Application of Queuing Theory to Health Care Delivery

N. M. Eze[1] ; W. B. Yahya[2]

[1]Department of Statistics,
University of Nigeria, Nigeria.

[2]Department of Statistics,
University of Ilorin, Ilorin, Nigeria.
E-mail: nnaemeka.eze@unn.edu.ng[1]

***Abstract* — This study analyzed the Queuing system experienced at healthcare delivery center in Nigeria. The primary goal of this study is to ascertain the congestion rate of patients waiting in line at the aforementioned center and to proffer solution(s) to handle this congestion and make recommendation(s) to enhance the management of the system. The queuing model used in this study is M/M/2 model. Where the first and second M are the inter-arrival and service time respectively and 2 is the number of server. Chi-square distribution was used to test for the Goodness of fit of the inter-arrival and service time and the results showed that both were exponentially distributed. In order to ascertain whether the system is efficient, primary data were collected, analysed, and interpreted. The researcher obtained the mean inter-arrival time (λ) = 0.1745mins and the mean service time (μ) = 0.1677mins. Whereas the measure of proportion between λ and μ which is known as the traffic intensity (ρ) was obtained to be 0.52. Since traffic intensity is less than 1, it indicates that the system is efficient.**

Keywords: *Queue, Goodness-of-fit, Inter-arrival Time, Service Time, Traffic Intensity.*

I. Introduction

Waiting in lines or "queues" seems to be a general observable fact or occurrence in our day to day life. Think of the many times one has to wait in line to get serviced in a facility, be it a booking counter, grocery store checkout, hospital, banks, pharmacy, petrol stations, cafeteria, bus terminals just to mention but a few. The frustration associated with those waits can really be unbearable, especially when there is a long waiting line. "Queue" is a word borrowed straight from French and the French word queue originally comes from the Latin word *coda which* means "tail" and this is the original meaning of the word "queue". Queue is defined as a line or sequence of animate or inanimate entity awaiting their turn to receive service or to proceed. "Queues are formed when the resources are limited" (J.E Beasley 2010). A queue is said to be formed, when the arrival rate of customers at a facility exceeds the processing system rate. Arrival may come in singly or in batches; they may come in consistent spaced or in a random order. The origin of queues started from A.K. Erlang in the year 1909, where he published his first book on queueing theory. Erlang was a Danish engineer who worked for Copenhagen Telephone Exchange in Denmark. He established a conceptual and methodological framework of queueing theory for application to telephone traffic which can be regarded as a precursor of much modern theory of stochastic processes. This simple but physical realistic formulation has provided the benchmark for this theory. "In queueing theory, we study situations where units of some kind arrive at a service facility for receiving service of some description, some of the units have to wait for service and depart after service" (kashyap and Chaudhry). Queuing theory has many applications and has been used extensively by service industries because the results are often used when making decisions about the resources needed to provide services to customers.

Long waiting time in any hospital is considered as an indicator of poor quality in service and needs improvement. This will reduce customer's demand and eventually revenue and profit (Biju, Naeema and Faisal, 2011). So, the purpose of this study is to give a general background into queueing theory concentrating basically on hospital management with respect to health care delivery system so as to proffer solutions to help decongest waiting line in the hospital.

II. Literature Review

Erlang (1909) studied a conceptual and methodological framework of queueing theory for application to telephone traffic. In his work, he showed that the number of calls during an arbitrary time interval, assuming calls originate at random, follows a poison law and the intervals between calls were then exponentially distributed.

Longenecker (1969) stated that waiting lines form when number of customers are more than the facilities in the system and suggested that management team must determine the customer arrival rate and their service time in order to achieve balance between cost of service and cost associated with waiting for service.

Karlin and Taylor (1981) categorized the queuing process into 3; input process, queuing discipline and service distribution. They were of the opinion that if arrivals of customers follow a Poisson distribution, it implies that the inter-arrival time and service time follows exponential distribution.

Dhanavatan (2010) studied queues at ATM at a bank for two months during weekdays busy and free hours and also weekends busy and free hours and observed that generally, arrivals do not occur at a fixed regular interval of time but tends to be clustered for a duration of a week. Ogini (2014) studied the queuing system at First bank Nigeria plc, Laspotech, Isolo Campus Branch, Lagos state. The queuing system analyzed in his study was that of M/M/1 model, the exponential inter-arrival time (λ) and exponential service time (μ) were computed. The result indicated that there was a waiting line problem, implying that customers are on the average arriving at a faster rate than they are being served.

