

**NEGATIVE BINOMIAL DISTRIBUTION AS AN ARRIVAL PATTERN
TO A QUEUEING SYSTEM: A DATA-DRIVEN APPROACH**

Ashwini K¹, Sumathi Prabhu^{2*}

Department of Mathematics, Manipal Institute of Technology, Manipal Academy of Higher
Education, Manipal - 576104, India

E-mail: ¹ashwinikbhat08@gmail.com

²k.sumathi@manipal.edu (*Corresponding author)

Abstract

Queues form an integral part of daily life. Several studies have been done for better performance of systems to avoid queues or the waiting time of customers. The two important factors to be noted are the arrival pattern of customers and the service rate of the system. These are interrelated and are probabilistic. The literature on queueing theory considers Poisson distribution for the arrival pattern. The present paper focuses on observing the arrival pattern of students to a college library. Binomial, Poisson and negative binomial distributions are fitted to the data and has been found that the negative binomial distribution is a good fit to the data. Interarrival times are also analyzed.

Math. Subj. Classification 2020: 60K25, 60K30

Key Words and Phrases: Arrival pattern, Data-driven approach, Goodness of fit test, Interarrival time, Queue.

1. Introduction

In the modern era of advanced technology, empirical studies play a very important role in the design of any working system where queues are formed to meet the demand of the customers. Queues are common in any public or private sector such as hospitals, banks, government offices, malls and restaurants. They are formed where there is high demand for services. Extensive literature exists on queueing theory where the Poisson distribution is used for the arrival pattern and exponential for the interarrival time or service pattern. Good references for this are Gross et al. [6] and Bhat [2].

Recent works on the application of queueing models in health sector are Cho et al. [4], Aziati and Hamdan [1], Haque et al. [7], Zimmerman et al. [14], Sülz et al. [12], and Umar et al. [13]. The others include Siwach et al. [10] who studied queueing models for the data in a sugar mill and Dbeis and Al-Sahili [5] who analyzed the reneging behavior of customers in a drive-thru system. The arrival distribution considered in these works is Poisson.

In the era of modern technology, the evidence based, or data-driven studies are gaining prominence. Here, sampling forms a crucial part of any system to be designed and is less explored domain in queueing systems. Thus, with the intention of comparing theoretical work with real life system, a data-driven approach was considered. Triggered by the thought that the

existence of college libraries is redundant as the study material is available in the public domain, it was decided to observe the pattern of the students arriving at the college library of a higher education institution. A recent work on the application of queueing models in the library is that by Lyu et al. [9]. The arrival pattern used here is also Poisson.

The present paper builds a probability model for the arrival pattern of the students visiting the college library. The flow of the paper is as follows. Section 2 provides a description of the problem, followed by a preliminary analysis of the data. Section 3 describes the fitting of the distributions to the arrival pattern and testing for its goodness of fit. The interarrival time of the students is also studied and is given in section 4. The paper concludes with the findings of the study in section 5.

2. Problem Description

Library management is a less explored area in queueing theory. It should be noted that this also is a system where queueing theory can be applied for smooth functioning, with an optimal number of servers as most of the study materials are available online, thus making the existence of the libraries redundant.

The present study is based on the arrivals to the lending section of the library of an engineering college in Udupi district, Karnataka State, India. It is an institute with a high reputation. The library is equipped with all modern facilities for access to standard textbooks and journals.

The lending section of the library provides the students with access to textbooks, journals and periodicals. The students can borrow textbooks for a week or two or can use them for reference in the lending section itself. It is mandatory that the student possesses his/her identity card issued by the college authorities, at the time of entry to the library. The details on the identity card are recorded electronically along with the date and time of arrival to the lending section.

The students entering the lending section are of three categories, *viz.*, undergraduates, postgraduates and research scholars. The objective of the study is to determine an appropriate probability distribution to the arrival pattern of the students. With this information, interarrival time is also analyzed.

