GSJ: Volume 7, Issue 5, May 2019
ISSN 2320-9186

Global Scientific JOURNALS

GSJ: Volume 7, Issue 5, May 2019, Online: ISSN 2320-9186
www.globalscientificjournal.com

IDENTIFYING STATIONARY PERIODS AND DISTRIBUTION SELECTION FOR QUEUEING PROCESS

Oke, J. A.

FEDERAL SCHOOL OF STATISTICS, IBADAN, OYO STATE, Nigeria. okejamiade@yahoo.com

KeyWords
queueing process, test for stationarity, test for invariance, distribution selection.

ABSTRACT

The study adopted a systematic plan to observe and record the events of customer arrivals (C(t)), interarrival times (I_t) and service times (S_t). The observations were recorded hourly at different periods (morning, afternoon and evening) daily in a week at Nigerian National Petroleum Corporation (NNPC) Mega Station Port Harcourt Enugu Express, Emene, Enugu. Stationary periods were identified by considering two adjacent time intervals (0, t_1) and (t_1, t_2), the invariance of I_t and S_t were checked for the different periods. Test of stationarity and invariance were tested with non parametric tests of Mann-Whitney-Wilcoxon, and Friedman Two-way ANOVA respectively, which found stable for all periods. The I_t and S_t fitted exponential distribution in all the periods through the Anderson Dalling test of goodness of fit. At the stationary periods, the most suitable fitted exponential distribution for I_t and S_t were selected respectively for the distribution of the queue elements.

GSJ: Volume 7, Issue 5, May 2019
ISSN 2320-9186

INTRODUCTION

Queues are phenomena of everyday life. Human and non human wait in queues, students wait at bus stands to enter school bus; they wait at banks to pay school fees; vehicles wait at filling stations to buy fuel; customers wait at mall to obtain food items; they wait in front of point of sale terminals to pay for the items bought; etc. Queueing theory is the study of waiting in various guises (Frederick and Gerald, 2001). It uses queueing models to represent the various types of queueing systems (systems that involve traffic or queues of some kind) that arise in practice. Queueing theory is defined as the development of mathematical models to describe various types of queueing systems so that it may be possible to predict how the system will perform in a given demand situation Eze et al (2005).

There are basic elements responsible for queueing, they are: input process; service mechanism; system capacity; and queueing discipline. According to Bhat (2008), the nomenclature for symbolical representation of the traffic system with ranges of system elements are recognized by these basics. Input/service/number of server is the three elements symbol mostly adopted for representations in queueing theory, although there are six symbols in the general specification of queueing which is known as Kendal's notation (**I**/**S**/**s**/**b**/**P**/**Sd**). The **I** specifies inter-arrival time distribution, **S** specifies service time distribution, **s** specifies number of servers, **b** specifies number of buffer (the system capacity), **P** specifies population size and **Sd** specifies service discipline. The inter-arrival and service time distributions (i.e. **I**/**S** of the Kendal's notation above) can be fitted any of exponential (**M)**, Erlang with parameter k (**E_k**), hyperexponential with parameter k (**H_k**), deterministic (**D**), or general i.e. any distribution (**G**).

Consequently, if we have queueing system with infinite buffer capacity, infinite population size and service discipline is First Come First Serve (FCFS), the **b**/**P**/**Sd** is omitted in the Kendal's notation above and thereby resulted to the three elements symbol **I**/**S**/**s** only. For example, we can have M/M/1 (the two Ms are exponential inter-arrival and service time distributions with 1 server) or M/G/1 (M exponential inter-arrival and G general service time distribution with 1 server) or M/M/s (Ms exponential inter-arrival and service time distribution with s servers).

In building a suitable probability model for queue system, we start with its elements. Of the four elements mentioned earlier, number of servers, system capacity and discipline are normally deterministic (unless, the number of available servers becomes a random variable). But there are uncertainty related to arrivals and service process which result in the underlying process being stochastic. The similarity of arrival and service processes can be brought out by identifying similar components, such as inter-arrival times and service times. Therefore, the possibilities of using certain probability distributions to represent the process of inter-arrival times and service times are main concern of any queueing study.

