UNIT 13 QUEUEING MODELS

Objectives

After studying this unit, you should be able to :

- identify the occurrence of queueing in weal life situations.
- describe the characteristics of a queueing problem.
- use the statistical methods necessary to analyse queueing problems.
- apply the common queueing models in suitable problems.
- estimate the optimum parameters of a queueing model with respect to cost and service level.

Structure

13.1 INTRODUCTION

The **queueing problem** is identified by the presence of a group of customers who arrive **randomly** to receive some service. The customer upon arrival may be attended to immediately or may have to wait until the server is free. The service time required to serve the customers is also a **statistical** variable. This methodology can be applied in the field of business (banks, booking counters), industries (servicing of machines), government {railway or post-office counters), transportation (airport, harbour) and everyday life (elevators, restaurants, doctor's chamber).

The queueing models are basically relevant to service oriented organisations and suggest ways and means to improve the efficiency of the service. An improvement of service level is always possible by increasing the number of employees. Apart from increasing the cost an immediate consequence of such a step is unutilised or **idle time** of the servers. In addition, it is unrealistic to assume that a large-scale increase in staff is possible in an organisation. Queueing methodology indicates the optimal usage of existing manpower and other resources to improve the service. It can also indicate the cost implications if the existing service facility has to be improved by adding more servers.

The relationship between queueing and service rates can be diagrammatically illustrated using the cost curves shown in Figure 13.1.

At a slow service rate, queues build up and the cost of queueing increases. An ideal service unit will minimise the operating cost of the entire system.

13.2 CHARACTERISTICS OF A QUEUEING MODEL

A **queueing system** can be described by the following components:

Arrival

The statistical pattern of the arrival can be indicated through the probability distribution of the number of arrivals in an interval. This is a **discrete random variable**

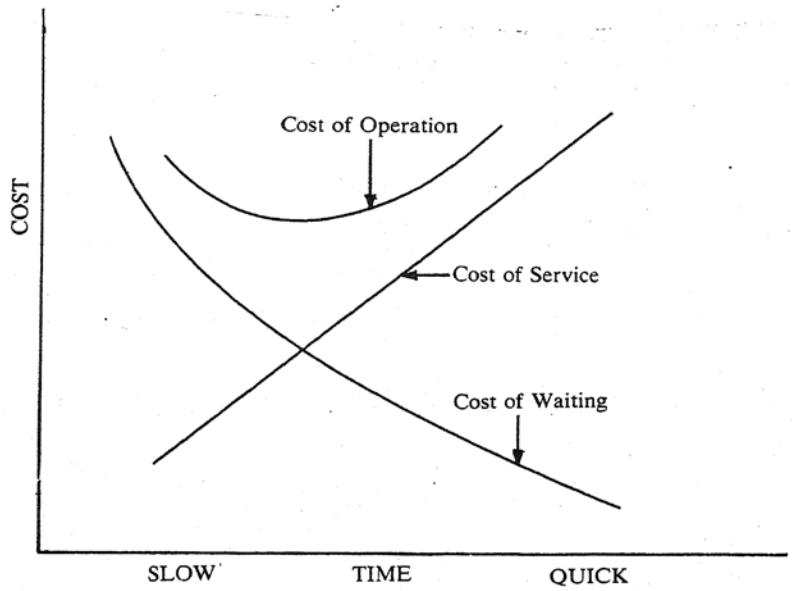

Figure 13.1: Cost Structure of a Queueing Problem

(Block 3, Unit 10, Section 3 of MS 8). Alternatively, the probability distribution of the time between successive arrivals (known as **interarrival time)** can also be studied to ascertain the stochastic aspect of the problem. This variable is **continuous** in nature (Block 3, Unit 11 of MS 8).

The probability distribution of the arrival pattern can be identified through analysis of past data. The discrete random variable indicating the number of arrivals in a time interval and the continuous random variable indicating the time between two successive arrivals **(interarrival time)** are obviously interrelated. A remarkable result in this context is that if the number of arrivals follows a **Poisson distribution** (Block 3, Unit 10, Section 5 of MS 8), the corresponding interarrival time follows an **exponential distribution** (Block 3, Unit 11, Section 3 of MS 8). This property is frequently used to derive elegant results on queueing problems.

Service

The time taken by a server to complete service is known as **service time.** The service time is a **statistical variable** and can be studied either as the number of services completed in a given period of time or the completion period of a service. The data on actual service time should be analysed to find out the probability distribution of service time.

Server

The service may be offered through a single server such as a ticket counter or through several channels such as a train arriving in a station with several platforms.

Sometimes the service is to be carried out sequentially through several phases known as **multiphase service.** In government, the papers move through a number of phases in terms of official heirarchy till they arrive at the appropriate level where a decision can be taken.

Time spent in the queueing system

The time spent by a customer in a queueing system is the sum of waiting time before service and the service time.

Queue discipline

The queue discipline indicates the **order** in which members of the queue are selected for service. It is most frequently assumed that the customers are served on a first' come first serve basis. This is commonly referred to as **FIFO** (first in, first out) system. Occasionally, a certain group of customers receive_ priority in service over others even if they arrive late. This is commonly referred to as **priority queue,** The queue discipline does not always take into account the order of arrival. The server chooses one of the customers to offer service at random. Such a system is known as service in random order (SIRO). While allotting an item with high demand and limited

supply such as a test match ticket or share of a public limited company, SIRO system is the only possible way of offering service when it is not possible to identify the order of arrival.

