It has a number of desirable properties viz. unbiasedness, consistency and sufficiency. It lends itself to testing of hypothesis on ρ. The sampling distribution is also tractable – a major characteristic which many frequentist estimators lack. The sampling scheme is also straightforward to implement and therefore, any queuing practitioner would find our estimator appealing. The only issue that remains to be settled is some guidelines regarding what should be the sample size. In this section, we shall deal with this issue.

The guideline that we shall prescribe should work for a range of ρ's. Since in most cases $\rho > 0.5$ [15], we carried out the simulation for three typical choices of λ and μ.

1) $\lambda = 12, \mu = 15$, thereby implying that $\rho = 0.80$,
2) $\lambda = 10, \mu = 16$, thereby implying that $\rho = 0.625$,
3) $\lambda = 16, \mu = 18$, thereby implying that $\rho = 0.88$.

We recall that we need two samples for practically implementing our estimator. One of these is the sample of inter-arrival times of size n_1 and the other is the sample of service times of size n_2. We shall provide a guiding rule for what should be n_1 and n_2 using simulation. To this end, we shall compare various choices of n_1 and n_2 as follows:

$n_1 = (25,\ 50,\ 75,\ 100,\ 125,\ 150)$ and $n_2 = (25,\ 50,\ 75,\ 100,\ 125, 150)$.

In practice, we have 36 pairs of sample sizes to compare:

$(25,\ 25), (25,\ 50), \ldots, (25,\ 150), (50, 25),\ ...,\ (150, 150)$.

We shall perform tests of homogeneity between them. Essentially, we are interested in testing equivalence between these pairs. *Equivalence testing is a statistical method designed to provide evidence that groups are comparable by demonstrating that the mean differences found between groups are small enough that they are considered practically unimportant* [29]. *Homogeneity of variance testing is a statistical method designed to provide evidence that groups are comparable by demonstrating that the variations found between groups are small enough that they are considered practically insignificant* [27]. If for any two pairs, our test shows non-significant results, then we will accept the hypothesis of equality of variance. There will be no need to increase the sample size beyond that. The specific pair for which no significance is attained shall be our recommended sample size.

We perform the Brown–Forsythe test for homogeneity of variance. Our choice is dictated by the fact that it is more robust and is relatively insensitive to departures from normality (Brown and Forsythe [10]).

To begin, we shall compare the pairs $(25, 50)$ and $(25, 75)$ for the case $\lambda = 12, \mu = 15$. For this purpose, we generate a random sample of inter-arrival times using a simulation technique under Markovian setup with parameter $\lambda = 12$. Similarly, we generate a random sample from service time using a simulation technique under Markovian setup with parameter $\mu = 15$. The simulation technique for drawing random numbers is the method of inverse transformation, discussed by Taha [37]. Using this procedure, we simulate 10,000 estimates of the traffic intensity for each pair of sample sizes using the estimator (4). The tables presenting the results of our analysis are given in the Appendix (Tables A1–A3).

8. Discussion

Tables A1–A3 in the appendix contain the summary of results from the Brown–Forsythe test for $\lambda = 12$, $\mu = 15$ ($\rho = 0.80$), $\lambda = 10$, $\mu = 16$ ($\rho = 0.625$), and λ = 16, $\mu = 18$ ($\rho = 0.88$), respectively. The null hypothesis of homogeneity of variance is accepted for n_1 and n_2 pairs $(100, 25)$ and higher or $(25, 100)$ and higher. Thus there will be no need to increase the sample size beyond that. We compared our estimators with those given in [5] and [15]. The expression of maximum likelihood estimator of ρ obtained by Basak and Choudhury [5] is

$$\hat{\rho}_{ML} = \frac{\sqrt{n_2^2 + 4\left(y + 2n\right)\left(y + n_2\right)} - n_2}{2\left(y + 2n\right)}$$

where $y = \sum_{i=1}^{n_2} x_i$, n is the number of the M/M/1 queues observed, x_i is the non-empty queue size of n_2 observations.

We first computed our estimator (given in (4)) for different sample sizes and ρ. The process was repeated 5,000 times to evaluate the mean square errors (MSE). The simulation repeated 5,000 times by us corresponds to that carried out by Basak and Choudhury [5]. The results of our calculations are given in Table 1 (columns 2–4) together with the results from Basak and Choudhury [5] (columns 5–7).