There are types of queue; Single Queue with Single Service Point; Single Queue with Multiple Service Points; Multiple Queues with Multiple Service Points; and Multiple Queues with Single Service Point. In this study, Single Queue with Multiple Service Points where one waiting line is formed but there are several servers was adopted, each server capable of meeting the demand of individual in the queue.

Ger and Avishai (2001) in their work raised issues with queueing models, they argued that most work on queueing lack practical solution. They discouraged comparison between analytical models and simulation model. The relevance of appropriate method of data collection and analysis for such system performance was pointed out. They warned that assumed stationarity could be problematic if the system does not relax fast enough. It was opined that system performance can only be tracked if system of queue is study over a short interval of time and study in peak period should be avoided for it is extremely sensitive to changes in its underlying parameters.

In order to contribute to the existing studies, this work focus on identification of stationary periods which were problem found in most queueing works which are deemed necessary before assuming M/M/S queueing. Fredrick and Gerald (2001), Ger and Avishai (2001), and Bhat (2008) emphasized that this is the only period where probability distribution can be assumed.

The crucial intention of the investigation of queueing systems as described by Bhat (2008) is to comprehend the manners of their fundamental processes and able to give conversant and sharp decisions in their administration. Among other concerned problems of studying queueing system which face most researchers were the steady state or stationary period and how can the period be identified?

Let X(t) be a markov chain and number of customers at time t in a queueing system. If the system was observed for a very short interval of time h from time t, the probability X(t) in (t, t+h) is $P_n(t+h)$.

According to Mitrofanova (2007), the probability that there will be an increase of size 1 when h is short and $h \rightarrow 0$ is $P(X(t + h) - X(t) = 1| X(t) = n) = \lambda_n h + o(h)$; $P(X(t + h) - X(t) = -1| X(t) = n) = \mu_n h + o(h)$ when the system decrease by size 1, and $P(X(t + h) - X(t) = 0| X(t) = n) = 1 - (\lambda_n + \mu_n)h + o(h)$ for the probability that more than one event will occur. The λ_n and μ_n are birth and death rate respectively.

GSJ: Volume 7, Issue 5, May 2019
ISSN 2320-9186
Using Kolmogorov forward equation for the birth and death processes;

$$P_n(t+h) = \sum_{k=0}^{\infty} P_{n-k}(t)P_k(h)$$

$$= P_n(t) - (\lambda_n + \mu_n)hP_n(t) + \lambda_{n-1}hP_{n-1}(t) + \mu_{n+1}hP_{n+1}(t) + o(h) \quad 1$$

$$\Rightarrow \frac{P_n(t+h) - P_n(t)}{h} = -(\lambda_n + \mu_n)P_n(t) + \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t) + \frac{o(h)}{h} \quad 2$$

$$\Rightarrow \lim_{h\to 0} \frac{P_n(t+h) - P_n(t)}{h} = P_n'(t) = -(\lambda_n + \mu_n)P_n(t) + \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t) \quad 3$$

Equation (3) is the differential equation for the birth and death process.
In the study of queue, we consider only stationary system (i.e. system of equilibrium).
That is when $P_n(t) = P_n$
Equation (3) becomes;

$$0 = -(\lambda_n + \mu_n)P_n + \lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1} \quad 4$$

For M/M/S

$$\lambda_n = \lambda$$

$$\mu_n = n\mu \qquad n \le s$$

$$\mu_n = s\mu \qquad n > s$$

$$\therefore \quad P_n = \begin{cases} \left(\frac{\lambda^n}{\mu.2\mu....n\mu}\right)P_0 = \frac{\lambda^n}{n!\mu^n}P_0 = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 = \frac{\gamma^n}{n!}P_0 & n \le s \\ \left(\frac{\lambda^n}{\mu...(s-1)\mu)\mu.s\mu.\mu}\right)P_0 = \frac{\lambda^n}{s!\mu^s(s\mu s^{n-s}}P_0 = \frac{\gamma^n}{s!s^{n-s}}P_0 & n > s \end{cases} \quad 5$$

Consequently, the objective of this study is to identify the stationary periods, test the stationarity of the queueing process and fit theoretical probability distributions into interarrival time I_t and service time S_t at the stationary periods.