Size of a population

The collection of potential customers may be very large or of a moderate size. In a railway booking counter the total number of potential passengers is so large that although theoretically finite it can be regarded as infinity for all practical purposes. The assumption of infinite population is very convenient for analysing a queueing model. However, this assumption is not valid where the customer group is represented by a few looms in a spinning mill that require operator facility from time to time. If the population size is finite then the analysis of queueing model becomes more involved.

Maximum size of a queue

Sometimes only a finite number of customers. are allowed to stay in the system although the total number of customers in the population may or may not be finite. For example, a doctor may make appointments with **k** patients in a day. If the number of patients asking for appointment exceeds k, they are not allowed to join the queue. Thus, although the size of the population is infinite, the maximum number permissible in the system is k.

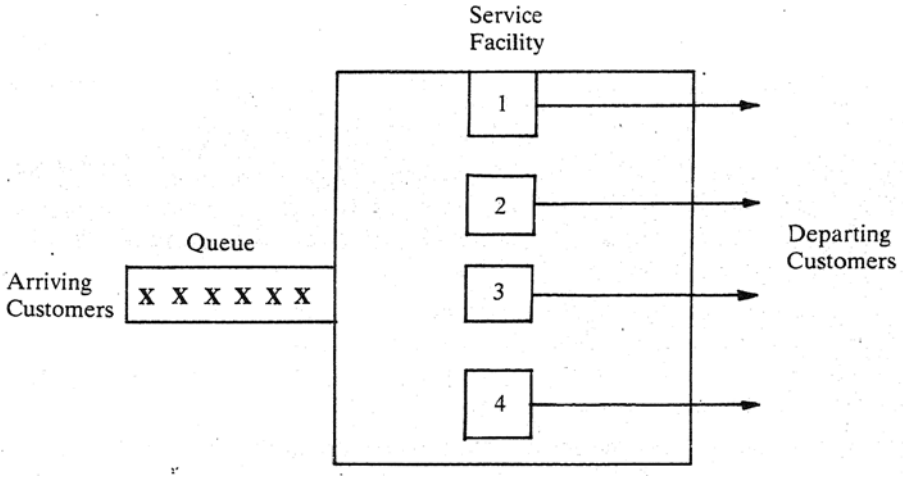

Figure 13.2: Schematic Representation of a Queueing Problem

13.3 NOTATIONS AND SYMBOLS

Kendall (Kendall, 1951) has introduced a set of notations which have become standard in the literature of queueing models. A general queueing system is denoted by (a/b/c) : (d/e) where

a = probability distribution of the interarrival time.

b = probability distribution of the service time.

c = number of servers in the system.

d = maximum number of customers allowed in the system.

e = queue discipline.

In addition, the size of the population as mentioned in the previous section is important for certain types of queueing problem although not explicitly mentioned in the Kendall's notation. Traditionally, the exponential distribution in queueing problems is denoted by M. Thus (M/M/1) : (∞/FIFO) indicates queueing system when the interarrival times and service times are exponentially distributed having one server in the system with first in first out discipline and the number of customers allowed in the system can be infinite.

In general, the behaviour of a queueing system will depend upon time. Such a system is said to be in a **transient state.** This usually occurs at an early stage of formation of queues where its behaviour is still dependent upon the initial conditions. When sufficient time has elapsed since the beginning of the operation (mathematically as

the time approaches to infinity) the behaviour of the system may become **independent** of time. The system then is said to be in a steady state. Only steady state queueing problems will be analysed in this unit.

The following symbols will be used while studying queueing models.

n = number of units in the system who are waiting for service and who are being served.

$P_n(t)$ = transient state probabilities of exactly n customers in the system at time t assuming that the system has started its operation at time zero.

P_n = steady state probability of having n customers in the system.

^{X} = mean effective arrival rate (number of customers arriving per unit time).

μ = mean service rate per busy server (number of customers served per unit tine) C = number of parallel servers.

W(t) = density function of the waiting time.

W_s = expected waiting time per customer in the system including the service time.

W_q = expected waiting time per customer in the queue excluding the service time.

L_s = expected. number of customers in the system including those who are receiving service.

L_q = expected number of customers in the queue excluding those who are receiving service.

13.4 STATISTICAIL METHODS IN QUEUEING

The arrival of customers, the departure of customers after service and the waiting time of a customer before being served in a queueing system are random phenomena. In order to develop appropriate statistical model of this problem the following assumptions are made

1) If N(t) denotes the number of arrivals or departures after service in the time interval (0,t) then N(t) has **independent increments.** This can be stated statistically as follows. If $t_1<t_2<t_3$ then $\{N(t_3) - N(t_2)\}$ and $\{N(t_2) - N(t_1)\}$ are **independent random variables** (Block 4, Unit 14, Section 2 of MS 8)

2) If there are n units in the system, the probability of exactly one arrival from t to $t+\Delta t$ is $\lambda_n\Delta t+o(\Delta t)$. Likewise, the probability that exactly one departure after service will occur from t to $t+\Delta t$ is $\mu_n\Delta t+o(\Delta t)$. The quantity $o(\Delta t)$ represents a function of Δt such that

$$\lim_{t-0} \frac{o(\Delta t)}{t} = 0.$$

3) If there are n customers at time t, the probability that the number of arrivals and departures combined will exceed one during the time interval t to $t+\Delta t$ is $o(\Delta t)$.