2.1 Preliminary Analysis of the data

The data used for the study corresponds to eight continuous working days of the college library. The collection can be done in two ways. The first method is to observe the system until a specific number of arrivals occurs and the second method is to observe the system for a fixed interval of time and note down the number of arrivals (Gross et al. [6], Bhat [2]). In the first method, the number of arrivals is not a random variable since it is a prefixed value and in the second method, number of arrivals is a random variable since the time interval is fixed. The present study considers the second method since our interest is to estimate the distribution corresponding to the arrival pattern.

The arrival to the lending section of the college library from 09:00 AM-05:00 PM is noted. The number of arrivals on the eight days under consideration are 93, 49, 67, 69, 72, 22, 143 and 152 respectively. Thus, the total number of arrivals was 667. Figure 1 represents the bar diagram of number of arrival of students over eight days of observation.

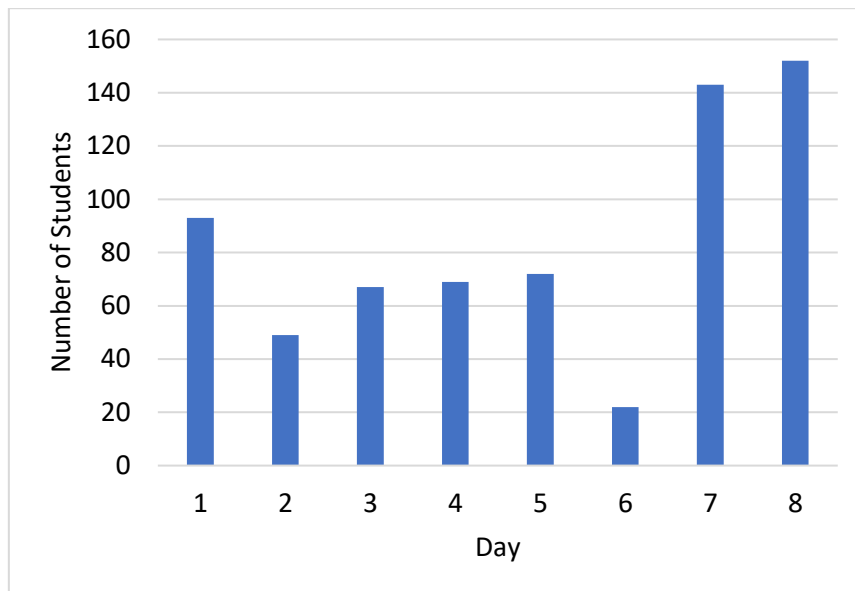


Figure 1. Bar chart representing the number of arrivals for the eight working days

From Figure 1, it can be observed that the number of students arriving on the 6th day is very low. It was 22 and the day was a Saturday. The highest was on Monday. The numbers are 93 and 143. Thus, the average number arriving per day is 83.375 and 10.42 per hour. The data recorded is for eight hours each day.

3. Distribution of the arrivals

Since the number of arrivals over a period is a counting process, the distributions considered were binomial, Poisson and negative binomial. Hence, to check whether each of these distributions is a good fit or not, the frequency distribution of the arrivals over the total period observed was formed. The observation was eight hours each day. The number of days observed was eight. Thus, the total number of hours observed was forty-eight.

Tables 1 and 2 show the frequency distribution of arrivals corresponding to 10 minutes and 15 minutes of the total period of observation. It is seen from table 1 that there were no arrivals in 114 10-minute intervals, one arrival in 103 10-minute intervals and so on. Similarly, from table 2, there were zero arrivals in 53 15-minute intervals, one in 52 15-minute intervals, and so on.

Note that $\sum X_i O_i = 667$, the total number of arrivals, in both tables 4 and 5, where X_i is the number of arrivals per unit of time and O_i is the number of intervals having X_i arrivals. The average number of arrivals is 1.737 per 10 minutes with a variance 3.3345, while the

corresponding figures are respectively, 2.6055 and 6.5201 for a 15-minute interval. Thus, the average number of arrivals is 10.42 per hour in both cases.

3.1 Preliminaries of the distributions

An experiment with two and only two possible outcomes is a Bernoulli trial. A random variable X is said to follow a Bernoulli distribution with parameter p if

$$X = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p \end{cases} \quad \text{and } 0 \leq p \leq 1.$$

$X = 1$ may be termed a success and $X = 0$, a failure.

