23

QUEUEING THEORY
(Waiting Line Models)

23.1. INTRODUCTION

In everyday life, it is seen that a number of people arrive at a cinema ticket window. If the people arrive "too frequently" they will have to wait for getting their tickets or sometimes do without it. Under such circumstances, the only alternative is to form a queue, called the ***waiting line,*** in order to maintain a proper discipline. Occasionally, it also happens that the person issuing tickets will have to wait, (*i.e.* remains idle), until additional people arrive. Here the arriving people are called the ***customers*** and the person issuing the tickets is called a ***server.***

Another example is represented by letters arriving at a typist's desk. Again, the letters represent the *customers* and the typist represents the *server.* A third example is illustrated by a machine breakdown situation. A broken machine represents a *customer* calling for the service of a repairman. These examples show that the term *customer* may be interpreted in various number of ways. It is also noticed that a service may be performed either by moving the *server* to the *customer* or the *customer* to the *server.*

Thus, it is concluded that waiting lines are not only the lines of human beings but also the aeroplanes seeking to land at busy airport, ships to be unloaded, machine parts to be assembled, cars waiting for traffic lights to turn green, customers waiting for attention in a shop or supermarket, calls arriving at a telephone switch-board, jobs waiting for processing by a computer, or anything else that require work done on and for it are also the examples of costly and critical delay situations. Further, it is also observed that arriving units may form one line and be serviced through only one station (as in a doctor's clinic), may form one line and be served through several stations (as in a barber shop), may form several lines and be served through as many stations (*e.g.* at check out counters of supermarket).

Servers may be in parallel or in series. When in parallel, the arriving customers may form a single queue as shown in *Fig.* 23.1 or individual queues in front of each server as is common in big post-offices. Service times may be constant or variable and customers may be served singly or in batches (like passengers boarding a bus).

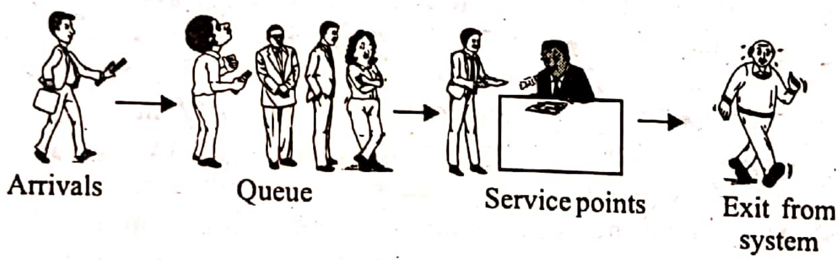


Fig. 23.1 (a). Queueing system with single queue and single service station..

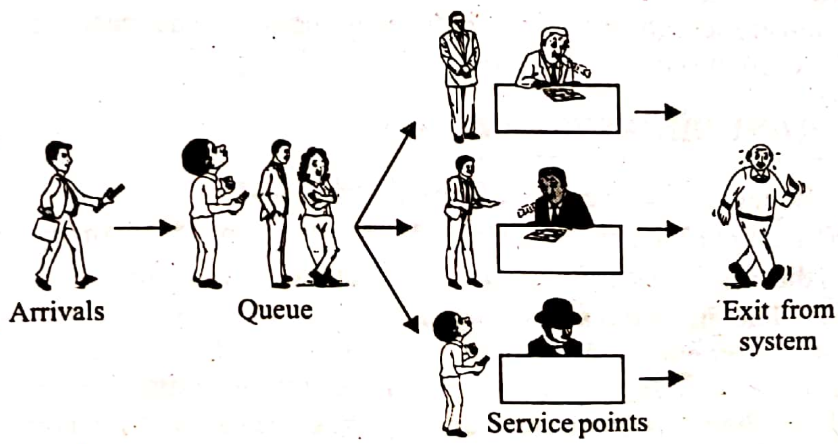


Fig. 23.1 (b). Queueing system with single queue and several service stations.

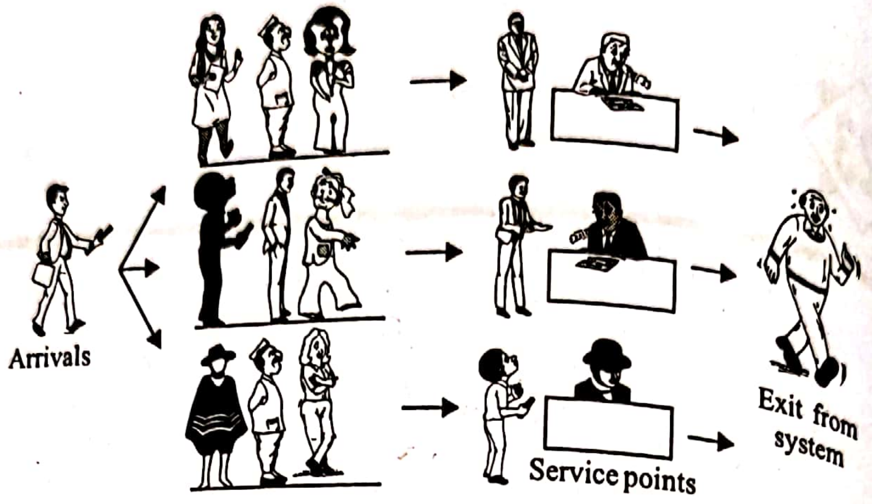


Fig. 23.1 (c). Queueing system with several queues and several service

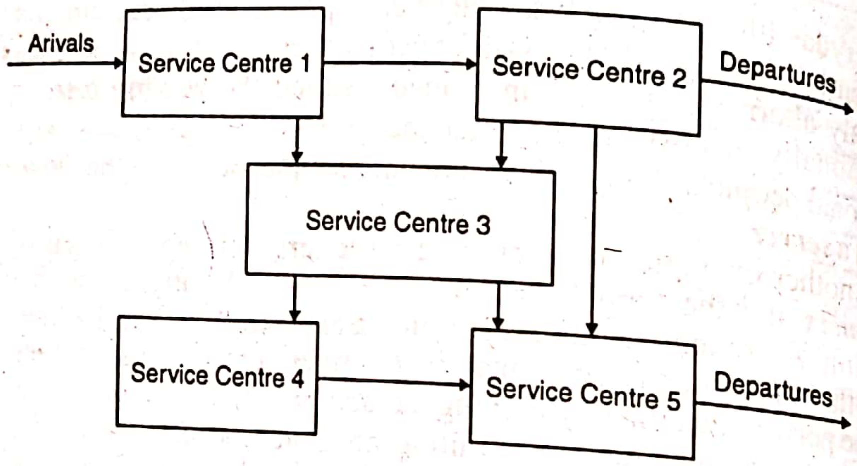


Fig. 23.2. A machine shop as a complex queue.

Fig. 23.2 illustrates how a machine shop may be thought of as a system of queues forming in front of a number of service centres, the arrows between the centres indicating possible routes for jobs processed in the shop. Arrivals at a service centre are either new jobs coming into the system or jobs, partially processed, from some other service centre. Departures from a service centre may become the arrivals at another service centre or may leave the system entirely, when processing on these items is complete.