Poisson Process

It can be shown (Taha, 1971) when $\lambda_n = \lambda$ and $\mu_n=0$ for all n that under conditions (1), (2) and (3), the probability $P_n(t)$ of having exactly n customers in the system (assuming that there is no customer in the system at time 0) is given by

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}; \quad n = 0,1,2,.....$$

Thus, if the service component is ignored the number of arrivals in the interval (0,t) has the probability distribution of a Poisson variable with parameter λt which is the product of the **arrival rate** and the length of the interval t. This probability law is known as the **Poisson Process.**

Relationship between Poisson Process and Exponential Probability Distribution

a) Consider a queueing model in which the number of arrivals in an interval of length t follows a **Poisson Process** with **arrival rate** λ**.** If the interarrival times are **independent** random variables, they must follow an exponential distribution with density f(t) where

$$f(t)= \lambda e^{-\lambda t} ; t>0.$$

It can be readily shown by integration that E(t) = $\frac{1}{\lambda}$. Thus, if the arrival rate λ =20/hour the average time between two successive arrivals is $\frac{1}{20}$ hour or 3 minutes.

b) If the interarrival times in a queueing system are independently, identically distributed **exponential** random variables then the number of arrivals in an interval follows a **Poisson Process** with arrival rate identical with parameter of the exponential distribution.

c) Consider a random variable X with an exponential probability distribution. Then for s, t>0

$$P(X>t+s \mid X>s) = P(X>t)$$

This property is important for solving queueing problems. Assuming that when a customer arrives the service is in progress for s units of time. The chance that the service will continue for at least an additional t units is identical with the probability that the service will prolong for at least t units after a fresh start.

One way to look at this property is that the time already spent in the system has no relevance to the additional time the customer is likely to spend in the system further. Thus, this property is sometimes referred to **as lack of memory or forgetfulness** of the exponential distribution.

Activity 1

Show by actual computation that if a random variable t has exponential distribution with parameter X, its expected value is $\frac{1}{\lambda}$.

Activity 2

Prove the lack of memory property of the exponential distribution.

……
……
……

13.5 THE M/M/l SYSTEM

In this queueing model, it is assumed that the number of customers arriving in a time interval t follows a Poisson Process with parameter λ. Equivalently, the interval between any two successive arrivals is exponentially distributed with parameter **A.** The time taken to complete a single service is exponentially distributed with parameter μ. The number of server is one. Although not explicitly stated both the population and the queue size can be infinity. The order of service is assumed to be FIFO.

It may be shown (Taha, 19717 that under the conditions stated in the previous section, the **steady state** probabilities P_n of the queueing system exists if $\frac{\lambda}{\mu}$. The probabilities are :

P_n = P (No. of customers in the system=n)

$$= (\frac{\lambda}{\mu})^n (1-\frac{\lambda}{\mu}); \quad n= 1,2,\ldots\ldots\ldots$$

$$P_o = 1 - \frac{\lambda}{\mu}.$$

This probability distribution is known as **Geometric Probability Distribution.**

The quantity $\frac{\lambda}{\mu}$ is usually known as **traffic intensity** and is usually denoted by π.

The condition under which the steady state solution is available can be expressed as p<1.

On re-examining this condition, we observe that this is equivalent to the condition that the expected service time is less than the expected interarrival time. If on an average, service is quicker than the average gap between arrivals then the steady state solution exists as stated earlier. If on the other hand, expected arrivals are too quick in comparison with expected service no steady state solution is available.

The expected number of customers in the system is given by

$$L_s = E(n) = \sum_{n=1}^{\infty} nP_n = \sum_{n=1}^{\infty} n(1-\frac{\lambda}{\mu})(\frac{\lambda}{\mu})^n$$

$$= \frac{\lambda}{\mu-\lambda} = \frac{\rho}{1-\rho}.$$

The expected number of customers in the queue is

$$L_q = \sum_{n=1}^{\infty} (n-1)P_n = \sum_{n=1}^{\infty} nP_n - \sum_{n=1}^{\infty} P_n$$

$$= \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{\rho^2}{1-\rho}.$$

Example 1

At a certain petrol pump, customers arrive according to a Poisson process with an average time of 5 minutes between arrivals. The service time is exponentially distributed with mean time = 2 minutes. On the basis of this information find out

a) What would be the average queue length?

b) What would be the average number of customers in the queueing system?

Solution 1

This is an M/M/1 queueing model.

Average inter-arrival time $= \frac{1}{\lambda} = 5$ minutes $= \frac{1}{12}$ hour

$\lambda = 12$/hour

Average service time $= \frac{1}{\mu} = 2$ minutes $= \frac{1}{30}$ hour

$\mu = 30$/hour

Hence $\lambda < \mu$ and the steady state solution exists.