Table 1. Estimates and (MSEs) of ρ using estimators given in (4) (columns 2–4) and [5] (columns 5–7)

ρ	n					
	50	100	200l	50	100	200
0.5	0.500080 (0.010559)	0.498276 (0.005057)	0.499000 (0.002443)	0.492379 (0.002954)	0.496542 (0.001402)	0.498042 (0.000708)
0.8	0.797365 (0.026612)	0.798926 (0.012399)	0.801630 (0.006503)	0.795285 (0.000815)	0.797708 (0.00035)	0.798164 (0.000158)
0.9	0.898658 (0.034092)	0.898810 (0.015888)	0.900996 (0.008266)	0.890558 (0.000276)	0.892443 (0.000169)	0.899422 (0.00011)

The maximum likelihood estimator for ρ obtained by Chowdhury and Mukherjee [15] is

$$\hat{\rho}_{\text{mle}} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A}$$

where

$$A = N + n_{00} + n_{10} - n_0 - 1$$

$$B = N + \sum_{i=2}^{\infty} \sum_{j=i-1}^{\infty} (j - i + 1) n_{ij} + n_{00} + n_{10} + \sum_{j=0}^{\infty} j(n_{0j} + n_{1j}) + 1$$

$$C = n_0 + \sum_{j=0}^{\infty} j\left(n_{0j} + n_{1j}\right) + \sum_{i=2}^{\infty} \sum_{j=i-1}^{\infty} \left(j - i + 1\right) n_{ij}$$

$$N = \sum_{i=2}^{\infty} \sum_{j=i-1}^{\infty} n_{ij}, n_{00} = \sum_{j=0}^{\infty} n_{0j}, n_{10} = \sum_{j=0}^{\infty} n_{1j}$$

and n_{ij} is the observed number of transitions in N_t from state i to state j.

For comparison with the maximum likelihood estimator of ρ given by Chowdhury and Mukherjee [15], we followed their simulation procedure. We first computed our estimator (cf. (4)) for different sample sizes and ρ. The process was repeated 1000 times to evaluate the root mean square errors (RMSE). The simulation repeated 1000 times by us corresponds to that carried out in [15]. The results of our calculations are given in Table 2 (columns 2–4) together with the results of [15] (columns 5–7).

Table 2. Estimates and (RMSEs) of ρ using estimators given in (4) (columns 2–4) and [15] (columns 5–7)

ρ	n					
	30	50	100	30	50	100
0.5	0.498067 (0.133917	0.497466 (0.101799)	0.496688 (0.076015)	0.4823 (0.0234151)	0.4941 (0.0169929)	0.4956 (0.0102145)
0.7	0.700486 (0.181277)	0.701708 (0.144778)	0.700464 (0.096699)	0.6739 (0.0314587)	0.6812 (0.0214703)	0.6998 (0.0154366)
0.9	0.900931 (0.232086)	0.895995 (0.175890)	0.899186 (0.125476)	0.8688 (0.0245871)	0.8809 (0.0124587)	0.8912 (0.0095487)

From Table 1, a trade-off can be observed. While our estimator provides estimates closer to actuals compared to those reported in [5], we lose out on MSE. A similar trade-off can be observed from Table 2. Our estimator provides estimates much closer to actuals compared to those in [15], but we lose out on RMSE. From these two tables, we can conclude that our estimator provides estimates of traffic intensity which are closer to actuals and hence are better than those available in the literature. The other advantage is the ease of implementation.

9. Conclusions

A method of estimating the traffic intensity of the M/M/1 queue has been presented. The advantage of the method lies in its simplicity. The estimator is shown to satisfy a number of properties of a good estimator. Confidence intervals are constructed. The most powerful critical region for the estimator of traffic intensity is also constructed. Regarding sample size, from the above discussion, the hypothesis of homogeneity of variance is accepted for n_1 and n_2 pairs $(100, 25)$ and higher or $(25, 100)$ and higher. Thus the hypothesis of homogeneity of variance can be accepted for a total sample of size approximately 125. We, therefore, prescribe that a minimum total sample size of 125 may be used in such a manner that one of the components (n_1 or n_2) is at least 25.

A caveat is in order. The conclusion that we have drawn is on the bases of simulation. We invite other researchers to carry out similar exercises and confirm or improve upon our conclusion.

Acknowledgement

The authors are grateful to the anonymous reviewers for their valuable comments and suggestions made on the previous draft of this manuscript.

References

[1] Almeida, M. A. C., and Cruz, F. R. B. A note on Bayesian estimation of traffic intensity in single-server Markovian queues. *Communications in Statistics-Simulation and Computation 47*, 9 (2018), 2577–2586.

[2] Armero, C., and Bayarri, M. J. Bayesian prediction in M/M/1 queues. *Queueing Systems 15*, 1-4 (1994), 401–417.