Queueing theory is concerned with the statistical description of the behaviour of queues with finding, *e.g.*, the probability distribution of the number in the queue from which the mean and variance of queue length and the probability distribution of waiting time for a customer, or the distribution of a server's busy periods can be found. In operational research problems involving queues, investigators must measure the existing system to make an objective assessment of its characteristics and must determine how changes may be made to the system, what effects of various kinds of changes in the system's characteristics would be, and whether, in the light of the costs incurred in the systems, changes should be made to it. A model of the queueing system under study must be constructed in this kind of analysis and the results of queueing theory are required to obtain the characteristics of the model and to assess the effects of changes, such as the addition of an extra server or a reduction in mean service time.

Perhaps the most important general fact emerging from the theory is that the degree of congestion in a queueing system (measured by mean wait in the queue or mean queue length) is very much dependent on the amount of irregularity in the system. Thus congestion depends not just on mean rates at which customers arrive and are served and may be reduced without altering mean rates by regularizing arrivals or service times, or both where this can be achieved.

23.2. QUEUEING SYSTEM

A queueing system can be completely described by
(*a*) the input (or arrival pattern), (*b*) the service mechanism (or service pattern),
(*c*) the 'queue discipline' and (*d*) customer's behaviour.

(a) The input (or arrival pattern). The input describes the way in which the customers arrive and join the system. Generally, the customers arrive in a more or less random fashion which is not worth making the prediction. Thus, the arrival pattern can best be described in terms of probabilities and consequently the probability distribution for inter-arrival times (the time between two successive arrivals) or the distribution of number of customers arriving in unit time must be defined.

The present chapter is only dealt with those queueing systems in which the customers arrive in *'Poisson'* or *'completely random'* fashion (see *sec. 23.7-1*). Other types of arrival pattern may also be observed in practice that have been studied in queueing theory. Two such patterns are observed, where

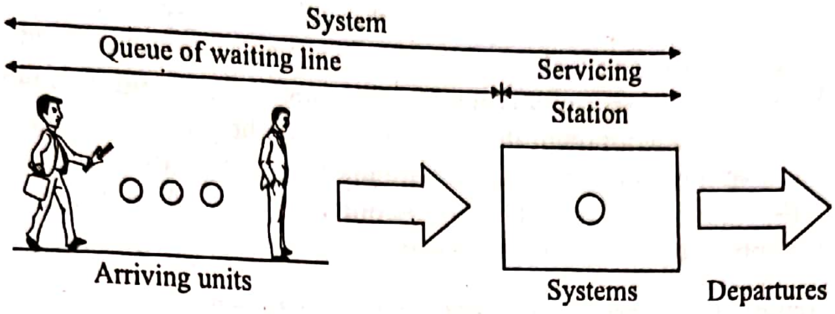


Fig. 23.3. A queueing system with single service station.

(*i*) arrivals are of regular intervals;

(*ii*) there is general distribution (perhaps normal) of time between successive arrivals.

(b) The service mechanism (or service pattern). It is specified when it is known how many customers can be served at a time, what the statistical distribution of service time is, and when service is available. It is true in most situations that service time is a random variable with the same distribution for all arrivals, but cases occur where there are clearly two or more classes of customers (*e.g.* machines waiting repair) each with a different service time distribution. Service time may be constant or a random variable. Distributions of service time which are important in practice are ***'negative exponential distribution'*** and the related ***'Erlang (Gamma) distribution'***. Queues with the negative exponential service time distribution are studied in the following sections.

In the present chapter, only those queueing systems are discussed in which the service time follows the *'Exponential and Erlang (Gamma)'* probability distributions (see ***sec. 23.7-1*** **to** ***23.7-8***).

(c) The queue discipline. The queue discipline is the rule determining the formation of the queue, the manner of the customer's behaviour while waiting, and the manner in which they are chosen for service. The simplest discipline is ***"first come, first served"***, according to which the customers are served in the order of their arrival. For example, such type of queue discipline is observed *at a ration shop, at cinema ticket windows, at railway stations,* etc. If the order is reversed, we have the ***"last come, first served"*** discipline, as in the case of a big godown the items which come last are taken out first. An extremely difficult queue discipline to handle might be ***"service in random order"*** or ***"might is right"***.

Properties of a queueing system which are concerned with waiting times, in general, depend on queue discipline. For example, the variance of waiting time will be much greater with the queue discipline ***'first come, last served'*** than with ***'first come, first served'***, although mean waiting time will remain unaffected.

The following notations are used for describing the nature of service discipline.

FIFO → First In, First Out or *FCFS* → First Come, First Served

LIFO → Last In, First Out or *FILO* → First In, Last Out.

SIRO → Service in Random Order

This chapter shall be concerned only with the customers which are served in the order in which they arrive at the service facility, that is, ***'first come, first served'*** discipline.

(d) Customer's behaviour. The customers generally behave in four ways :

(*i*) **Balking.** A customer may leave the queue because the queue is too long and he has no time to wait, or there is not sufficient waiting space.

(*ii*) **Reneging.** This occurs when a waiting customer leaves the queue due to impatience.

(*iii*) **Priorities.** In certain applications some customers are served before others regardless of their order of arrival. These customers have *priority* over others.

(*iv*) **Jockeying.** Customers may *jockey* from one waiting line to another. It may be seen that this occurs in the supermarket.

(e) 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 queuing model. Howerver, this assumption is not valid where the customer group is represented by few machines in workshop that require operator facility from time to time. If the population size is finite then the analysis of queueing model becomes more involved.

(f) Maximum Length 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 have 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.

Q. 1. Explain briefly the main characteristics of queueing system.
2. Describe the fundamental components of a queueing process and give suitable examples. [C.A. (Nov) 92]
3. List the factors that constitute the basic elements of a queueing model. For each of these enumerate the alternatives posible. Represent this diagramatically to cover all possible implimentations of a queueing model. [IGNOU 99 (Dec.)]

23.3. QUEUEING PROBLEM

In a specified queueing system, the problem is to determine the following :

(a) Probability distribution of queue length. When the nature of probability distributions of the arrival and service patterns is given, the probability distribution of queue length can be obtained. Further, we can also estimate the probability that there is no queue.

(b) Probability distribution of waiting time of customers. We can find the time spent by a customer in the queue before the commencement of his service which is called his ***waiting time.*** The total time spent by him in the system is the waiting time plus service time.

(c) The busy period distribution. We can estimate the probability distribution of busy periods. If we suppose that the server is free initially and customer arrives, he will be served immediately. During his service time, some more customers will arrive and will be served in their turn. This process will continue in this way until no customer is left unserved and the server becomes free again. Whenever this happens, we say that a **busy period** has just ended. On the other hand, during **idle periods** no customer is present in the system. A busy period and the idle period following it together constitute a *busy cycle.* The study of the busy period is of great interest in cases where technical features of the server and his capacity for continuous operations must be taken into account.

23.4. TRANSIENT AND STEADY STATES

Queueing theory analysis involves the study of a system's behaviour over time. ***A system is said to be in "transient state" when its operating characteristics (behaviour) are dependent on time.*** This usually occurs at the early stages of the operation of the system where its behaviour is still dependent on the initial conditions. However, since we are mostly interested in the *"long run"* behaviour of the system, mainly the attention has been paid toward *"steady state"* results.