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{144}{30\times 18} = \frac{4}{15}.$$

$$L_s = \frac{\lambda}{\mu-\lambda} = \frac{12}{18} = \frac{2}{3}.$$

Probability Distribution of waiting time in an M/M/I system

Let T be the time spent by a customer in the system. Based on first in first out service discipline if an arriving customer finds **It** person ahead of him in the queueing system then

$$T = t_1 + t_2 + \ldots\ldots\ldots\ldots + t_n + t_{n+1},$$

where t_1 is the **additional time** taken by the customer in service, t_2,……, t_n are the service times of the other customers ahead of him and t_{n+1} is the service time of the arriving customer. By **lack** off **memory** property of the exponential distribution, t_1 is also distributed as an exponential variable with parameter μ. Thus, T is the sum of (n+1) **independently identically exponentially** distributed random variables. The density function of T is a **Gamma distribution** whose density function is W(t|n+1) where

$$W(t|n+1) = \frac{\mu(\mu t)^n}{n!} e^{-\mu t}; \quad t>0.$$

A special form of Gamma distribution namely chi-square distribution has already been introduced in Block 4, Unit 16 of MS 8. If $\mu t = \frac{Z}{2}$, $n+1 = \frac{r}{2}$, then $W(t \mid n+1)$ becomes **the density function of chi-square distribution with r degrees of freedom.**

The density function of the total time W(t) can be computed by first multiplying the expression of $W(t|n+1)$ with the probability that there are n customers in the system and then summing over all values of n from 0 to ∞. The density function W(t) is, thus,

$$W(t) = (\mu - \lambda)\, e^{-(\mu-\lambda)t}; \quad t>0.$$

Thus, the total time of a customer in the system follows **an exponential probability distribution with parameter** μ-λ

It may be also of some interest to compute the probability distribution of T* of waiting time of an incoming customer before he receives the service. This probability distribution has two components. The customer starts receiving service immediately after his arrival if there is no. customer in the system. Thus

$$P(T^* = 0) = P_0 = 1 - \frac{\lambda}{\mu}.$$

If there are n customers in the system ($n \geq 1$) when a customer arrives then following a similar line of argument, the density function of waiting time before receiving the service W*(t) is given by

$$W^*(t) = \frac{\lambda}{\mu}(\mu-\lambda)\, e^{-(\mu-\lambda)t} \quad (t>0)$$

The expected time of a customer in the system W_s and the expected waiting time before receiving the service W_q are

$$W_s = \frac{1}{\mu-\lambda}$$

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

Example 2 (Continuation of the Example 1)

c) What is the average time spent by a car in the petrol pump?

d) What is the average waiting time of a car before receiving petrol?

Solution

c) Average time spent in the petrol pump

$$= \frac{1}{\mu-\lambda} = \frac{1}{18}\text{ hour} = 3.33\text{ minutes}$$

d) Average waiting time of a car before receiving petrol

$$= \frac{\lambda}{\mu(\mu-\lambda)} = \frac{12}{30} \times \frac{1}{18}\text{ hour} = \frac{1}{45}\text{ hour} = 1.33\text{ minutes.}$$

Activity 3

Derive the expression of W(t).

……
……
……
……

Activity 4

In the problem discussed in the section the management of the petrol pump is willing to open a second pump if the average waiting time before service is 5 minutes. What should be the minimum arrival rate for which such a decision should be taken?

……
……
……
……

13.6 THE M/M/C SYSTEM

In this queueing model it is assumed that the interarrival times are independently, identically exponentially distributed random variables with a common parameter A. The service times are independently, identically, exponentially distributed random variables with a common parameter μ. Although not explicitly stated the order of service is first in first out and both the population size and the queue size can be infinity. The number server in the system is c (> 1) so that service can be provided simultaneously to c customers.

The **steady state probabilities** exist if

$$\rho = \frac{\lambda}{c\mu} < 1.$$

These probabilities are given by :

$$\frac{1}{P_0} = \sum_{n=0}^{c-1} \frac{(\lambda/\mu)^n}{n!} + \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!} \times \frac{1}{1-\rho}$$

$$P_n = \frac{(\lambda/\mu)^n}{n!} P_0 \quad 0 \leq n \leq c.$$

$$= \frac{(\lambda/\mu)^n}{c!\, c^{n-c}} P_0 \quad n \geq c.$$

The expected number of customers in the queue

$$L_q = P_0 \frac{(\lambda/\mu)^c}{c!} \frac{\rho}{(1-\rho)^2}.$$

The expected waiting time in the queue

$$W_q = \frac{1}{\lambda} L_q.$$

The expected waiting time in the system

$$W_s = W_q + \frac{1}{\mu}.$$

The expected number of the customer in the system

$$L_s = \lambda W_s = L_q + \frac{\lambda}{\mu}.$$

Example 3

A petroleum company is considering expansion of its one unloading facility at its refinery. Due to random variations in weather, loading delays and other factors, ships arriving at the refinery to unload crude oil arrive at a rate of 5 ships per week. The service rate is 10 ships per week. Assume arrivals follow a Poisson Process and the service time is exponential.

a) Find the average time a ship must wait before beginning to deliver its cargo to the refinery.

b) If a second berth is rented what will be the average number of ships waiting before being unloaded?

c) What would be the average time a ship would wait before being unloaded with two berths?

d) What is the average number of idle berths at any specified time?