III. Materials And Methods

A. Data Description

The data were collected primarily from Royal Cross Hospital and Maternity in Enugu, Nigeria. The arrival time, the time service starts and the time service ends of each patient were obtained respectively. The data were collected for a period of ten (10) clinic days. The arrival time is the time the patient stands before the front desk and the staff on duty writes down the patient's name and issues a tally to the patient.
Moreover, the inter-arrival times were simply obtained from subtracting successive arrival times while waiting times were obtained from the difference between the arrival time and the time service ended of each patient and the service times were obtained from the time taken to attend to a patient. The unit of time was recorded in "minutes".

B. Methodology

In this study, M/M/2 model was used to analyze the queuing system
Where

- The first M is the Exponential distribution of inter-arrival times, with parameter λ.
- The second M is the Exponential distribution of service times, with parameter μ.
- 2 is the number of server.

The study is aimed at adopting statistical measures to find out the efficiency of the service delivery to patients that visit antenatal clinics in Nigeria and the objectives of the study are to:

- fit a distribution of the inter-arrival time of patients.
- fit a distribution of the service time of patients.
- determine the traffic intensity while checking for the efficiency of the system.
- to compute some performance measures.

Below are the basic formulas used in carrying out this analysis:

1. Mean arrival rate, $\lambda = \frac{1}{\bar{X}}$ 1
$$where\ \bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$
2. Mean service rate, $\mu = \frac{1}{\bar{Y}}$ 2
$$where\ \bar{Y} = \frac{\sum_{i=1}^{n} f_i y_i}{\sum_{i=1}^{n} f_i}$$
3. Traffic Intensity $\rho = \frac{\lambda}{S\mu}$ 3
where s = number of server = 2
4. Probability of the system being empty:
5. $$P_0 = \left[\sum_{n=0}^{s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{s^s}{S!}\frac{\rho^s}{1-\rho}\right]^{-1} \quad 4$$
$$= \left[\sum_{n=0}^{s-1} \frac{1}{n!}(s\rho)^n + \frac{s^s}{S!}\frac{\rho^s}{1-\rho}\right]^{-1} \quad 5$$
6. The expected number of patients in the queue:
$$E(N_q) = \frac{\rho(\rho s)^s P_0}{(1-\rho)^2 s!} \quad 6$$
7. The expected waiting time in the queue:
$$E(W_q) = \frac{(\rho s)^s P_0}{s!(\mu s)(1-\rho)^2} \quad 7$$

IV. Data Analysis And Results

A. Fitting A Distribution to the Inter-Arrival Time of Patients

First we obtain the class interval for this distribution:
Class size= $\frac{\text{Range}}{\text{Number of classes}} = \frac{24-1}{6} = 3.8 \cong 4$

Table 1: Frequency table for inter-arrival time

Inter-arrival time (mins)	Mid-point (x_i)	Observed frequency $O_i = f_i$	$f_i x_i$
$0<t\le4$	2	156	312
$4<t\le8$	6	89	534
$8<t\le12$	10	42	420
$12<t\le16$	14	28	392
$16<t\le20$	18	8	144
$20<t\le24$	22	3	66
Total		326	1868

From the table above, the mean is given as;

$$\bar{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} = \frac{1868}{326} = 5.7301$$

Hence, $\lambda = \frac{1}{\bar{X}} = \frac{1}{5.7301} = 0.1745$

Test of hypothesis
H_0: The inter-arrival time follows an exponential distribution.
H_1: The inter-arrival time does not follow an exponential distribution.
$\alpha = 0.05$
Test Statistic:

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{n-k-1}$$

Decision rule: Reject H_0 if P-value < α, do not reject if otherwise.
Calculations: According to the hypothesis stated above, an exponential
distribution with parameter λ has its pdf as

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \ge 0 \\ 0, & \text{if otherwise} \end{cases}$$

The Probabilities, P_i, are obtained thus:

$$P_i = P(a < t \le b) = \int_a^b f(t)dt = \int_a^b \lambda e^{-\lambda(t)}dt$$

$$= \lambda \int_a^b e^{-\lambda(t)}dt = \frac{-\lambda}{\lambda} e^{-\lambda(t)} \Big|_a^b = \left(e^{-\lambda a} - e^{-\lambda b}\right)$$

Where i = 1,2,3,4,5,6
Such that the probability for each interval is calculated below
The interval between 0 and 4
$P_1 = P(0 < t \le 4) = e^{-0.1745(0)} - e^{-0.1745(4)} = 1 - 0.4976 = 0.5024$
.
.
.
The interval between 20 and 24
$P_6 = P(20 < t \le 24) = e^{-0.1745(20)} - e^{-0.1745(24)} = 0.0305 - 0.0152 = 0.0153$

Given that N (total number of observation) = 326 the expected frequencies ($e_i's$) for each class becomes:
$e_1 = NP_1 = 326 * 0.5024 = 163.7824$
.
.
.
$e_6 = NP_6 = 326 * 0.0153 = 4.9878.$

Table 2: The expected frequency table for inter-arrival time.