A steady state condition is said to prevail when the behaviour of the system becomes independent of time. Let $P_n(t)$ denote the probability that there are n units in the system at time t. In fact, the change of $P_n(t)$ with respect to t is described by the derivative $[dP_n(t)/dt]$ or $P_n'(t)$. Then the queueing system is said to become *'stable'* eventually, in the sense that the probability $P_n(t)$ is independent of time, that is, remains the same as time passes ($t \to \infty$). Mathematically, in steady state

$$\lim_{t \to \infty} P_n(t) = P_n \text{ (independent of } t) \Rightarrow \lim_{t \to \infty} \frac{dP_n(t)}{dt} = \frac{dP_n}{dt} \Rightarrow \lim_{t \to \infty} P_n'(t) = 0.$$

In some situations, if the arrival rate of the system is larger than its service rate, a steady state cannot be reached regardless of the length of the elapsed time. In fact, in this case the queue length will increase with time and theoretically it could build up to infinity. Such case is called the ***"explosive state".***

In this chapter, only the steady state analysis will be considered. We shall not treat the ***'transient'*** and *'explosive'* states.

Q. 1. What is queueing problem ? Explain queueing system, transient and steady state. [Garhwal M.Sc. (Stat.) 96]
2. What is a queueing theory problem ? Describe the advantages of queueing theory to a business executive with a view to persuading him to make use of the same in management. [Garhwal M.Sc. (Stat.) 95]
3. What do you understand by a queue ? Give some important applications of queueing theory. [Garhwal M.Sc. (Stat.) 92, 91]
4. Write an essay on various characteristics of a queueing system. [Virbhadra 2000]

23.5. A LIST OF SYMBOLS

Unless otherwise stated, the following symbols and terminology will be used henceforth in connection with the queueing models. The reader is reminded that a queueing system is defined to include the *queue* and the *service stations* both. (see *Fig.* 23.3).

n = number of units in the system

$P_n(t)$ = transient state probability that exactly n calling units are in the queueing system at time t

E_n = the state in which there are n calling units in the system

P_n = steady state probability of having n units in the system

λ_n = mean arrival rate (expected number of arrivals per unit time) of customers (when n units are present in the system)

μ_n = mean service rate (expected number of customers served per unit time when there are n units in the system)

λ = mean arrival rate when λ_n is constant for all n

μ = mean service rate when μ_n is constant for all $n \geq 1$

s = number of parallel service stations

$\rho = \lambda/\mu s$ = traffic intensity (or utilization factor) for servers facility, that is, the expected fraction of time the servers are busy

$\phi_T(n)$ = probability of n services in time T, given that servicing is going on throughout T

Line length (or queue size)
= number of customers in the queueing system

Queue length
= line length (or queue size)–(number of units being served)

$\Psi(w)$ = probability density function (p.d.f.) of waiting time in the system

L_s = expected line length, *i.e.*, expected number of customers in the *system*

L_q = expected queue length, *i.e.*, expected number of customers in the *queue*

W_s = expected waiting time per customer in the system

W_q = expected waiting time per customer in the queue

$(W \mid W > 0)$ = expected waiting time of a customer who has to wait

$(L \mid L > 0)$ = expected length of non-empty queues, *i.e.*, expected number of customers in the queue when there is a queue

$P(W > 0)$ = probability of a customer having to wait for service

$\binom{n}{r}$ = the binomial coefficient ${}^nC_r$.

$$= \frac{n!}{r!\,(n-r)!} = \frac{n\,(n-1)\ldots(n-r+1)}{r!} \quad \text{for } r \text{ and } n \text{ non-negative integers } (r \leq n).$$

23.6. TRAFFIC INTENSITY (OR UTILIZATION FACTOR)

An important measure of a simple queue $(M \mid M \mid 1)$ is its *traffic intensity*, where

$$\text{Traffic intensity } (\rho) = \frac{\text{mean arrival rate}}{\text{mean service rate}} = \frac{\lambda}{\mu}$$

i.e.,
$$\rho = \frac{1/\mu}{1/\lambda} = \frac{\text{Mean service time}}{\text{Mean inter–arrival time}}$$

The unit of traffic intensity is *Erlang*.

Here it should be noted carefully that a necessary condition for a system to have settled down to steady state is that $\rho < 1$ or $\lambda/\mu < 1$ or $\lambda < \mu$, *i.e.*, *arrival rate < service rate*.

If this is not so, *i.e.*, $\rho > 1$, the arrival rate will be greater than the service rate and consequently, the number of units in the queue tends to increase indefinitely as the time passes on, provided the rate of service is not affected by the length of queue.

23.7. PROBABILITY DISTRIBUTIONS IN QUEUEING SYSTEMS

The arrival pattern of customers at a queueing system varies between one system and another, but one pattern of common occurrence in practice, which turns out to be relatively easy to deal with mathematically, is that of *'completely random arrivals'*. This phrase means something quite specific, and we discuss what does it mean before dealing in the subsequent sections with a variety of queueing systems. In particular, we show that, if arrivals are *'completely random'*, the number of arrivals in unit time has a *Poisson distribution*, and the intervals between successive arrivals are distributed *negative exponentially*.

23.7-1. Distribution of Arrivals 'The Poisson Process' (Pure Birth Process)

In many situations the objective of an analysis consists of merely observing the number of customers that enter the system. The model in which only arrivals are counted and no departures take place are called *pure birth models*. The term *'birth'* refers to the arrival of a new calling unit in the system, and the *'death'* refers to the departure of a served unit. As such *pure birth* models are not of much importance so far as their applicability to real life situation is concerned, but these are very important in the understanding of completely random arrival problems.

Theorem 23.1. (*Arrival Distribution Theorem*). *If the arrivals are completely random, then the probability distribution of number of arrivals in a fixed time-interval follows a Poisson distribution.*

[Agra 99; Meerut (Stat.) 98; Garhwal M.Sc. (Math.) 94; Raj. Univ. (Math) 93]

Proof. In order to derive the arrival distribution in queues, we make the following three assumptions (sometimes called the ***axioms***).

1. Assume that there are n units in the system at time t, and the probability that exactly one arrival (birth) will occur during small time interval Δt be given by $\lambda \Delta t + O(\Delta t)$, where λ is the arrival rate independent of t and $O(\Delta t)$ includes the terms of higher order of Δt.
2. Further assume that the time Δt is so small that the probability of more than one arrival in time Δt is $O(\Delta t)^2$, *i.e.*, almost zero.
3. The number of arrivals in non-overlapping intervals are statistically independent, *i.e.*, the process has independent increments.

We now wish to determine the probability of n arrivals in a time interval of length t, denoted by $P_n(t)$. Clearly, n will be an integer greater than or equal to zero. To do so, we shall first develop the differential-difference equations governing the process in two different situations.

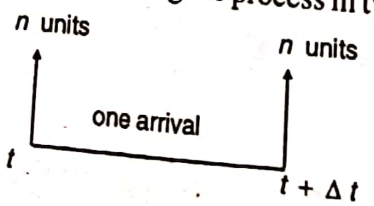


Fig. 23.4.

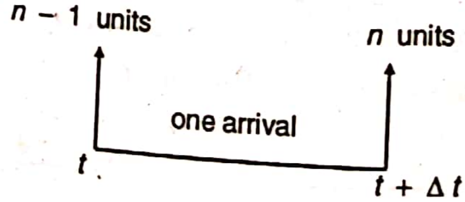


Fig. 23.5.