Solution

a) The expected waiting time W_q before a ship begins to unload crude oil is

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{5}{10\times 5} = \frac{1}{10} \text{ week.}$$

b) With a second berth the queueing model is an M/M/2 system.

$$\rho = \frac{\lambda}{2\mu} = \frac{5}{20} = \frac{1}{4}.$$

$$\frac{1}{P_0} = 1 + \frac{\lambda}{\mu} + \frac{(\lambda/\mu)^2}{2} \times \left(\frac{1}{1-\frac{1}{4}}\right)$$

$$= 1 + \frac{1}{2} + \frac{1}{8} \times \frac{4}{3} = \frac{5}{3}$$

$$\therefore P_0 = \frac{3}{5}.$$

$$L_q = \frac{\frac{3}{5} \times (\frac{1}{2})^2 \times \frac{1}{4}}{2(1-\frac{1}{4})^2} = \frac{1}{30}$$

c) $W_q = \frac{L_q}{\lambda} = \frac{1}{150}$ week

d) If there is no ship in the system then both the berths are idle. If there is only one ship in the system one of the berths is empty. Hence the expected number of idle berths is

$$2P_0 + P_1 = 2 \times \frac{3}{5} + \frac{1}{2} \times \frac{3}{5} = 1.5.$$

The M/M/∞ System

The infinite server queueing system is the limiting behaviour of an M/M/C model when each customer is **served immediately on arrival.** Such a situation is conceivable when the customers are using self-service. There is no waiting before the start of the service so that $L_q = 0$, $W_q = 0$. Under assumptions identical to those in M/M/C system, as C approaches to infinity the steady state probabilities are given by:

$$\frac{1}{P_0} = \sum_{n=0}^{\infty} (\lambda/\mu)^n/n! = e^{\lambda/\mu}$$

$$P_0 = e^{-\lambda/\mu}$$

$$P_n = e^{-\lambda/\mu} (\lambda/\mu)^n/n!$$

Thus, the steady state probability P_n indicating that there are n customers in the system is given by a **Poisson distribution** with parameter A/μ. The expected number of customers in tl system $L_S = E(n) = \frac{\lambda}{\mu}$. The expected time spent by a customer in the system is 1/μ.

Example 4

A cafeteria with self-service has an arrival rate of 12 per hour. The average time taken by a person to collect and eat his meal is 20 minutes. Assuming that the interarrival times are exponentially distributed how many seats must the cafeteria have to accommodate each customer with 95% probability.

Solution

This is an M/M/co queueing system. λ=12/hour μ = 3/hour. Hence the number of customers in the cafeteria follows a Poisson distribution with parameter λ/μ=4. It is thus necessary to find the value of k so that $P(n \leq k)$=.95. This value of k is 7.

Activity 5

A Bank has two counters working on savings accounts. The first counter deals with withdrawals only. The second counter deals with the deposit of the clients. It has been found that the service time distributions for both deposits and withdrawals are exponential with mean service time of 3 minutes per customer. Depositors are found.

to arrive according to a Poisson process with an arrival rate of 16/hour. Clients who want to withdraw arrive according to a Poisson process with an arrival rate of 14/hour. What would be the effect on the average waiting time of the depositors and the average waiting time for the customers who arrive for withdrawal if each counter can deal with both withdrawals and deposits?

……
……
……

13.7 THE M/E_K/I SYSTEM

In many queueing models the service consists of a number of phases. For example, the repair of a certain machine may require a number of sequential steps. The time taken at each phase of repair is a random variable following **in exponential diaribution** with parameter μ. Let there be k phases, the service time of the ith phase being S_i (i = 1, 2 , ..., k). The total service time S can be expressed as

$$S = S_1 + S_2 + \ldots\ldots\ldots + S_k.$$

It is also assumed that S_i (i = 1, 2,……, k) are independently distributed. Hence the probability distribution of S is a Gamma distribution with parameter k and μ as introduced in Section 13.5. The Gamma distribution is also, a member of **Erlang** of distribution and is denoted by E_K when k exponential variables are involved.

Apart from the justification of the Erlang family of service distribution as indicated above the family can approximate a variety of service time distributions. In fact, both the exponential service time distribution and constant **service time distribution** are special cases of Erlang service time distribution with k=l and k=- respectively. In between there are service time distributions of various types.

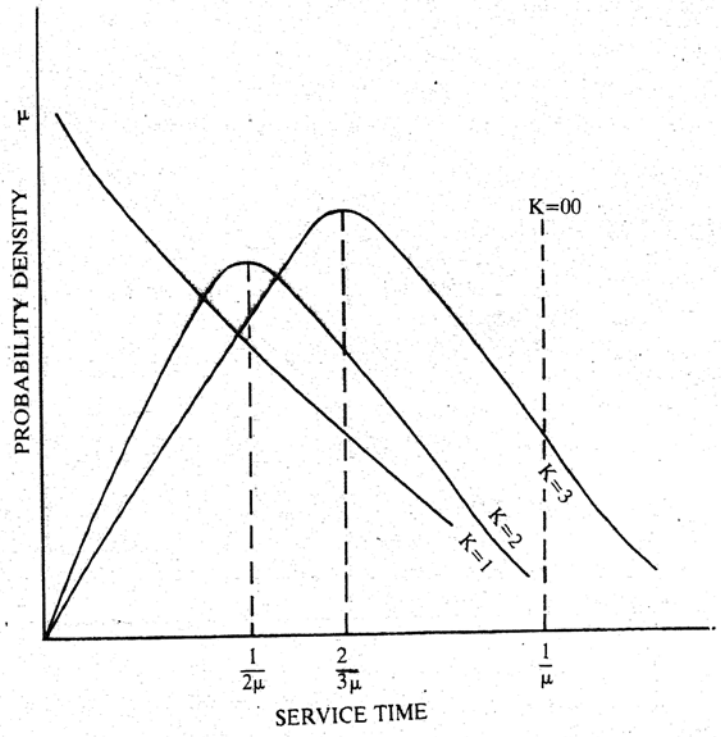

Figure 13.3 : The Erlang Family of Service Distributions

Each member of the Erlang distribution has the same mean $\frac{1}{\mu}$. The **mode** is at t=0 for k=1. It is at t = $\frac{1}{2\mu}$ for k=2. In general the mode is located at t = $\frac{k-1}{k\mu}$. The variance of the kth member of the family is $1/k\mu^2$. Thus, given the data of service distribution, we may approximate it by a member of Erlang family using a suitable value of k.