Inter-arrival time in (mins)	Observed frequency O_i	Probability P_i	Expected frequency $e_i = NP_i$	$\frac{(O_i - e_i)^2}{e_i}$
0<t≤4	156	0.5024	163.7824	0.3698
4<t≤8	89	0.2499	81.4674	0.6965
8<t≤12	42	0.1244	40.5544	0.0515
12<t≤16	28	0.0619	20.1794	3.0309
16<t≤20	8 } 11	0.0308	10.0408 } 15.03	1.0799
20<t≤24	3	0.0153	4.9878	
Total	**N = 326**	**0.9847**		**5.2286**

Therefore, $\chi^2_{cal} = 5.2286$ and P-value(5.2286, 3) = 0.1558 where 3 is obtained from n-k-1 (i.e., 5-1-1)

Conclusion: Since P-value = 0.1558 > α = 0.05, we do not have sufficient evidence to reject H_0. So, the researcher accepts the null hypothesis and conclude that the inter-arrival time of patients follows an exponential distribution at α = 0.05

A. Fitting a Distribution of The Service Time of Patients

We obtain the class interval for this distribution:

Class size = $\frac{\text{Range}}{\text{Number of classes}} = \frac{24-1}{5} = 4.6 \cong 5$

Table 3: Frequency table for service time.

Service time (mins)	Mid-point y_i	Observed frequency $O_i = f_i$	$f_i y_i$
0<t≤5	2.5	182	455
5<t≤10	7.5	98	735
10<t≤15	12.5	35	437.5
15<t≤20	17.5	19	332.5
20<t≤24	21.5	2	43
Total		336	2003

From the table above, the mean is given as;

$$\overline{Y} = \frac{\sum_{i=1}^{n} f_i y_i}{\sum_{i=1}^{n} f_i} = \frac{2003}{336} = 5.9613$$

Hence, $\mu = \frac{1}{\overline{Y}} = \frac{1}{5.9613} = 0.1677$

Test of hypothesis

H_0: The service time follows an exponential distribution.

H_1: The service time does not follow an exponential distribution.

$\alpha = 0.05$

Test Statistic:

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{n-k-1}$$

Decision rule: Reject H_0 if P-value < α, do not reject if otherwise.

Calculations: According to the hypothesis stated above, an exponential distribution with parameter μ has its pdf as

$$f(t) = \begin{cases} \mu e^{-\mu t}, & t \geq 0 \\ 0, & \text{if otherwise} \end{cases}$$

The Probabilities, P_i, are obtained thus:

$$P_i = P(a < t \leq b) = \int_a^b f(t)dt = \int_a^b \mu e^{-\mu(t)}dt$$

$$= \mu \int_a^b e^{-\mu(t)}dt = \frac{-\mu}{\mu} e^{-\mu(t)} \Big|_a^b = (e^{-\mu a} - e^{-\mu b})$$

Where i = 1,2,3,4,5

Such that the probability for each interval is calculated below

The interval between 0 and 5

$P_1 = P(0 < t \leq 5) = e^{-0.1677(0)} - e^{-0.1677(5)} = 1 - 0.4324 = 0.5676$

.

.

.

$P_5 = P(20 < t \leq 25) = e^{-0.1745(20)} - e^{-0.1745(25)} = 0.0349 - 0.0151 = 0.0198$ Given that N (total number of observation) = 336 the expected frequencies ($e_i's$) for each class becomes:

$e_1 = NP_1 = 336 * 0.5676 = 163.7824$

.

.

.

e_5 $e_1 = NP_5 =$

336* 0.0308 = 6.6528

Table 4: The expected frequency table for service time

Service time (mins)	**Observed frequency O_i**	**Probability P_i**	**Expected frequency** $e_i = NP_i$	$\frac{(O_i - e_i)^2}{e_i}$
0<t≤5	182	0.5676	190.7136	0.3981
5<t≤10	98	0.2455	82.488	2.9171
10<t≤15	35	0.1061	35.6496	0.0118
15<t≤20	19	0.0459	15.4224	0.8299
20<t≤24	2	0.0198	6.6528	3.2541
Total	**336**	**0.9849**	**330.9264**	**7.411**

Therefore, χ^2_{cal} = 7.411 and P-value(7.411, 3) = 0.05989 where 3 is obtained from n-k-1 (i.e. 5-1-1)

Conclusion: Since P-value = 0.05989 > α = 0.05, we do not have sufficient evidence to reject H_0. So, the researcher accepts the null hypothesis and conclude that the service time of patients follows an exponential distribution at α = 0.05

B. Traffic Intensity

This has to do with checking for the efficiency of the system which is denoted by ρ. It is a measure of how busy a system is, (that is the measure of the inter-arrival time rate over the service time rate of the system.)