Case I. ***When n > 0.*** For $n > 0$, there may be two mutually exclusive ways of having n units at time $t + \Delta t$.

(*i*) There are n units in the system at time t and no arrival takes place during time interval Δt. Hence, there will be n units at time $t + \Delta t$ also. This situation is better explained in *Fig. 23.4*.

Therefore, the probability of these two combined events will be

$= $ Prob. of n units at time $t \times$ Prob. of no arrival during $\Delta t = P_n(t) \cdot (1 - \lambda \Delta t)$ ...(23.1)

[since prob. of exactly one arrival in $\Delta t = \lambda \Delta t$, prob. of no arrival becomes $= 1 - \lambda \Delta t$.]

(*ii*) *Alternately*, there are $(n-1)$ units in the system at time t, and one arrival takes place during Δt. Hence there will remain n units in the system at time $t + \Delta t$. This situation is better explained in *Fig. 23.5*.

Therefore, the probability of these two combined events will be

$= $ Prob. of $(n-1)$ units at time $t \times$ Prob. of one arrival in time $\Delta t = P_{n-1}(t) . \lambda \Delta t$ ...(23.2)

Note. Since the probability of more than one arrival in Δt is assumed to be negligible, other alternatives do not exist.

Now, adding above two probabilities [given by (23.1) and (23.2)], we get the probability of n arrivals at time $t + \Delta t$, *i.e.*

$$P_n(t + \Delta t) = P_n(t)\,(1 - \lambda \Delta t) + P_{n-1}(t)\,\lambda \Delta t \qquad ...(23.3)$$

Case 2. *When n = 0.*

$$P_0(t + \Delta t) = \text{Prob. } [\textit{no unit at time } t] \times \text{Prob. } [\textit{no arrival in time } \Delta t]$$

$$\therefore \quad P_0(t + \Delta t) = P_0(t)\,(1 - \lambda \Delta t). \qquad ...(23.4)$$

Rewriting the equations (23.3) and (23.4) after transposing the terms $P_n(t)$ and $P_0(t)$ to left hand sides, respectively, we get

$$P_n(t + \Delta t) - P_n(t) = P_n(t)\,(-\lambda \Delta t) + P_{n-1}(t)\,\lambda \Delta t, \qquad n > 0 \qquad ...(23.3)'$$

$$P_0(t + \Delta t) - P_0(t) = P_0(t)\,(-\lambda \Delta t) \qquad n = 0 \qquad ...(23.4)'$$

Dividing both sides by Δt and then taking limit as $\Delta t \to 0$,

$$\lim_{\Delta t \to 0} \frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = -\lambda P_n(t) + \lambda\, P_{n-1}(t) \qquad ...(23.5)$$

$$\lim_{\Delta t \to 0} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda\, P_0(t) \qquad ...(23.6)$$

Since by definition of first derivative, $\lim_{\Delta t \to 0} \frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = \frac{d\,P_n(t)}{dt} = P_n'(t)$,

the equations (23.6) and (23.5) respectively can be written as

$$P_0'(t) = -\lambda P_0(t), \qquad n = 0 \qquad ...(23.7)$$

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t), \qquad n > 0 \qquad ...(23.8)$$

This is known as the *system of differential-difference equations.*

To solve the equations (23.7) and (23.8) by iterative method :

Equation (23.7) can be written as

$$\frac{P_0'(t)}{P_0(t)} = -\lambda \quad \text{or} \quad \frac{d}{dt}[\log P_0(t)] = -\lambda \qquad ...(23.9)$$

Integrating both sides w.r.t. 't',

$$\log P_0(t) = -\lambda t + A \qquad ...(23.10)$$

The constant of integration can be determined by using the boundary conditions :

$$P_n(0) = \begin{cases} 1 & \text{for} \quad n = 0 \\ 0 & \text{for} \quad n > 0. \end{cases}$$

Substituting $t = 0$, $P_0(0) = 1$ in (23.10), find $A = 0$. Thus, (23.10) gives

$$\log P_0(t) = -\lambda t \quad \text{or} \quad P_0(t) = e^{-\lambda t} \qquad ...(23.11)$$

Putting $n = 1$ in (23.8),

$$P_1'(t) = -\lambda P_1(t) + \lambda P_0(t)$$

or

$$P_1'(t) + \lambda P_1(t) = \lambda e^{-\lambda t}. \qquad ...(23.12)$$

Since this is the linear differential equation of first order, it can be easily solved by multiplying both sides of this equation by the integrating factor, I.F. $= e^{\int \lambda dt} = e^{\lambda t}$.

Thus, eqn. (23.12) becomes

$$e^{\lambda t}\,[P_1'(t) + \lambda P_1(t)] = \lambda \quad \text{or} \quad \frac{d}{dt}[e^{\lambda t}\,P_1(t)] = \lambda$$

Now integrating both sides w.r.t. 't'
$$e^{\lambda t} P_1(t) = \lambda t + B, \quad \text{...(23.13)}$$
where B is the constant of integration.

In order to determine the constant B, put $t = 0$ in (23.13), and get
$$P_1(0) = 0 + B \text{ or } B = 0 \quad [\because P_1(0) = 0]$$

Substituting $B = 0$ in (23.13),
$$P_1(t) = \frac{\lambda t\, e^{-\lambda t}}{1\,!} \quad \text{...(23.14)}$$

Similarly, putting $n = 2$ in (23.8) and using the result (23.14), we get the equation
$$P_2'(t) + \lambda P_2(t) = \lambda \frac{(\lambda t)\, e^{-\lambda t}}{1\,!} \quad \text{or} \quad \frac{d}{dt}[e^{\lambda t} P_2(t)] = \frac{\lambda\,(\lambda t)}{1\,!}.$$

Integrating w.r.t. 't'
$$e^{\lambda t} P_2(t) = \frac{(\lambda t)^2}{2\,!} + C,$$

Put $t = 0$, $P_2(0) = 0$ to obtain $C = 0$. Hence
$$P_2(t) = \frac{(\lambda t)^2\, e^{-\lambda t}}{2\,!}, \text{ for } n = 2 \quad \text{...(23.15)}$$

Similarly, obtain
$$P_3(t) = \frac{(\lambda t)^3\, e^{-\lambda t}}{3\,!}, \text{ for } n = 3$$

Likewise, in general,
$$P_m(t) = \frac{(\lambda t)^m\, e^{-\lambda t}}{m\,!} \text{ for } n = m. \quad \text{...(23.16)}$$

If, anyhow, it can be proved that the result (23.16) is also true for $n = m + 1$, then by induction hypothesis result (23.16) will be true for general value of n.

To do so, put $n = m + 1$ in (23.8) and get
$$P'_{m+1}(t) + \lambda P_{m+1}(t) = \lambda \frac{(\lambda t)^m\, e^{-\lambda t}}{m\,!} \quad \text{[using the results (23.16)]}$$
or
$$\frac{d}{dt}[e^{\lambda t} P_{m+1}(t)] = \frac{(\lambda t)^m\,(\lambda)}{m\,!}.$$

Integrating both sides,
$$e^{\lambda t} P_{m+1}(t) = \frac{(\lambda t)^{m+1}}{(m+1)\, m\,!} + D,$$

Again, putting $t = 0$, $P_{m+1}(0) = 0$, we get $D = 0$. Therefore,
$$\therefore \quad P_{m+1}(t) = \frac{(\lambda t)^{m+1}\, e^{-\lambda t}}{(m+1)\,!}.$$

Hence, in general,
$$P_n(t) = \frac{(\lambda t)^n\, e^{-\lambda t}}{n\,!}, \quad \text{...(23.17)}$$
which is a ***Poisson distribution formula.*** This completes the proof of the theorem.