The total number of successes in a fixed number of Bernoulli trials is a binomial distribution. Suppose the interest is in finding the number of Bernoulli trials required to get a fixed number of successes. Then, this formulation leads to negative binomial distribution. Thus, in a sequence of independent Bernoulli (p) trials, if the random variable X denotes the trial at which the r^{th} success occurs, r is a fixed integer, then X has a negative binomial distribution with parameters r and p .

The probability mass functions (p. m. f) of the binomial, Poisson and negative binomial distribution are shown in Table 3. In the present study, the random variable X denotes the number of arrivals in a fixed interval of time and the success is the arrival of the student to the library.

3.2 Fitting of the distribution and testing for its Goodness of fit

Since the arrival process is a counting process, the discrete distributions binomial, Poisson and negative binomial were fitted to the data. This was done by first estimating the parameters of the distributions. The estimation was done by equating the observed moments, viz., the mean and the variance, to the theoretical moments, and then solving the equations. This method is also known as the method of moments. The computed values are shown in Table 4. Then, using the recurrence relations given in Table 3, the probabilities and the expected frequencies corresponding to the three distributions were obtained. The expected frequencies E_i or the hypothetical frequencies are shown in Tables 1 and 2.

Testing for the goodness of fit was done using the chi-squared test. The null hypothesis is H_0 : The theoretical distribution is a good fit to the observed frequency distribution. The alternative hypothesis need not be specified as this is a significance test. The test statistic is

$$\chi^2_{calc} = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

The test procedure is to reject the null hypothesis at the specified level of significance (L.O.S.) if the calculated value of the test statistic is greater than the threshold value $\chi^2_{critical}$. This value is also known as the critical value and is obtained from the chi-squared tables. Alternatively, the size of the test, i.e., p-value, may also be obtained. Here, the test procedure

is to reject the null hypothesis if the p-value is less than the L.O.S. In this study, the L.O.S. considered is 5%.

While calculating the test statistic, the adjacent classes will be pooled if the expected frequencies E_i of any of the classes is less than 5. The corresponding changes are reflected in the observed frequencies O_i too. Thus, the degrees of freedom (d.f.) will be equal to the number of classes after pooling minus the number of estimated parameters of the distribution.

The results of χ^2 test for both 10 minutes and 15 minutes intervals are shown in Table 4. MS-Excel 2019 and R software are used for calculations. From Table 4, it is clear that $\chi^2_{calc} > \chi^2_{critical}$, or p-value < L.O.S = 0.05, both in the case of 10 minutes and 15 minutes interval for Binomial and Poisson distributions, indicating that these distributions are not a good fit to the data. But this is the reverse in case of negative binomial accepting it as a good fit. This is also visible from the graphs shown in Figures 2 and 3.

Table 1. Frequency Distribution of Number of Arrivals in 10 minutes interval

Number (X_i) of Arrivals per 10 Minutes	Observed Frequency (O_i) (Number of 10-minute intervals)	Expected Frequency (E_i) (Number of 10-minute intervals)		
		Binomial	Poisson	Negative Binomial
0	114	59	68	112
1	103	119	117	101
2	62	111	102	70
3	43	63	59	44
4	32	24	26	26
5	15	06	09	14
6	06	01	03	08
7	05	00	01	04
8	02	00	00	02
9	01	00	00	01
10	00	00	00	01
11	00	00	00	00
12	01	00	00	00
More than 12	00	00	00	00
Total	384	384	384	384