Material and Method

The data used for this study was a primary data collected from customers who bought Premium Motor Spirit (PMS) called petrol using direct observation method for three periods of the day i.e. morning, afternoon and evening for one week at NNPC Mega station Enugu-Port Harcourt express way, Emene, Enugu.

The NNPC Enugu Mega Station opens at 6 AM in the morning and closes at 6 PM in the evening. Therefore, a systematic collection of hourly observations was adopted. These hours were: 8.30 – 9.30 AM (morning); 12.30 – 1.30 PM (afternoon) and 4.30 – 5.30 PM (evening) With this, 21 periods' observations were recorded in seven days. The station has two gates, one for entrance and the other for exit. The service capacity for PMS (petrol) in the station is 12 pumps (servers). These 12 pumps are arranged in parallel of three (3) lines. A line contains two (2) machines with two (2) pumps each. Customers form queue along the road leading to the entrance gate.

In order to get the system elements, a traffic form was designed to capture:
C(t) = number of customers in one minute
A_t = customer arrival time
I_t= interarrival time (i.e. $A_t – A_{t-1}$)
E_t = entry to service time
D_t = departure time
S_t = service time (i.e. $D_t – E_t$)

GSJ: Volume 7, Issue 5, May 2019
ISSN 2320-9186
The number of arrivals from two adjacent time intervals $(0, t_1)$ and (t_1, t_2) was observed for several time periods. Let $X_1, X_2, \dots, X_n$ be the number of arrivals during the first interval for n periods, and let $Y_1, Y_2, \dots, Y_m$ be the second interval for m periods (usually m = n). In line with Conover (1971) and Randle and Wolfe (1979), the stationarity of the queueing process was tested with Mann-Whitney-Wilcoxon test on the observed arrivals from the two intervals. If F and G represent the distributions of the X's and Y's, respectively, then the hypothesis to be tested is F = G against the alternative $F \neq G$, for which the Mann-Whitney-Wilcoxon statistic can be used. The Mann-Whitney-Wilcoxon statistic was defined by Hogg and Craig (1970) as

$$Z_u = \frac{U - \frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \qquad \text{(for n or m > 20)} \qquad 6$$

under the null hypothesis F(z) = G(z) and decision rule that: if the p value for the test statistic (Z_u) is greater than 0.05 (level of significance) accept H_0 otherwise reject.
where

$$U = T - \frac{n(n+1)}{2}$$

If we pooled $X_1, X_2, \dots, X_n$ and $Y_1, Y_2 \dots, Y_m$ together, we will get m + n items. T is the sum of the ranks of $X_1, X_2, \dots, X_n$ among the m + n items $X_1, X_2, \dots, X_n, Y_1, \dots, Y_{m-1}, Y_m$, once the combined sample has been ordered.
To test the invariance of the system variables, we adopted the nonparametric technique of Friedman Two-Way Analysis of Variance by Ranks because of the skewed distributions of interarrival times and service times (Frederick and Gerald, 2001). Also there are two factors, the first is the periods and the other is the days.
For the Friedman test, the data were casted in a two-way table having 3 rows and 7 columns. The rows represent periods of the day (i.e. morning, afternoon and evening), and the columns represent the days of the week (i.e. Monday, Tuesday, …, Sunday). In as much as the columns contain equal number of cases, an equivalent statement would be that under H_0 the mean ranks of various columns would be about equal and decision is taken base on decision rule that: if the p value for the test statistic is greater than 0.05 (level of significance) accept H_0 otherwise reject. The Friedman test determines whether the rank totals (R_j) differ significantly. To make this test, we compute the value of a statistic which Friedman denotes as χ_r^2. It can be shown that χ_r^2 is distributed approximately as Chi square with df = k – 1 (Siegel, 1956), when

$$\chi_r^2 = \frac{12}{Nk(k+1)}\sum_{j=1}^{k} R_j^2 - 3N(k+1) \qquad 7$$

where N = number of rows; k = number of columns; R_j = sum of ranks in jth column.