Note. After carefully understanding the above procedure, the students can much reduce the number of steps by solving the differential equation of the standard from : $y' + P(x)\, y = Q(x)$, using the formula
$$y \,.\, e^{\int P dx} = \int Q(x)\,(e^{\int P dx})\, dx + C,$$
where $e^{\int P dx}$ is the integrating factor (I.F.)

Alternative Method : *Generating Function Technique.*

The system of equations (23.7) and (23.8) is
$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n > 0 \quad \text{...(i)}$$
$$P_0'(t) = -\lambda P_0(t), \quad n = 0 \quad \text{...(ii)}$$

We define the generating function of $P_n(t)$ as, $P(z, t) = \sum_{n=0}^{\infty} P_n(t)\, z^n$. Also, $P'(z, t) = \sum_{n=0}^{\infty} P_n'(t)\, z^n$.

Multiplying both sides of (i) by z^n and taking summation for $n = 1, 2, ..., \infty$, we get

$$\sum_{n=1}^{\infty} z^n P_n'(t) = -\lambda \sum_{n=1}^{\infty} z^n P_n(t) + \lambda \sum_{n=1}^{\infty} P_{n-1}(t) z^n \quad \text{...(iii)}$$

Now adding (ii) and (iii), we get

$$\sum_{n=0}^{\infty} z^n P_n'(t) = -\lambda \sum_{n=0}^{\infty} z^n P_n(t) + \lambda \sum_{n=0}^{\infty} z^{n+1} P_n(t)$$

or
$$P'(z, t) = -\lambda P(z, t) + \lambda z P(z, t) \quad \text{or} \quad \frac{P'(z, t)}{P(z, t)} = \lambda (z-1)$$

or
$$\frac{d}{dt} [\log P(z, t)] = \lambda (z-1)$$

Integrating both sides, $\log P(z, t) = \lambda (z-1) t + E.$ ...(iv)

To determine E, we put $t = 0$ to get $\log P(z, 0) = E$

But,
$$P(z, 0) = \sum_{n=0}^{\infty} z^n P_n(0) = P_0(0) + \sum_{n=1}^{\infty} z^n P_n(0)$$

$$= 1 + 0 = 1 \quad (\because P_0(0) = 1, \text{ and } P_n(0) = 0 \text{ for } n > 0)$$

Therefore, $E = \log P(z, 0) = \log 1 = 0.$

$\therefore$ eqn. (iv) becomes, $\log P(z, t) = \lambda(z-1)t$ or $P(z, t) = e^{\lambda(z-1)t}$.

Now, $P_n(t)$ can be defined as $P_n(t) = \frac{1}{n!} \left[\frac{d^n P(z, t)}{dz^n} \right]_{t=0}$

Using this formula, $P_0(t) = [P(z, t)]_{z=0} = e^{-\lambda t}$

$$P_1(t) = \left[\frac{d P(z, t)}{dz} \right]_{t=0} = [e^{\lambda(z-1)t} \lambda t]_{z=0} = \frac{e^{-\lambda t} \lambda t}{1!}$$

$$P_2(t) = \frac{1}{2!} \left[\frac{d^2 P(z, t)}{dz^2} \right]_{t=0} = \frac{(\lambda t)^2 e^{-\lambda t}}{2!}$$

... ...

In general, $$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}. \quad \text{...(v)}$$

Thus the probability of n arrivals in time 't' follows the ***Poisson law*** given by eqn. (v).

Q. 1. Show that 'n' the number of arrivals in a queue in time t follows the Poisson distribution, stating the assumptions clearly.

2. Show that the distribution of the number of births up to time 'T' in a simple birth process follows the Poisson law.

3. What do you understand by a queue ? Give some applications of queueing theory.

4. Explain what do you mean by Poisson process. Derive the Poisson distribution, given that the probability of single arrival during a small time interval Δt is $\lambda \Delta t$ and that of more than one arrival is negligible.

[JNTU (B. Tech.) 2002; Meerut (Maths.) 96]

23.7-2. Properties of Poisson Process of Arrivals

It has already been derived that–*if n be the number of arrivals during time interval t, then the law of probability in Poisson process is given by*

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad n = 0, 1, \ldots, \infty \quad \text{...(23.18)}$$

where λt is the parameter.

(1) Since mean $E(n) = \lambda t$, and var. $(n) = \lambda t$, ...(23.19)

the average (expected) number of arrivals in unit time will be

$$E(n)/t = \lambda = \text{mean arrival rate (or input rate).}$$

(2) If we consider the time interval $(t, t + \Delta t)$, where Δt is sufficiently small, then

$$P_0(\Delta t) = \text{Prob } [\textit{no arrival in time } \Delta t]$$

Putting $n = 0$ and $t = \Delta t$ in (23.18)

$$P_0(\Delta t) = \frac{e^{-\lambda \Delta t}}{0\,!} = e^{-\lambda \Delta t} = 1 - \lambda \Delta t + \frac{(\lambda \Delta t)^2}{2\,!} - \ldots = 1 - \lambda \Delta t + O(\Delta t)$$

where the term $O(\Delta t)$ indicates a quantity that is negligible compared to Δt. More precisely, $O(\Delta t)$ represents any function of Δt such that

$$\lim_{\Delta t \to 0} \frac{O(\Delta t)}{\Delta t} = 0.$$

[For example, $(\Delta t)^2$ can be replaced by $O(\Delta t)$ because $\lim_{\Delta t \to 0} \frac{(\Delta t)^2}{\Delta t} = 0$. This notation will be very useful for summarizing the negligible terms which do not enter in the final result]

$$\therefore \quad P_0(\Delta t) = 1 - \lambda \Delta t \qquad \ldots(23.20)$$

which means that the probability of no arrival in Δt is $1 - \lambda \Delta t$. In the similar fashion, $P_1(\Delta t)$ can be written as

$$P_1(\Delta t) = \frac{(\lambda \Delta t)\, e^{-\lambda \Delta t}}{1\,!} \qquad [\text{putting } n = 1,\ t = \Delta t \text{ in (23.18)}]$$

$$= \lambda \Delta t \left[1 - \lambda \Delta t + \frac{(\lambda \Delta t)^2}{2\,!} - \ldots \right] \qquad = \lambda \Delta t + O(\Delta t).$$

Neglecting the term $O(\Delta t)$,
$$P_1(\Delta t) = \lambda \Delta t\,, \qquad \ldots(23.21)$$
which means that the probability of one arrival in time Δt is $\lambda \Delta t$.