The expected number of customers in the queue L_q, expected waiting time before being served W_q, the expected number of customers in the system L_s and the expected time spent in the system W_s are given by

$$L_q = \frac{1+k}{2k}\frac{\lambda^2}{\mu(\mu-\lambda)}$$

$$W_q = \frac{1+k}{2k}\frac{\lambda}{\mu(\mu-\lambda)}$$

$$W_s = W_q + \frac{1}{\mu}$$

$$L_s = \lambda W_s$$

As k approaches to infinity the queueing system represents an arrival pattern which follows **Poisson Process with arrival rate and a constant service time** p. per customer. The expressions of L_q, W_q, W_s and L_s are given by

$$L_q = \frac{\lambda^2}{2\mu(\mu-\lambda)}$$

$$W_q = \frac{\lambda}{2\mu(\mu-\lambda)}$$

$$W_s = \frac{\lambda}{2\mu(\mu-\lambda)} + \frac{1}{\mu}$$

$$L_s = \frac{\lambda^2}{2\mu(\mu-\lambda)} + \frac{\lambda}{\mu}$$

Example 5

Repair of certain type of machine requires three steps to be completed sequentially. The time taken to perform each step follows an exponential distribution with mean $6\frac{2}{3}$ minutes and is independent of each other. The machine breakdown follows a Poisson process with a rate of 1 per 2 hours. Assuming that there is only one repairman find out

a) The expected idle time of a machine.

b) The average waiting time of a brokendown machine in a queue.

c) The expected number of brokendown machines in the queue..

d) The average number of machines which are ndt in operation.

Solution

This is an $M/E_k/1$ system with k=3, $\lambda = \frac{1}{2}$/hour. Since there are 3 phases, the total service system can be considered to have three exponential phases with each with mean $\frac{1}{3\mu} = 6\frac{2}{3} = \frac{20}{3}$ minutes.

Hence μ = 3/hour.

$$W_q = \frac{4}{6} \times \frac{\left(\frac{1}{2}\right)}{3\left(3-\frac{1}{2}\right)} = \frac{2}{45} \text{ hour} = 2 \text{ minutes } 40 \text{ seconds.}$$

a) $W_s = (\frac{2}{45} + \frac{1}{3})$ hour = 22 minutes 40 seconds.

b) W_q = 2 minutes 40 seconds.

c) $$L_q = \frac{4}{6} \times \frac{\left(\frac{1}{2}\right)^2}{3\left(3-\frac{1}{2}\right)} = \frac{1}{45}$$

d) $$L_s = \frac{1}{2} \times \frac{17}{45} = \frac{17}{90}$$

Activity 6

A booking counter takes 10 minutes to book a ticket for each customer. If the customers are arriving according to a Poisson process with a rate of 4 per hour, find out

a) Expected queue length.

b) Expected waiting time of a customer in the queue.

c) Expected time a customer spends in the system.

d) Expected number of customers in the system.

……
……
……

13.8 DECISION PROBLEMS IN QUEUEING

The study of queueing models helps us to find the ways and means for improving performance of the given system. A common viewpoint is to base the decision on a cost model which **minimises the sum of costs of service and of waiting time per unit time.** Generally, it is assumed that the cost of waiting is directly proportional to the total time that the customers spend in the system, both waiting and in service. The service can usually be ascertained from the various available records.

Consider a single server model with an arrival rate **A** and service rate μ. It is assumed that the service rate μ is controllable and it is required to determine its optimum value based on an appropriate model.

Let

C_i = cost per unit increase in p. per unit time.

C_2 = cost per unit time per person of waiting.

The total cost of average waiting and service per unit time is T(μ), where

$$T(\mu) = C_1\mu + C_2L_s$$

In an M/M/l : ∞/FIFO system it follows that

$$T(\mu) = C_1\mu + \frac{C_2\lambda}{\mu-\lambda}.$$

The optimum value of μ is obtained by minimising T(μ) with respect to p. by calculus methods. This value is given by

$$\mu = \lambda + \left(\frac{C_2\lambda}{C_1}\right)^{1/2}.$$

Example 6

A manufacturing process consists of two stages. The output of stage I (which is the input of stage II) is Poisson with average value k=100 units per day. The number of units which can be processed in stage II follows Poisson distribution with a rate of μ per day. The cost of unit increase in μ in stage II process is Rs. 600 per day. If a unit manufactured in stage I is not immediately processed in stage II, the cost of storing and other expenses are estimated as Rs. 6 per day per unit. What should be the optimum rate μ of manufacturing in stage II so that the total cost of waiting at the end of stage I and the cost of manufacturing of stag II is minimised?

Solution

Here, λ = 100, C_I = 600, C_2 = 6

The optimum value of μ is = $100 + \sqrt{\frac{6}{600}\times 100}$ = 101.

13.9 SUMMARY

A queueing problem is characterised by a flow of customers arriving randomly at one or more service facilities. The customers upon arriving at the facility may be served immediately or, if willing, may have to wait until the facility is available.