In order to identify the distribution of a given data, Minitab (a statistical package) can be used to identify distribution any given data using the following procedure (Frost, 2012):
Stat>Quality Tools>Individual Distribution Identification (in Minitab)
This handy tool easily compares how well one data fit 16 different distributions. It produces a lot of output both in the Session window and graphs (Frost, 2012). It was explained further that there were measures to check in the output.
Anderson-Darling statistic (AD): lower AD values indicate a better fit. It is generally valid to compare AD values between distributions and go with the lowest
P-value: One needs a high p-value. A low p-value (e.g., < 0.05) indicates that the data do not follow that distribution.
The A-D statistic controls the hypothesis that the sample derives from a distribution which is described by the fitted density function, using the A^2 statistic:

$$A^2 = N - S \qquad 8$$

where $S = \sum_{i=1}^{N} \frac{2i-1}{N}\left[\text{In}F(x_i) + \text{In}(1 - F(x_{N+1-i}))\right]$ and $x_1, \dots, x_N$ are the sample values sorted in order of magnitude. Gavriil et al (2006) stated that A-D statistic is considered more dependable than any other statistic since it emphasizes at the upper tails of the distribution functions where the larger discrepancies are expected.

GSJ: Volume 7, Issue 5, May 2019
ISSN 2320-9186
The presentations and computations in the next section will be facilitated using the following spreadsheet packages: Ms Excel, SPSS and Minitab.

ANALYSIS AND RESULT

The third-ten minutes of the 21 periods (morning, afternoon and evening for seven days of the week) were considered and data for the test were obtained within two adjacent time intervals (t_0 = 0, t_1 = 5mins] and (t_1 = 5mins, t_2 = 10mins]. Hence, number of arrivals in between $A(t_0)$ and $A(t_1)$ is $C(t_i)$ (X_i for easy transcription) and number of arrivals in between $A(t_1)$ and $A(t_2)$ is $C(t_j)$ (Y_j for easy transcription).

We mean, $A(t_1) – A(t_0) = C(t_i) = X_i \qquad i = 1, 2, \ldots, n$

$A(t_2) – A(t_1) = C(t_j) = Y_j \qquad j = 1, 2, \ldots, m$

$n = m = 21$

The observed values for X's and Y's are displayed in table 3.1 below.

Table 3.1: Observed Number of Arrivals at $[t_0, t_1]$ and $[t_1, t_2]$

S/N	X	Y
1	5	18
2	12	8
3	6	4
4	6	5
5	16	6
6	8	11
7	17	13
8	8	13
9	13	6
10	9	3
11	15	17
12	9	6
13	10	10
14	16	12
15	11	7
16	9	9
17	6	13
18	4	6
19	2	12
20	12	7
21	9	13

Using Mann-Whitney-Wilcoxon test, the result of the test by SPSS is displayed in Table 3.2 below.

H_0: F = G

H_1: F ≠ G

Decision rule: reject H_0 if p value for Mann-Whitney-Wilcoxon test statistics is less than $\alpha = 0.05$, otherwise accept.

Table 3.2: Mann-Whitney-Wilcoxon Test Statistics

Statistics	C(t)
Mann-Whitney U	215.500
Wilcoxon W	446.500
Z	-.126
p-value	.899

GSJ: Volume 7, Issue 5, May 2019
ISSN 2320-9186
Under the null hypothesis (H_0) that F = G, we would accept the H_0 and conclude that the process is stationary at the two adjacent time intervals considered as p-value (0.899) is greater than $\alpha = 0.05$. It is at this interval of times that we will determine our distributions for I_t and S_t.