Similarly,
$$P_2(\Delta t) = \frac{(\lambda \Delta t)^2\, e^{-\lambda \Delta t}}{2\,!} = (\lambda \Delta t)^2 \left[1 - \lambda \Delta t + \frac{(\lambda \Delta t)^2}{2\,!} - \ldots \right] = O(\Delta t).$$

Again neglecting the term $O(\Delta t)$, we have
$$P_2(\Delta t) = 0, \qquad \ldots(23.22)$$
ans so on. Thus, it is concluded from the property of Poisson process that the probability of more than one arrival in time Δt is negligibly small, provided the terms of second and higher order of Δt are considered to be negligibly small. Symbolically,

$$P_n(\Delta t) = \text{negligibly small for all } n > 1. \qquad \ldots(23.23)$$

23.7-3. Distribution of Inter-Arrival Times (Exponential Process)

Let T be the time between two consecutive arrivals (called the inter-arrival time), and $a(T)$ denotes the probability density function of T. Then the following important theorem can be proved.

Theorem 23.2. *If n, the number of arrivals in time t, follows the Poisson distribution,*

$$P_n(t) = (\lambda t)^n\, e^{-\lambda t}/n\,!\,, \qquad \ldots(23.24)$$

then T (the inter-arrival time) obeys the negative exponential law

$$a(T) = \lambda e^{-\lambda T} \qquad \ldots(23.25)$$

and vice-versa.

[Kanpur 2000; Garhwal M.Sc. (Stat.) 95; Raj .Univ. (M.Phil) 91]

Proof. Suppose that t_0 = instant of an arrival initially.

Since there is no arrival in the intervals $(t_0, t_0 + T)$ and $(t_0 + T, t_0 + T + \Delta T)$, therefore $(t_0 + T + \Delta T)$ will be the instant of subsequent arrival.

Therefore, putting $t = T + \Delta T$ and $n = 0$ in (5.24),

$$P_0(T + \Delta T) = \frac{[\lambda(T + \Delta T)]^0 \,.\, e^{-\lambda(T + \Delta T)}}{0\,!} = e^{-\lambda(T + \Delta T)}$$

$$= e^{-\lambda T} \,.\, e^{-\lambda \Delta T} = e^{-\lambda T}\, [1 - \lambda \Delta T + O(\Delta T)]$$

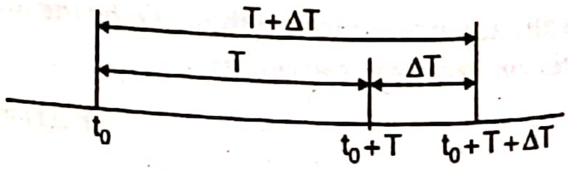


Fig. 23.6

Since $P_0(T) = e^{-\lambda T}$ from (23.24),

$$P_0(T + \Delta T) = P_0(T)\, [1 - \lambda \Delta T + O(\Delta T)]$$

or
$$P_0(T + \Delta T) - P_0(T) = P_0(T)\, [-\lambda \Delta T + O(\Delta T)].$$

Dividing both sides by ΔT,

$$\frac{P_0(T+\Delta T)-P_0(T)}{\Delta T} = -\lambda P_0(T) + \frac{O(\Delta T)}{\Delta T} P_0(T)$$

Now taking limit on both sides as $\Delta T \to 0$,

$$\lim_{\Delta T \to 0} \frac{P_0(T+\Delta T)-P_0(T)}{\Delta T} = \lim_{\Delta T \to 0}\left[-\lambda P_0(T) + \frac{O(\Delta T)}{\Delta T} P_0(T)\right]$$

or

$$\frac{dP_0(T)}{dT} = -\lambda P_0(T). \left[\text{since } \lim_{\Delta T \to 0} \frac{O(\Delta T)}{\Delta T} = 0\right] \quad ...(23.26)$$

But, L.H.S. of (23.26) is denoting the *probability density function* of T, say $a(T)$. Therefore,

$$*a(T) = **\lambda P_0(T). \quad \text{(see footnote)} \quad ...(23.27)$$

But, from equation (23.24), $P_0(T) = e^{-\lambda T}$. Putting this value of $P_0(T)$ in (23.27),

$$a(T) = \lambda e^{-\lambda T} \quad ...(23.28)$$

which is the *exponential law of probability* for T with mean $1/\lambda$ and variance $1/\lambda^2$, *i.e.*,

$$E(T) = 1/\lambda, \ \text{Var. } (T) = 1/\lambda^2.$$

In a similar fashion, the converse of this theorem can be proved.

Q. 1. Give the axioms characterizing a Poisson process. If the number of arrivals in some time interval follows a Poisson distribution, show that the distribution of the time interval between two consecutive arrivals is exponential.
[Delhi M.A/M.Sc. (Stat.) 95; Raj. Univ. (M. Phil) 91]

2. Show that if the inter-arrival times are negative exponentially distributed, the number of arrivals in a time period is a Poisson process and conversely.

3. If the intervals between successive arrivals are i.i.d. random variables which follow the negative exponential distribution with mean $1/\lambda$, then show that the arrivals form a Poisson Process with mean λt. **[Garhwal M.Sc. (Stat.) 91]**

4. Show that inter-arrival times are distributed exponentially, if arrival is a Poisson process. Prove the converse also.
[Delhi M.A/M.Sc (OR) 92.]

5. State the three axioms underlying the exponential process. Under exponential assumptions can two events occur during a very small interval. **[Meerut 2002]**

23.7-4. Markovian Property of Inter-arrival Times

Statement. *The Markovian property of inter-arrival times states that at any instant the time until the next arrival occurs is independent of the time that has elapsed since the occurrence of the last arrival. That is to say,*

$$Prob.\ [T \geq t_1 \mid T \geq t_0] = Prob.\ [0 \leq T \leq t_1 - t_0]$$

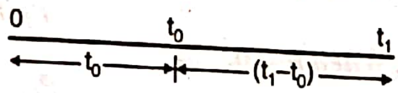


Fig. 23.7

Proof. Consider

$$\text{Prob. } [T \geq t_1 \mid T \geq t_0] = \frac{\text{Prob. } [(T \geq t_1) \text{ and } (T \geq t_0)]}{\text{Prob. } [T \geq t_0]} \quad \text{(formula of conditional probability)} \quad ...(23.29)$$

Since the inter-arrival times are exponentially distributed, the right hand side of equation (23.29) can be written as

$$\frac{\int_{t_0}^{t_1} \lambda e^{-\lambda t}\, dt}{\int_{t_0}^{\infty} \lambda e^{-\lambda t}\, dt} = \frac{e^{-\lambda t_1} - e^{-\lambda t_0}}{-e^{-\lambda t_0}}$$

$$\therefore \quad \text{Prob. } [T \geq t_1 \mid T \geq t_0] = 1 - e^{-\lambda(t_1 - t_0)} \quad ...(23.30)$$

---

* According to probability distributions $d/dx\,[F(x)] = f(x)$, where $F(x)$ is the *'distribution function'* and $f(x)$ is the *'probability density function'*. Hence by the similar argument, we may write $d/dT\,[P_0(T)] = a(T)$, where $P_0(T)$ is the probability distribution function for no arrival in time T, and $a(T)$ is denoting the corresponding probability density function of T.

** Since *'probability density function'* is always non-negative, so neglect the negative sign from right side of equation (23.26).

But, $\quad \text{Prob. } [0 \le T \le t_1 - t_0] = \int_0^{t_1 - t_0} \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda(t_1 - t_0)}$. ...(23.31)

Thus, by virtue of equations (23.30) and (23.31), it can be concluded that

$$\text{Prob. } [T \ge t_1 \mid T \ge t_0] = \text{Prob. } [0 \le T \le t_1 - t_0].$$

This proves the *Markovian property* of inter-arrival times.