[3] Armero, C., and Bayarri, M. J. Prior assessments for prediction in queues. *Journal of the Royal Statistical Society: Series D (The Statistician) 43*, 1 (1994), 139–153.

[4] Armero, C., and Bayarri, M. J. Dealing with uncertainties in queues and networks of queues: A Bayesian approach. In *Multivariate analysis, design of experiments, and survey sampling*, S. Ghosh, Ed., CRC Press, 1999, pp. 603–632.

[5] Basak, A., and Choudhury, A. Bayesian inference and prediction in single server M/M/1 queuing model based on queue length. *Communications in Statistics - Simulation and Computation 50*, 6 (2021), 1576–1588.

[6] Basawa, I. V., Lund, R., and Bhat, U. N. Estimating function methods of inference for queueing parameters. *Lecture Notes-Monograph Series 32* (1997), 269–284.

[7] Basawa, I. V., and Prabhu, N. U. Estimation in single server queues. *Naval Research Logistics Quarterly 28*, 3 (1981), 475–487.

[8] Basawa, I. V., and Prabhu, N. U. Large sample inference from single server queues. *Queueing Systems 3*, 4 (1988), 289–304.

[9] Bhat, U. U., and Rao, S. S. A statistical technique for the control of traffic intensity in the queuing systems M/G/1 and GI/M/1. *Operations Research 20*, 5 (1972), 955–966.

[10] Brown, M., and Forsythe, A. Robust tests for the equality of variances. *Journal of the American Statistical Association 69*, 346 (1974), 364–367.

[11] Choudhury, A., and Basak, A. Statistical inference on traffic intensity in an M/M/1 queueing system. *International Journal of Management Science and Engineering Management 13*, 4 (2018), 274–279.

[12] Choudhury, A., and Borthakur, A. C. Bayesian inference and prediction in the single server Markovian queue. *Metrika 67*, 3 (2008), 371–383.

[13] Choudhury, A., and Medhi, P. Performance evaluation of a finite buffer system with varying rates of impatience. *İstatistik - Journal of The Turkish Statistical Association 6*, 1 (2013), 42–55.

[14] Chowdhury, S., and Mukherjee, S. P. Estimation of waiting time distribution in an M/M/1 queue. *Opsearch 48*, 4 (2011), 306–317.

[15] Chowdhury, S., and Mukherjee, S. P. Estimation of traffic intensity based on queue length in a single M/M/1 queue. *Communications in Statistics - Theory and Methods 42*, 13 (2013), 2376–2390.

[16] Chowdhury, S., and Mukherjee, S. P. Bayes estimation in M/M/1 queues with bivariate prior. *Journal of Statistics and Management Systems 19*, 5 (2016), 681–699.

[17] Clarke, A. B. Maximum likelihood estimates in a simple queue. *The Annals of Mathematical Statistics 28*, 4 (1957), 1036–1040.

[18] Cruz, F. R. B, Almeida, M. A. C., D'Angelo, M. F. S. V., and Woensel, T. Traffic intensity estimation in finite Markovian queueing systems. *Mathematical Problems in Engineering 2018* (2018), 018758.

[19] Cruz, F. R. B., Quinino, R. C., and Ho, L. L. Bayesian estimation of traffic intensity based on queue length in a multi-server M/M/s queue. *Communications in Statistics-Simulation and Computation 46*, 9 (2017), 7319–7331.

[20] Dave, U., and Shah, Y. K. Maximum likelihood estimates in a M/M/2 queue with heterogeneous servers. *Journal of the Operational Research Society 31*, 5 (1980), 423–426.

[21] Deepthi, V., and Jose, J. K. Bayesian estimation of an M/M/R queue with heterogeneous servers using Markov chain Monte Carlo method. *Stochastics and Quality Control 35*, 2 (2020), 57–66.

[22] Deepthi, V., and Jose, J. K. Bayesian estimation of M/Ek/1 queueing model using bivariate prior. *American Journal of Mathematical and Management Sciences 40*, 1 (2021), 88–105.

[23] Dutta, K., and Choudhury, A. Estimation of performance measures of M/M/1 queues – a simulation-based approach. *International Journal of Applied Management Science 12*, 4 (2020), 265–279.

[24] Jose, J. K., and Manoharan, M. Bayesian estimation of rate parameters of queueing models. *Journal of Probability and Statistical Science 12*, 1 (2014), 69–76.

[25] Lilliefors, H. W. Some confidence intervals for queues. *Operations Research 14*, 4 (1966), 723–727.