From above, it is evident that the system is at stationary at the third-ten minutes of the periods. We took equal sample of size 20 at these periods and computed mean for I_t and S_t.
Putting the means of I_t in periods (rows) and days (columns), we have
Table 3.3: Means of I_t in periods and Days

Periods	Days						
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Morning	0.47	0.52	0.56	0.47	0.46	0.5	0.48
Afternoon	0.5	0.53	0.48	0.5	0.46	0.52	0.51
Evening	0.52	0.46	0.55	0.47	0.51	0.48	0.55

H_0: there is no significant difference in the means of I_t at the station during periods of days in a week
H_1: there is significant difference in the means of I_t at the station during periods of days in a week
Decision rule: reject H_0 if p value for Friedman's test statistics is less than $\alpha = 0.05$, otherwise accept.
Table 3.4: Result of Friedman Test on Means of I_t

Test Statistics	
N	3
Chi-Square	6.218
Df	6
p-value.	.399

Putting the means of S_t in periods (rows) and days (columns), we have
Table 3.5: Means of S_t in periods and Days

Periods	Days						
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Morning	2.53	2.56	2.65	2.45	2.45	2.46	2.52
Afternoon	2.53	2.47	2.47	2.42	2.53	2.53	2.47
Evening	2.44	2.53	2.50	2.59	2.52	2.55	2.45

Test of hypothesis
H_0: there is no significant difference in the means of S_t at the station during periods of days in a week
H_1: there is significant difference in the means of S_t at the station during periods of days in a week
Decision rule: reject H_0 if p value for Friedman's test statistics is less than $\alpha = 0.05$, otherwise accept.

Table 3.6: Result of Friedman Test on Means of S_t

Test Statistics	
N	3
Chi-Square	2.226
Df	6
p-value	.898

The two results show that the means of Y_t and S_t are invariant over the days of the week and times of the day at the stationary periods. It implies that the events Y_t and S_t remain the same at some interval of times from Monday to Sunday and any period of the days (in this case, third ten minutes of the hourly study at the station).

GSJ: Volume 7, Issue 5, May 2019
ISSN 2320-9186

The distributions of Y_t and S_t at stationary periods were fitted exponential distribution and tested for goodness-of-fit. Anderson-Darling goodness-of-fit was tested on the fitted distributions and the results are presented below.

Table 3.7: Test of goodness of fit result

Day	Period	I_t			S_t		
		MLE Scale (mean)	AD statistic	P value	MLE Scale (mean)	AD statistic	P value
Mon	Morning	0.465	0.763	0.217	2.534	0.757	0.221
	Afternoon	0.503	1.129	0.077	2.531	1.233	0.058
	Evening	0.518	1.052	0.095	2.438	1.275	0.052
Tue	Morning	0.521	1.254	0.055	2.556	1.217	0.061
	Afternoon	0.531	0.876	0.155	2.467	1.261	0.054
	Evening	0.461	1.162	0.07	2.534	0.757	0.221
Wed	Morning	0.56	0.71	0.254	2.651	1.209	0.062
	Afternoon	0.478	1.218	0.06	2.467	1.261	0.054
	Evening	0.545	1.272	0.052	2.498	1.209	0.062
Thu	Morning	0.465	0.763	0.217	2.445	1.176	0.068
	Afternoon	0.503	1.29	0.077	2.418	0.854	0.166
	Evening	0.467	1.129	0.077	2.589	0.777	0.208
Fri	Morning	0.461	1.162	0.070	2.445	1.176	0.068
	Afternoon	0.455	0.561*	0.400*	2.534	0.757	0.221
	Evening	0.513	1.19	0.065	2.518	0.669*	0.288*
Sat	Morning	0.503	1.129	0.077	2.462	1.112	0.081
	Afternoon	0.518	0.85	0.168	2.534	0.757	0.221
	Evening	0.475	0.816	0.185	2.55	0.81	0.189
Sun	Morning	0.475	0.816	0.185	2.517	1.171	0.068
	Afternoon	0.513	1.19	0.065	2.467	1.261	0.054
	Evening	0.549	0.887	0.151	2.445	1.176	0.068

*lowest AD statistic with highest p value

In the above table, I_t and S_t fitted exponential distribution for all days and periods during the stationary periods since all p values are greater than 0.05 for all AD statistic. The most fitted exponential distribution is the one with lowest AD value and highest p value (0.561 and 0.400 for I_t; 0.669 and 0.288 for S_t). Therefore, the distribution selection for I_t and S_t was exponential distribution with parameter scale 0.455 and 2.518 respectively.