Q. State and prove the Markovian property of inter-arrival times (*i.e.* of exponential distribution).

23.7-5. Distribution of Departures (or Pure Death Process)

In this process assume that there are N customers in the system at time $t = 0$. Also, assume that no arrivals (births) can occur in the system. Departures occur at a rate μ per unit time, *i.e.*, output rate is μ. We wish to derive the distribution of departures from the system on the basis of the following three *axioms* :

(1) Prob. [*one departure during* Δt] $= \mu \Delta t + O(\Delta t)^2 = \mu \Delta t \quad [\because O(\Delta t)^2$ is negligible]

(2) Prob. [*more than one departure during* Δt] $= O(\Delta t)^2 \doteq 0$.

(3) The number of departures in non-overlapping intervals are statistically independent and identically distributed random variable, *i.e.*, the process $N(t)$ has independent increments.

First obtain the differential difference equation in three mutually exclusive ways :

Case I. ***When*** $0 < n < N$. Proceeding exactly as in the *Pure Birth Process*,

$$P_n(t + \Delta t) = P_n(t)\,[1 - \mu \Delta t] + P_{n+1}(t)\, \mu \Delta t \quad \text{...(23.32)}$$

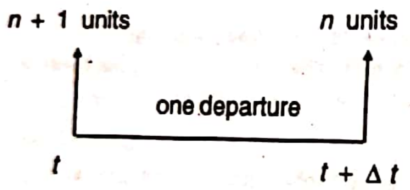


Fig. 23.8

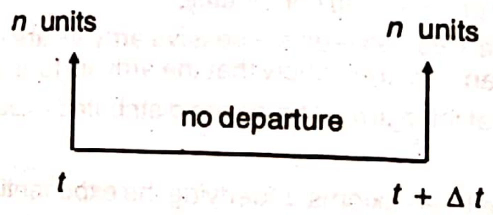


Fig. 23.9

Case II. ***When*** $n = N$. Since there are exactly N units in the system, $P_{n+1}(t) = 0$,

$\therefore$
$$P_N(t + \Delta t) = P_N(t)\,[1 - \mu \Delta t] \quad \text{...(23.33)}$$

Case III. ***When*** $n = 0$.

$$P_0(t + \Delta t) = P_0(t) + P_1(t)\, \mu \Delta t \quad \text{...(23.34)}$$

Since there is no unit in the system at time t, the question of any departure during Δt does not arise. Therefore, probability of no departure is unity in this case.

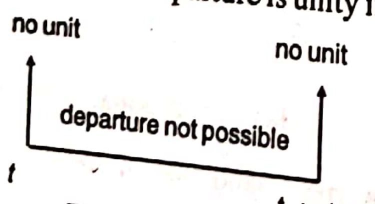


Fig. 23.10.

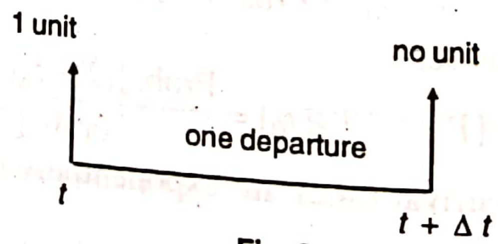


Fig. 23.11.

Now, re-arranging the terms and dividing by Δt, and also taking the limit $\Delta t \to 0$ the equations (23.33), (23.32) and (23.34), respectively, become

$$P_N'(t) = -\mu P_N(t), \qquad n = N \quad \text{...(23.35)}$$

$$P_n'(t) = -\mu P_n(t) + \mu P_{n+1}(t), \qquad 0 < n < N \quad \text{...(23.36)}$$

$$P_0'(t) = \mu P_1(t), \qquad n = 0 \quad \text{...(23.37)}$$

To solve the system of equations (23.35), (23.36) and (23.37) :

Iterative method can be used to solve the system of three equations.

Step 1. From equation (23.35) obtain

$$\frac{P_N'(t)}{P_N(t)} = -\mu \quad \text{or} \quad \frac{d}{dt} \log P_N(t) = -\mu.$$

Integrating both sides of this equation,

$$\log P_N(t) = -\mu t + A \qquad \text{...(23.38)}$$

To determine 'A', use the boundary condition $P_N(0) = 1$, and thus get $A = 0$ ($\because \log 1 = 0$).

Therefore, equation (23.38) becomes

$$\log P_N(t) = -\mu t \quad \text{or} \quad P_N(t) = e^{-\mu t} \qquad \text{...(23.39)}$$

Step 2. In equation (23.36), put $n = N - 1$, and the value of $P_N(t)$ from equation (23.39),

$$P'_{N-1}(t) = -\mu P_{N-1}(t) + \mu P_N(t)$$

or
$$P'_{N-1}(t) = -\mu P_{N-1}(t) + \mu e^{-\mu t} \qquad \text{[from equation (23.39)]}$$

or
$$P'_{N-1}(t) + \mu P_{N-1}(t) = \mu e^{-\mu t} \qquad \text{...(23.40)}$$

The solution of this equation is given by

$$P_{N-1}(t)\, e^{\mu t} = \int \mu e^{-\mu t} e^{\mu t}\, dt + B \qquad (\because \text{I.F.} = e^{\mu t})$$

or
$$P_{N-1}(t) = \mu t\, e^{-\mu t} + Be^{-\mu t} \qquad \text{...(23.41)}$$

To determine B, put $t = 0$, $P_{N-1}(t) = 0$ in (23.41) and get $B = 0$. Therefore,

$$P_{N-1}(t) = \frac{\mu t\, e^{-\mu t}}{1\,!}$$

Step 3. Putting $n = N - 2$ in equation (23.36) and proceeding exactly as in *Step 2*,

$$P_{N-2}(t) = \frac{e^{-\mu t}(\mu t)^2}{2\,!}$$

Step 4. Now, putting $n = N - 3, N - 4, ..., N - i$, and using induction process

$$P_{N-3}(t) = \frac{e^{-\mu t}(\mu t)^3}{3\,!}$$

$$\cdots \qquad \cdots$$

$$P_{N-i}(t) = \frac{e^{-\mu t}(\mu t)^i}{i\,!}, \quad i = 0, 1, 2, ..., N-1$$

$$\cdots \qquad \cdots$$

In general, on letting $n = N - i$

$$P_n(t) = \frac{e^{-\mu t}(\mu t)^{N-n}}{(N-n)\,!}, \quad n = 1, 2, ..., N \qquad \text{...(23.42)}$$

Step 5. In order to find $P_0(t)$, use the following procedure,

Since
$$1 = \sum_{n=0}^{N} P_n(t) = P_0(t) + \sum_{n=1}^{N} P_n(t)$$

$\therefore$
$$P_0(t) = 1 - \sum_{n=1}^{N} P_n(t) = 1 - \sum_{n=1}^{N} \frac{(\mu t)^{N-n} e^{-\mu t}}{(N-n)\,!} \qquad \text{...(23.43)}$$

Finally, combining the results (23.42) and (23.43)

$$P_n(t) = \begin{cases} \dfrac{(\mu t)^{N-n} e^{-\mu t}}{(N-n)\,!}, & \text{for } n = 1, 2, ..., N \\[2ex] 1 - \displaystyle\sum_{n=1}^{N} \frac{(\mu t)^{N-n} e^{-\mu t}}{(N-n)\,!}, & \text{for } n = 0 \end{cases} \qquad \text{...(23.44)}$$

Thus, the number of departures in time t follows the *'Truncated Poisson Distribution'*.