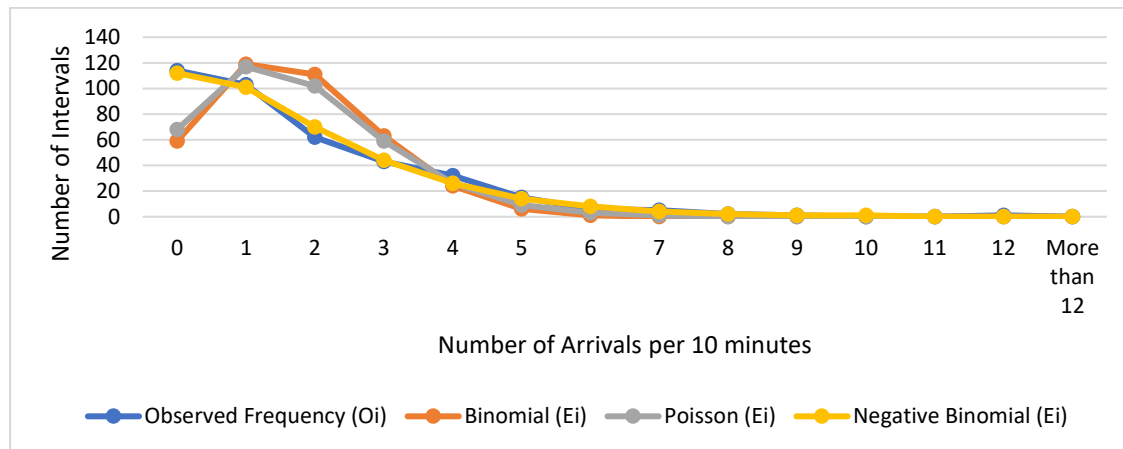


Figure 2. Observed and Expected Frequency in 10-minutes intervals

Table 2. Frequency Distribution of Number of Arrivals in 15 minutes interval

Number (X_i) of Arrivals per 15 Minutes	Observed Frequency (O_i) (Number of 15-minute intervals)	Expected Frequency (E_i) (Number of 15-minute intervals)		
		Binomial	Poisson	Negative Binomial
0	53	15	19	52
1	52	46	49	54
2	43	68	64	45
3	33	61	56	33
4	31	39	36	24
5	14	18	19	16
6	12	06	08	11
7	08	02	03	07
8	01	00	01	05
9	02	00	00	03
10	03	00	00	02
11	01	00	00	01
12	01	00	00	01
13	01	00	00	01
14	00	00	00	00
15	00	00	00	00
16	01	00	00	00

More than 16	00	00	00	00
Total	256	256	256	256

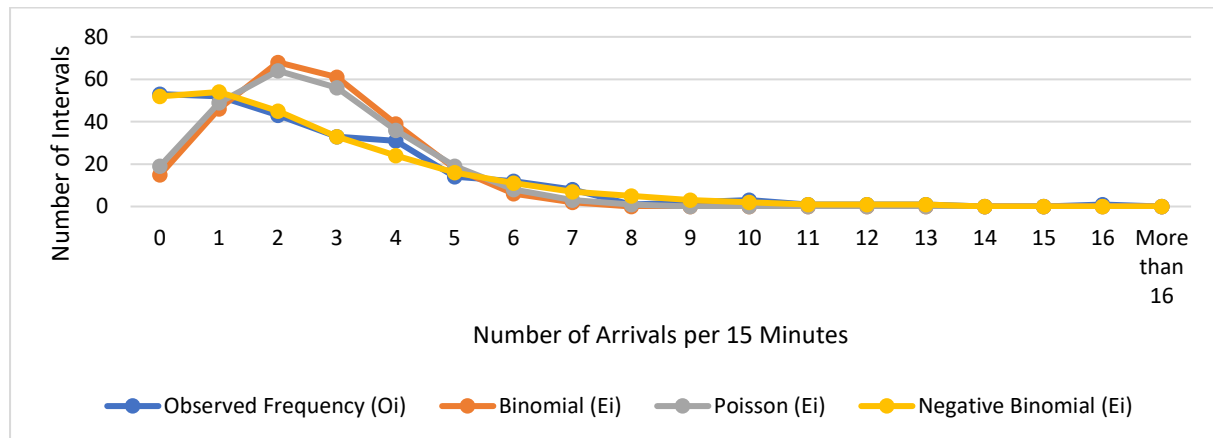


Figure 3. Observed and Expected Frequency in 15-minutes intervals

Table 3. Discrete Distributions

Distribution	Binomial	Poisson	Negative Binomial
p.m.f.	$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \dots, n; 0 < p < 1$	$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots; \lambda > 0$	$p(x) = \binom{r+x-1}{x} p^r (1-p)^x, x = 0, 1, 2, \dots; r > 1; 0 < p < 1$
Parameter/s	n, p	λ	r, p
Mean	np	λ	$\frac{r(1-p)}{p}$
Variance	$np(1-p)$	λ	$\frac{r(1-p)}{p^2}$
Recurrence relation	$p(x) = \frac{(n-x+1)p}{x(1-p)} p(x-1), x = 1, 2, \dots, n; p(0) = (1-p)^n$	$p(x) = \frac{\lambda}{x} p(x-1), x = 1, 2, \dots; p(0) = e^{-\lambda}$	$p(x+1) = \frac{(x+r)}{(x+1)} (1-p) p(x), x = 0, 1, 2, \dots; p(0) = p^r$