The arrivals are characterised either by the **number of arrivals** in a specific period of time or by the time between two successive arrivals, known as **interarrival time.** The number of arrivals is a **discrete random variable** whereas the interarrival times are **continuous random variables.**

The service is characterised either by the **number of services completed** in a given time period or the **time taken to complete the service.** The number of services completed is a **discrete random variable** while the service time is **a continuous random variable.**

Other characteristics of a queueing model are the **number of servers, order of service, size of the population of the customers, maximum size of the queue.** The various **characteristics** of a queueing problem can be expressed in terms of a suitable notation known as **Kendall's notation.**

The time spent by an incoming customer in a queueing system is a random variable and is of interest to the decision maker. If the customer under consideration is a down machine it is inoperative during the period of service. There is, thus, a cost involved due to its lack of functioning. One of the objectives of studying a queueing problem is to find out the optimum service rate and the number of servers so that **the average cost of being in queueing system and the cost of service are minimised.** The time a customer spends in a system before the start of service is a random variable known as **waiting time.** The probability distribution of waiting time depends upon the probability distribution of inter arrival time and service time.

Let $P_n(t)$ indicate the probability of having n customers in the system at time t. Then if $P_{,,}(t)$ depends upon t, the queueing system is said to be in the **transient** state. After the queueing system has become operative for a considerable period of time the probability $P_n(t)$ may become **independent of t.** The probabilities are then known as **steady state probabilities.**

When the number of arrivals in a time interval of length t follows a Poisson distribution with mean Jet where **A** is, the rate of arrival, the arrivals are said to follow a **Poisson Process.** In this case the interarrival times are independently, identically distributed random variables with an **exponential probability distribution with parameter X** and vice versa. If the service times are independently, identically distributed **exponential** random variables with **parameter** 11 and there is only one server in the queueing model, the resulting queueing model is denoted by M/M/1 according to Kendall's notation. If A<p., **the steady state probabilities exist** and P_n, the number of customers in the system follows a **geometric distribution** with parameter X/μ. The time spent by a customer in the system taking into account both waiting and service time is an **exponential distribution** with parameter μ-h. The probability distribution of the waiting time before service can also be derived in an identical manner.

In an M/M/C system, the probability distribution of arrivals and service are identical with those of M/M/I system but there are C servers who are offering service simultaneously. The steady state probabilities exist if $\frac{1}{c\mu}$<1. In an **M/E_k/l system,** the number of arrivals follow a Poisson process but the service times have k different phases each of which is exponentially distributed with a common parameter. For k = we obtain a queueing model where the number of arrivals follows a Poisson process and the **service time is constant.**

The optimum service rate of a queueing model can be determined by calculus method minimising the sum of costs of service and of waiting per unit time.

13.10 KEY WORDS

A Queueing Model is a suitable model to represent a service oriented problem where customers arrive randomly to receive some service, the service time being also a random variable.

The International Time is the time between two successive arrivals.

A Server is a person or a mechanism through which service is offered.

The Waiting Time of a customer is the time spent in a queueing system before the service starts.

The Time Spent by a customer in a queueing system is the sum of the waiting time and the service time.

The Queue Discipline is the order in which the members of the queue are offered service.

The FIFO is the first in first out queue discipline.

Kendall's Notation is a system of notation according to which the various characteristics of a queueing model are identified.

The Transient State of a queueing system is the state where who probability of the number of customers in the system depends upon time.

The Steady state to of a queueing system is the state where the probability of the number of ettstomers in the system is independent of t.

The Poisson Pam is a probabilistic phenomenon when the number of arrivals in an interval of length t follows a Poisson distribution with parameter At, where h is the rate of arrival.

The MIM/I Queueing Model is a queueing model where the arrivals follow a Poisson process, service times are exponentially distributed and there is one server.

The Gamma Distribution is the probability distribution of a random variable y such that y=Xt+X2+.....+X,, where X_is are independently, identically; exponentially distributed random variables.

The M/M/C Queueing Model is a queueing model where: the arrivals follow a Poisson *process, service times are exponentially* distributed and there are C servers.

The M/M/∞ Queueing Model is a queueing model where the customers are served immediately after arrival without any waiting when the arrivals follow a Poisson process and the service time is exponentially distributed.

An Erlang Family (E_K) of probability distribution is the probability distribution of a random variable which cap be expressed as the sum k independently, identically distributed exponential variables.

The M/E_k/l Queueing Model is a queueing model when the arrivals follow a Poisson process, service time follows an Erlang (k) probability distribution and the number of server is one.

13.11 SELF-ASSESSMENT EXERCISES

1) Customers at a box-office window, being manned by a single individual arrive according to a Poisson process with a rate of 30 per hour. The time taken to serve a customer has an exponential distribution with a mean of 90 seconds. Find the average waiting time of a customer.

2) In a bank operating from 10 AM to 2 PM the cheques are cashed in a single counter. Customers wishing to cash cheques arrive according to a Poisson process at the rate of 20 customers a day. The cashier at the counter takes on an average 10 minutes to ᶜash the cheque. The service time has been shown to be exponentially distributed.

   i) Compute the percentage of time the cashier is busy.

   ii) Compute the average time a customer is expected to wait.

   iii) Compute the average number of customers waiting in the queue.

3) Trucks arrive at a safety inspection station so that the interarrival times are exponentially distributed with a mean of. 113 hour. The times required for inspection are also exponentially distributed with a mean of 1/5 hour. Assume that the associated queueing system is in steady state.

Find out:

a) Expected number of trucks at the inspection station.
b) Expected number of trucks waiting to be inspected.
c) Expected time spent by a truck in the inspection station.
d) Expected waiting time of a truck before inspection.