Computationally, $\rho = \frac{\lambda}{s\mu}$

Hence,

$\rho = \frac{0.1745}{2(0.1677)} = \frac{0.1745}{0.3354} = 0.5203$

This shows that 52% of the total clinic working hours of Royal Cross Hospital and Maternity, antenatal unit is utilized. This indicates average service utilization.

C. **Computing some other performance measures are**:

Probability of the system being empty: $P_0 = \left[\sum_{n=0}^{s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{S^s}{S!}\frac{\rho^s}{1-\rho}\right]^{-1} = \left[\frac{1}{0!}(1.0406)^0 + \frac{1}{1!}(1.0406)^1 + \frac{2^2}{2!}\frac{0.5203^2}{1-0.5203}\right]^{-1} = (1.0406 + (2 * 2.1692))^{-1} = 0.1859$

Probability that the queue is non-empty: $P(n > 1) = 1 - \left(1 - \frac{\lambda}{\mu}\right) - \frac{\lambda}{\mu}\left(1 - \frac{\lambda}{\mu}\right)$

$= 1 - (1 - 1.0406) - 1.0406(1 - 1.0406) = 0.9984$

The expected number of patients in the queue: $E(N_q) = \frac{\rho(\rho s)^s P_0}{(1-\rho)^2 s!} = \frac{0.5203(0.5203\times 2)^2 \times 0.1859}{(1-0.5203)^2 2!}$

$= 0.2275$

The expected waiting time in the queue: $E(W_q) = \frac{(\rho s)^s P_0}{s!(\mu s)(1-\rho)^2} = \frac{(0.5203)^2 \times 0.1859}{2!(0.1677\times 2)(1-0.5203)^2}$

$= 0.3258\text{mins}(\text{i.e.}, 20\text{sec.})$

Table of results

Parameters	**Symbols**	**Results**
Mean inter-arrival time	λ	0.1745
Mean service time	μ	0.1677
Traffic intensity	ρ	0.5203
Probability of the system being empty	P_0	0.1859
Probability of the system being non-empty	$P(n > 1)$	0.9984
Expected number of customers in the queue	$E(N_q)$	0.2275
Expected waiting time in the queue	$E(W_q)$	0.3258 mins

III. Conclusion

The queuing system analyzed in this study is of the M/M/2 queuing model, where the exponential inter-arrival time λ = 0.1745 mins and the exponential service time μ = 0.1677 mins. From the analysis of this study, ρ was obtained to be <1 with a value of 0.52 which signifies that the system under consideration is 52% efficient (i.e. it utilizes 52% of its total working hours each day). Thus, arrivals are said to be within the server's capability..

IV. Recommendation

Computationally, the analysis carried out in this work shows that the mean inter-arrival time is greater than the mean service time. Therefore, there is actually a waiting line problem. To reduce this problem to its barest minimum, the researcher, therefore, recommend that the hospital should have a specific time to run the clinic and ensure that the consulting doctors are readily available so as to avoid cases of patient(s) coming too early and having to wait for hours before they receive service.

References

Erlang, A. K. (1909). The theory of probabilities and telephone conversations. *Nyt Tidsskrift Matematika, B.20*, 33-39

Biju, M. k., Naeema, K. and Faisal, U. (2010). Application of queuing theory in human resource management in health care. ICOQM-10: June 28-30.

Dhanavatan, A. v. (2010). Application of simulation technique in queuing model for ATM facility. *International Journal of Applied Engineering Research, Dindigul*, 1(30), 35-40.

Karlin, S. and Taylor, H. M. (1981). *A first course in stochastic processes*. Petroelli Book, New York.

Longnecker, J. G. (1969). *Principles of management and organizational behavior*. Merril Publishers, Columbus, Ohio.

Ogini, U.C (2014). *An statistical analysis of queuing system in a bank ATM Post: A Case Study of First Bank of Nigeria PLC, Lagos.* Unpublished B.sc Degree Project, Statistics Dept, UNN