Table 4. Estimated parameters and results of goodness-of-fit test

	Binomial Distribution	Poisson Distribution	Negative Binomial Distribution

	10 minutes interval	15 minutes interval	10 minutes interval	15 minutes interval	10 minutes interval	15 minutes interval
Estimated parameter	$\hat{p}=0.1448$	$\hat{p}=0.1628$	$\hat{\lambda}=1.737$	$\hat{\lambda}=2.6055$	$\hat{r}=1.8845$ $\hat{p}=0.5209$	$\hat{r}=1.7341$ $\hat{p}=0.3996$
χ^2_{calc}	145.8678	174.5278	80.7110	103.7015	3.1564	6.3626
d.f.	4	5	4	5	5	7
$\chi^2_{critical}$	9.49	11.10	9.49	11.10	11.10	14.10
p-value	1.56×10^{-30}	7.89×10^{-36}	1.23×10^{-16}	8.76×10^{-21}	0.6759	0.4981

4. Interarrival time analysis

Inter arrival time is the time between the successive arrival of customers to a system. The available data has the time of arrival of each student to the library. Inter arrival time is computed from the time of arrival by subtracting the consecutive time of arrival. It is known that, interarrival time is the reciprocal of arrival rate ([6], [2]). i.e., *Interarrival time* = $1/\text{arrival rate}$. The day-wise average arrivals are computed and are compared with corresponding day-wise average interarrival times. The results are presented in Table 5.

Table 5. Average arrival and interarrival time

Day	Average Arrival (per hour)	Average Time (in minutes)	Average Interarrival Time (in minutes)
1	11.625	5.16	0:05:01
2	6.125	9.79	0:09:24
3	8.375	7.16	0:07:02
4	8.625	6.96	0:06:45
5	9	6.67	0:06:38
6	2.75	21.82	0:21:17
7	17.875	3.36	0:03:16
8	19	3.16	0:03:04

The average time per arrival is almost the same as average interarrival time. Hence, the above-mentioned relation between these two is verified.

Since interarrival times are continuous, continuous probability distributions are tested to identify the suitable distribution. The analysis has been done with the help of the trial version of Minitab software. This reports the results of Anderson-Darling (AD) test and corresponding p-values for the specified distributions. The probability density functions of the continuous distributions used are as follows.

- Normal distribution: $f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right), -\infty < y < \infty; \sigma > 0$
- Lognormal distribution: $f(y) = \frac{1}{\sigma y\sqrt{2\pi}} \exp\left(\frac{-(\ln y - \mu)^2}{2\sigma^2}\right), y > 0; -\infty < \mu < \infty; \sigma > 0$
- Exponential distribution: $f(y) = \frac{1}{\sigma} \exp\left(\frac{-y}{\sigma}\right), y \geq 0, \sigma > 0$
- 2-parameter exponential distribution: $f(y) = \frac{1}{\sigma} \exp\left(\frac{-(y-\theta)}{\sigma}\right), y \geq \theta, \sigma > 0$
- Weibull distribution: $f(y) = \frac{\alpha}{\sigma} \left(\frac{y}{\sigma}\right)^{(\alpha-1)} \exp\left(-\left(\frac{y}{\sigma}\right)^\alpha\right), y \geq 0; \alpha, \sigma > 0$
- 3-parameter Weibull distribution: $f(y) = \frac{\alpha}{\sigma} \left(\frac{y-\mu}{\sigma}\right)^{(\alpha-1)} \exp\left(-\left(\frac{y-\mu}{\sigma}\right)^\alpha\right), y \geq \mu; \alpha, \sigma > 0$
- Smallest extreme value distribution: $f(y) = \frac{1}{\sigma} \exp\left(\frac{y-\mu}{\sigma}\right) \exp\left(-e^{\frac{y-\mu}{\sigma}}\right), -\infty < y < \infty; \sigma > 0$
- Largest extreme value distribution:

$$f(y) = \frac{1}{\sigma} \exp\left(\frac{-(y-\mu)}{\sigma}\right) \exp\left(-e^{-\frac{y-\mu}{\sigma}}\right), -\infty < y < \infty; \sigma > 0$$
- Gamma distribution: $f(y) = \frac{y^{\alpha-1}}{(\alpha-1)!\sigma^\alpha} \exp\left(\frac{-y}{\sigma}\right), y > 0; \alpha, \theta > 0$

In these distributions, Y denotes the interarrival time and the constants μ, σ, α and θ refer to the location, scale, shape and threshold parameters respectively (Johnson et al. [8]).

The AD test is used to test if the data comes from a specific distribution. The hypothesis used in this test is H_0 : The sample data follows the hypothesized distribution. The test statistic is defined as $A^2 = -N - S$, where $S = \sum_{i=1}^N \frac{(2i-1)}{N} [\ln F(Y_i) + \ln(1 - F(Y_{N+1-i}))]$, F is the cumulative distribution function of the specified distribution (Stephens [11]) and N represents the number of observations.

The decision was made by comparing the p-values with the L.O.S. The AD test is carried out for day-wise data and the results obtained are presented in Table 6.

Table 6. Distribution fitting summary for interarrival times

Distribution	Day-1	Day-2	Day-3	Day-4
--------------	-------	-------	-------	-------

	AD	p-value	AD	p-value	AD	p-value	AD	p-value
Normal	7.010	<0.005	3.049	<0.005	6.765	<0.005	3.393	<0.005
Lognormal	1.379	<0.005	1.253	<0.005	0.928	0.017	1.674	<0.005
Exponential	0.793	0.209	1.942	0.010	6.256	<0.003	0.394	0.648
2-Parameter Exponential	0.777	0.202	1.304	0.039	5.375	<0.010	0.279	>0.250
Weibull	0.323	>0.250	0.349	>0.250	0.235	>0.250	0.258	>0.250
3-Parameter Weibull	0.332	>0.500	0.340	>0.500	0.288	>0.500	0.224	>0.500
Smallest Extreme Value	11.555	<0.010	4.467	<0.010	10.667	<0.010	5.605	<0.010
Largest Extreme Value	2.917	<0.010	1.841	<0.010	3.716	<0.010	1.574	<0.010
Gamma	0.372	>0.250	0.244	>0.250	0.330	>0.250	0.232	>0.250
Distribution	Day-5		Day-6		Day-7		Day-8	
	AD	p-value	AD	p-value	AD	p-value	AD	p-value
Normal	6.214	<0.005	0.883	0.020	12.375	<0.005	14.397	<0.005
Lognormal	1.861	<0.005	1.222	<0.005	0.870	0.025	0.773	0.044
Exponential	1.561	0.026	2.280	0.004	4.154	<0.003	4.839	<0.003
2-Parameter Exponential	1.312	0.042	1.386	0.021	4.593	<0.010	5.538	<0.010
Weibull	0.492	0.223	1.155	<0.010	1.119	<0.010	0.821	0.034
3-Parameter Weibull	0.526	0.191	1.212	<0.005	0.841	0.033	0.474	0.248
Smallest Extreme Value	12.191	<0.010	0.906	0.019	22.697	<0.010	23.320	<0.010

Largest Extreme Value	1.778	<0.010	1.124	<0.010	6.983	<0.010	7.717	<0.010
Gamma	0.445	>0.250	0.998	0.018	1.562	<0.005	1.380	<0.005

The table shows that gamma followed by 3-parameter Weibull distributions are a good fit to the data.