[26] Medhi, P. Modelling customers' impatience with discouraged arrival and retention of reneging. *Operations Research and Decisions 31*, 3 (2021), 67–88.

[27] Ogbonna, C. J., Idochi, O., and Sylvia, I. O. Effect of sample sizes on the empirical power of some tests of homogeneity of variances. *International Journal of Mathematics Trends and Technology 65*, 6 (2019), 119–134.

[28] Rohatgi, V. K., and Ehsanes Saleh, A. K. Md. *An introduction to probability and statistics*, 3rd ed., John Wiley & Sons, 2015.

[29] Rusticus, S. A., and Lovato, C. Y. Impact of sample size and variability on the power and type I error rates of equivalence tests: A simulation study. *Practical Assessment, Research, and Evaluation 19*, (2014), 11.

[30] Schruben, L., and Kulkarni, R. Some consequences of estimating parameters for the M/M/1 queue. *Operations Research Letters 1*, 2 (1982), 75–78.

[31] Shortle, J. F., Thompson, J. M., Gross, D., and Harris, C. M. *Fundamentals of qeueing theory*, 5th ed., vol. 399, John Wiley & Sons, 2018.

[32] Srinivas, V., and Kale, B. K. ML and UMVU estimation in the M/D/1 queuing system. *Communications in Statistics – Theory and Methods 45*, 19 (2016), 5826–5834.

[33] Srinivas, V., Rao, S. S., and Kale, B. K. Estimation of measures in M/M/1 queue. *Communications in Statistics – Theory and Methods 40*, 18 (2011), 3327–3336.

[34] Srinivas, V., and Udupa, H. J. Best unbiased estimation and can property in the stable M/M/1 queue. *Communications in Statistics – Theory and Methods 43*, 2 (2014), 321–327.

[35] Suyama, E., Quinino, R. C., and Cruz, F. R. B. Simple and yet efficient estimators for Markovian multiserver queues. *Mathematical Problems in Engineering 2018* (2018), 3280846.

[36] Sztrik, J. *Basic queueing theory: Foundations of system performance modeling*. GlobeEdit, 2016.

[37] Taha, H. A. *Simulation modeling and SIMNET*. Prentice Hall international series in industrial and systems engineering. Prentice Hall, Englewood Cliffs, N.J, 1988.

[38] Zheng, S., and Seila, A. F. Some well-behaved estimators for the M/M/1 queue. *Operations Research Letters 26*, 5 (2000), 231–235.

Appendix

Table A1. Summary of the results of the Brown–Forsythe test for $\lambda = 12$, $\mu = 15$ ($\rho = 0.80$)

Sample size pairs compared (n_1, n_2)	p-value	Conclusions
(25, 50) and (25, 75)	5.429e–08	We can reject our null hypothesis of variance and conclude that two sample size pairs are not equivalent.
(25, 75) and (25, 100)	0.0179	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(25, 100) and (25, 125)	0.09428	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(25, 125), (25, 150)	0.7801	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(50, 25) and (75, 25)	5.894e–08	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(75, 25) and (100, 25)	7.763e–06	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(100, 25) and (125, 25)	0.1677	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(125, 25) and (150, 25)	0.6176	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.

Table A2. Summary of results of the Brown–Forsythe test for $\lambda = 10$, $\mu = 16$ ($\rho = 0.625$)

Sample size pairs compared (n_1, n_2)	p-value	Conclusions
(25, 50) and (25, 75)	7.387e–09	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(25, 75) and (25, 100)	0.04051	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(25, 100) and (25, 125)	0.05696	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(25, 125), (25, 150)	0.6952	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(50, 25) and (75, 25)	6.88e–13	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(75, 25) and (100, 25)	6.341e–05	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(100, 25) and (125, 25)	0.1407	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(125, 25) and (150, 25)	0.5794	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.

Table A3. Summary of results of the Brown–Forsythe test for λ = 16, μ=18 (ρ = 0.88)

Sample size pairs compared (n_1, n_2)	p-value	Conclusions
(25, 50) and (25, 75)	6.448e–09	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(25, 75) and (25, 100)	0.002814	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(25, 100) and (25, 125)	0.1815	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(25, 125), (25, 150)	0.7496	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(50, 25) and (75, 25)	2.437e–10	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(75, 25) and (100, 25)	0.00057	We can reject our null hypothesis of homogeneity of variance and conclude that two sample size pairs are not equivalent.
(100, 25) and (125, 25)	0.2155	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.
(125, 25) and (150, 25)	0.1414	We can accept our null hypothesis of homogeneity of variance and conclude that two sample size pairs are equivalent.