4) Telephone calls arrive at an office with two operators. If both operators are busy, the calls are automatically held on queue to be answered on a first come first serve basis. The time spent by each operator in taking the call is exponentially distributed with a mean service time of 6 minutes. The calls have exponentially distributed interarrival times and occur at a mean rate of 15 per, hour.
   a) What is the expected number of calls held waiting for service?
   b) What is the expected time an arriving call will spend in the queneing system taking into account both waiting *and service time?*
   c) What fraction of time are both operators busy?
5) A petroleum company is considering the expansion of its one unloading facility at its refinery. The ships arrive according to a Poisson process to unload crude oil at a rate of 5 ships per week. The service is carried out according to an exponential probability distribution with a rate of 10 ships per week.

   The company has under consideration a second unloading berth which can be rented for Its. 5000 a weak. The service rate for this berth will also be 10 per week. For each week a ship remains idle waiting in line or for unloading the company loses Its., 20,000.
   a) What is the average time a ship must wait before beginning to deliver its cargo to the refinery?
   b) If a second berth is rented, what will be the average number of ships waiting?
   c) What will be the average waiting time of a ship with a second berth?
   d) Is the benefit of reduced waiting time' worth the rental cost of the second berth?
6) In a self-service facility arrivals occur according to a Poisson process with a rate of 5 per hour. Service time per customer is exponentially distributed with mean 5 minutes.
   a) Find the expected number of customers in service.
   b) What is the percentage of time the facility is idle?
7) A barber with a one-man shop takes exactly 25 minutes to complete one haircut. If customers arrive according to a Poisson process at a rate of one every 40 minutes, how long on the average must a customer wait for service?
8) Repair of a certain type of machine which breaks down in a factory consists of five basic steps that must be performed sequentially. The time taken to perform each of the five steps is found to have an exponential distribution with mean 5 minutes and is independent of the other steps. If these machines break down according to a Poisson process at a rate of 2 per hour and if there is only one repairman, what is the average idle time for each machine that haS broken down?

13.12 ANSWERS

Activity 1

$f(t) = \lambda e^{-\lambda t}$ $(t>0)$

$$E(t) = \lambda \int_0^\infty te^{-\lambda t}\, dt$$

$$\therefore \frac{E(t)}{\lambda} = \int_0^\infty te^{-\lambda t}\, dt = \left[\frac{-te^{-\lambda t}}{\lambda}\right]_0^\infty + \frac{1}{\lambda}\int_0^\infty e^{-\lambda t}\, dt$$

$$= 0 + \frac{1}{\lambda^2}\left[-e^{-\lambda t}\right]_0^\infty = \frac{1}{\lambda^2}$$

$$\therefore E(t) = \frac{1}{\lambda}$$

Activity 2

$P(X>t+s \mid X>s)$

$$= \frac{P(X>t+s;\ X>s)}{P(X>s)}$$

$$= \frac{P(X>t+s)}{P(X>s)} = e^{-\mu(t+s)+\mu s}$$

$= e^{-\mu t} = P(X>t).$

Activity 3

$$W(t) = \sum_{n=0}^{\infty} \frac{\mu(\mu t)^n}{n!} e^{-\mu t} \rho^n (1-\rho)$$

$$= (1-\rho)\ \mu\ e^{-\mu t} \sum_{n=0}^{\infty} \frac{(\mu t \rho)^n}{n!}$$

$$= (1-\rho)\ \mu e^{-\mu t(1-\rho)}$$

$$= (\mu-\lambda)\ e^{-(\mu-\lambda)t}$$

Activity 4

$= 21\frac{3}{7}$ per hour

Activity 5

Initially 12 minutes for the depositors and 7 minutes for withdrawals, combined 3.85 minutes.

Activity 6

$L_q = \frac{2}{3} W_q = 10$ minutes $W_s = 20$ minutes $L_s = 1\frac{1}{3}$.

Self-assessment Exercises

1) $4\frac{1}{2}$ minutes

2) (i) 83.33% (ii) 50 minutes (iii) $4\frac{1}{6}$.

3) (a) $L_s = \frac{3}{2}$ (b) $L_q = \frac{9}{10}$ (c) $W_s = \frac{1}{2}$ hour (d) $W_q = \frac{3}{10}$ hour.

4) (a) $L_q = \frac{27}{14}$ (b) $W_s = \frac{24}{105}$ hour (c) $\frac{9}{14}$.

5) a) Average waiting time $= \frac{1}{10}$ week.

   b) Average number of ships waiting $= \frac{1}{30}$

   c) Average waiting time of a ship $= \frac{1}{150}$ week

   d) Present cost per week = Rs. 10000
   Cost per week with two berths = Rs. $5666\frac{2}{3}$.
   Benefit is worthwhile.

6) a) $\frac{5}{12}$ b) 66%

7) 20.8 minutes

8) 96 minutes.

13.13 FURTHER READINGS

Kendall, M. 1951.. *Some Problems in the Theory of Queues, JRSS (B),* Volume 13, No. 2, 117-139.

Saaty, T.L. 1961. *Elements of Queueing Theory with Applications,* McGraw Hill New York.

Sasieni, M.A. Yaspan and L. Friedman, 1959. *Operations Research : Methods and Problems,* J. Wiley and Sons : New York.

Taha, H.A. 1971. *Operations Research an Introduction,* The Macmillan Company New York.