Conclusion

This work sought to find stationary periods for a queueing system as one of the assumption of m/m/s queueing model and fitted suitable distribution for the main stochastic elements (the interarrival (I_t) and service times (S_t)) of the model at the identified stationary periods. Two adjacent times were observed for the periods and Mann-Whitney-Wilcoxon test was used to test the stationarity. The result shows that process is stationary with Mann-Whitney-Wilcoxon statistic, Z = -0.126 and p-value = 0.899. With the result, the hypothetical periods were confirmed stationary for the queueing model. Also, Friedman two-way ANOVA proved that the system variables were invariant with respect to days of the week and periods of the day. The results revealed that chi square statistics of Friedman with the p values for Y_t and S_t were 6.218 (p = 0.399) and 2.226 (p = 0.898) repectively.

At the stationary periods, all Y_t and S_t were fitted exponential distribution through Anderson-Darling (AD) goodness-of-fit for continuous distributions. It was found that the most suitable exponential distribution for Y_t and S_t was that with AD statistics 0.561, 0.669 and p-values 0.400, 0.288 respectively. From the result of test of goodness-of-fit, the maximum likelihood estimates for the parameters of exponential distribution of Y(t) and S(t) were approximately 0.5 and 2.5 respectively. It means averagely 2 customers

GSJ: Volume 7, Issue 5, May 2019
ISSN 2320-9186
arrive for service in one minute with interarrival times averagely 0.5 minute and each customer's service time is averagely 2.5 minutes.
It is recommended that strict adherence should be given to stationarity in the analysis of queueing study of m/m/s. This will give traceable and brilliant estimate of the system metrics.

References

Bhat, U. N. (2008). **An Introduction to Queueing Theory**, Springer Science Business Media, LLC, New York.

Conover, W. J. (1971). **Practical Nonparametric Statistics**, John Wiley and Sons, New York.

Eze, J. I., Obiegbu, M. E. and Jude-Eze, E. N. (2005). **Statistics and Quantitative Methods for Construction and Business Manager**,YUM-SEG Enterprises, Lagos.

Fredrick, H. S. and Gerald, L. J. (2001). **Introduction to Operations Research**, McGraw-Hill Higher Education, 7th edition, New York.

Frost, J. (2012). "How to identify the Distribution of your Data using Minitab", The Minitab Blog: Adventures in Statistics.
:http://blog.minitab.com/blog/adventures-in-statistics/how-to-identify

Gavriil, I., Grivas, G., Kassomenos, P. and A. and Spyrellis, N. (2006). "An Application of Theoretical Probability Distributions to the Study of PM_{10} and $PM_{2.5}$ Time Series in Athens, Greece"**,** Global Nest Journal, vol. 8, No. 3.

Ger, K. and Avishai, M. (2001). Queueing Models of Call Centres, An Introduction, available at www.cs.vu.nl/obp/callcentres

Hogg, R. V. and Craig, A. T. (1970). **Introduction to Mathematical Statistics**, Macmillan Publishing Co, Inc, New York.

Mitrofanova, A. (2007). "Continuous Times Markov Chains, Poisson Process, Birth and Death Process", Lecture Note, New York University.

Randles, R. H. and Wolfe, D. A. (1979). **Introduction to the Theory of Nonparametric Statistics**, John Wiley & Sons, New York.

Siegel, S. (1956). **Nonparametric Statistics for the Behavioural Sciences**, McGraw-Hill, International Student Edition, New York.