5. Conclusions

The Poisson distribution is frequently used to model arrivals, as evidenced by previous literature. Service times are assumed to be exponential. The system was considered to be in steady-state with an arrival rate of 10.42 per hour and a service rate of 11 per hour. Initially, a Markovian queue with a single server was developed. Students were discovered to be waiting for longer than an hour on average. This looks erroneous because it is quite improbable to occur in a library. This motivated us to validate the Poisson assumption for arrivals. As a result, the goodness of fit test for arrivals, for different discrete probability distributions was examined.

From the present study, it is found that negative binomial distribution can be considered as an alternative for Poisson for the arrival pattern of customers to a queueing system. When the arrival pattern is Poisson, it has been proved theoretically that the interarrival time is exponential. In the present study, it is observed that the 3-parameter Weibull is a good fit for the interarrival time distribution. The next alternative is gamma.

It must be noted that the negative binomial, gamma and the 3-parameter Weibull distributions belong to the exponential family just as the Poisson and exponential. Thus, further work in this direction is in progress so that the results can be generalized.

Data availability statement

As further analyses are being carried out on the data, the data can be provided by the corresponding author to interested learners based on reasonable request.

References

- [1] A.N. Aziati and N.S.B. Hamdan, Application of Queuing Theory Model and Simulation to Patient Flow at the Outpatient Department, In *Proceedings of the International Conference on Industrial Engineering and Operations Management Bandung, Indonesia*, (2018), 3016-3028.
- [2] U.N. Bhat, *An Introduction to Queueing Theory (Second Edition)*. Springer Science + Business Media, New York (2015).
- [3] G. Casella and R.L. Berger, *Statistical Inference (Second Edition)*. Duxbury Pacific Grove, CA (2002).

- [4] K.W. Cho, S.M. Kim, Y.M. Chae and Y.U. Song, Application of Queueing Theory to the Analysis of Changes in Outpatients' Waiting Times in Hospitals Introducing EMR. *Healthcare Informatics Research*, **23**, No 1 (2017), 35-42.
- [5] A. Dbeis and K. Al-Sahili, Enhancing queueing theory realism: analysis of reneging behavior impact on M/M/1 drive-thru service system. *Journal of Management Analytics*, (2024), 1–16.
- [6] D. Gross, J.F. Shortle, J.M. Thompson and C.M. Harris, *Fundamentals of Queueing Theory (Fourth Edition)*. Wiley Series in Probability and Statistics, New Delhi (2008).
- [7] M.O. Haque, S. Akter, A. Hossen, and Z. Rahman, A Case Study on Outpatient Waiting Time for Treatment with Single Servers Queueing Model at Public Eye Hospital in Bangladesh. *American Scientific Research Journal for Engineering, Technology, and Sciences*, **68**, No 1 (2020), 143–151.
- [8] N.L. Johnson, S. Kotz and N. Balakrishnan, *Continuous Univariate Distributions, Volumes I and II, 2nd. Ed.*, John Wiley and Sons (1994).
- [9] X. Lyu, F. Xiao and X. Fan, Application of Queueing Model in Library Service. *Procedia Computer Science*, **188**, (2021), 69-77.
- [10] V. Siwach, M.S. Tonk and H. Poonia, Simulation of Queues in Sugar Mills Using Monte Carlo Technique. In: Sharma, R. K., Pareschi, L., Atangana, A., Sahoo, B., Kukreja, V. K. (eds.), *Frontiers in Industrial and Applied Mathematics. FIAM 2021*. Springer Proceedings in Mathematics & Statistics, (2023), 410.
- [11] M.A. Stephens, EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, **69**, (1974), 730-737.
- [12] S. Sülz, A. Fügner, M. Becker-Peth and B. Roth, The potential of patient-based nurse staffing – a queueing theory application in the neonatal intensive care setting. *Health Care Management Science*, **27**, (2024), 239–253.
- [13] M.I. Umar, M.U. Adehi and A.M. Auwal, An Investigation of Multi-server Queueing Analysis to Assess Hospital Healthcare Systems' Operational Effectiveness. *Asian Journal of Probability and Statistics*, **26**, No 3 (2024), 80–89.
- [14] S.L. Zimmerman, A.R. Rutherford, A. van der Waall, M. Norena and P. Dodek, A queueing model for ventilator capacity management during the COVID-19 pandemic. *Health Care Management Science*, **26**, (2023), 200–216.