Q. Establish the probability distribution formula for Pure-Death Process.

23.7-6. Derivation of Service Time Distribution

Let T be the random variable denoting the service time and t the possible value of T.

Let $S(t)$ and $s(t)$ be the *cumulative density function* and the *probability density function* of T, respectively.

To find $s(t)$ for the Poisson departure case, it has been observed that the probability of no service during time 0 to t is equivalent to the probability of having no departure during the same period.

Thus, Prob. [*service time* $T \geq t$] = Prob. [*no departure during* t] = $P_N(t)$

where there are N units in the system and no arrival is allowed after N. Therefore, $P_N(t) = e^{-\mu t}$

$$S(t) = \text{Prob.}\ (T \leq t) = 1 - \text{Prob.}\ [T \geq t] \quad \text{or} \quad S(t) = 1 - e^{-\mu t}$$

Differentiating both sides, w.r.t. 't', we get

$$\frac{d}{dt} S(t) = s(t) = \begin{cases} \mu e^{-\mu t}, & t \geq 0 \\ 0, & t < 0 \end{cases}$$

Thus, it is concluded that the service time distribution is *'Exponential'* with mean $1/\mu$ and variance $1/\mu^2$.

Thus, mean service time $= 1/\mu$.

23.7-7. Analogy of Exponential Service Times with Poisson Arrivals

It has been proved in sec. **23.7-3** that if number of arrivals (n) follows the *Poisson distribution*, then the inter-arrival time (T) will follow the *exponential one*, and *vice-versa.*

In the like manner, it can also be shown that, if the time (t) to complete the service of a unit follows the *exponential distribution* given by the probability density function

$$s(t) = \mu e^{-\mu t} \qquad ...(23.45)$$

where μ is the mean servicing rate for a particular station, then the number (n) of departures in time T (if there were no enforced idle time) will follow the *Poisson distribution* given by

$$\varphi_T(n) = \text{Prob. [}n\text{ services in time }T\text{, if servicing is going on throughout }T] = (\mu T)^n\, e^{-\mu T}/n\,! \qquad ...(23.46)$$

Consequently, from (23.27), it can be shown that

$$\varphi_{\Delta T}(0) = \text{Prob. [}\textit{no service in } \Delta T] = 1 - \mu \Delta T \qquad ...(23.47)$$

and

$$\varphi_{\Delta T}(1) = \text{Prob. [}\textit{one service in } \Delta T] = \mu \Delta T. \qquad ...(23.48)$$

23.7-8. Erlang Service Time Distribution (E_k).

So far it is considered (in **23.7-3** and **23.7-6**) and seen that the inter-arrival time distribution and service time distribution both will follow the exponential assumptions given by

$$a(T) = \lambda e^{-\lambda t}, \text{ and } s(t) = \mu e^{-\mu t}, \text{ respectively.} \qquad ...(23.49)$$

These only give a one particular family of possible arrival and service time distribution, respectively.

A two parameter (μ and k) generalisation of the exponential family, which is of great importance in queueing problems is called the *Erlang* family of service time distribution (named for *A.K. Erlang*, the ***Danish*** telephone engineer. This is defined by its probability density function,

$$s(t, \mu, k) = (k\mu)^k\, t^{k-1}\, e^{-k\mu t}/(k-1)\,! = C_k\, t^{k-1}\, e^{-k\mu t} \qquad ...(23.50)$$

where $C_k = (k\mu)^k/(k-1)\,!$, $0 \leq t < \infty$, $k \geq 1$.

It should be noted carefully that (23.50) gives us the exponential distribution given by (23.49) for $k = 1$.

Let $t_1, t_2, t_3, ..., t_k$ be the servicing time for any customer in respective k phases, then the total service time t is given by

$$t = t_1 + t_2 + \ldots + t_k.$$

Also, each of the times $t_1, t_2, ..., t_k$ is independently and exponentially distributed with parameter $k.\mu$.

Hence, $P\,[t \leq t_1 + t_2 \ldots + t_k \leq t + \Delta t]$

$$= \iint \ldots \int p(t_1)\, p(t_2) \ldots p(t_k)\, dt_1\, dt_2 \ldots dt_k. \text{ for } t \leq t_i \leq t + \Delta t,\ i = 1, 2, ..., k.$$

$$= \iint \ldots \int (k\,\mu e^{-k\mu t_1} \ldots (k\,\mu e^{-k\mu t_k})\, dt_1 \ldots dt_k. \quad [\text{since } p(t_i) = k\,\mu e^{-k\mu t_i}]$$

$$= (k\,\mu)^k \iint \ldots \int e^{-k\mu \sum_{i=1}^{k} t_i}\, dt_1 \ldots dt_k.$$

Now applying ***Drichlet's*** theorem of multiple integrals,

$$= (k\,\mu)^k \frac{\Gamma 1^k}{\Gamma(k)}\, e^{-k\mu t}\, t^{k-1}$$

$$= \frac{(k\mu)^k}{\Gamma(k)} t^{k-1} e^{-k\mu t}, k \geq 0.$$

Note. The distribution is a modified χ^2 distribution with mean $(1/\mu)$ and $2k$ degrees of freedom.

Thus, if we have service times $t_1, t_2, t_3, ..., t_k$ in k phases which are exponentially distributed variables with a common mean $1/k\mu$, then $t = t_1 + t_2 + t_3 + ... + t_k$ has the *Erlangian (Gamma)* distribution with k phases and parameter μ.

Derivation of Erlangion Service Distribution

In the *Fig.* 23.12 for *Erlangian service time distribution* for $k = 3$ phases, it is observed that—

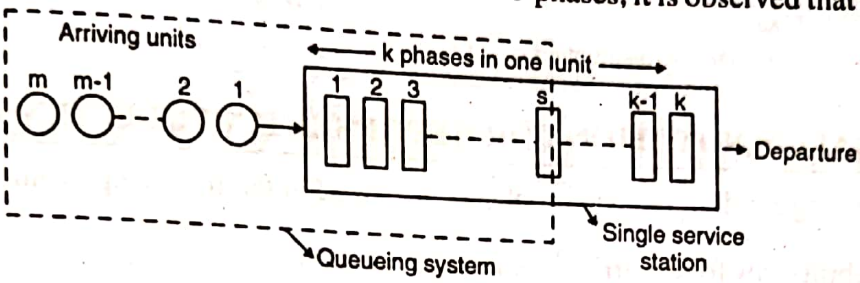


Fig. 23.12

(*i*) each phase of service is exponential;

(*ii*) a unit enters phase 3 first, then goes to 2, to 1 and out;

(*iii*) no other unit can enter phase 3 until the previous unit leaves phase 1.

Properties. The *Erlang* family of service time distribution has many interesting properties such as :

(1) All the members share the common mean $1/\mu$, that is, $E(t) = 1/K\mu$, and variance of t is given by $V(t) = 1/k\mu^2$

(2) One parameter family is obtained by setting $k = 1$.

(3) The mode is located at :

$t = 0$	for $k = 1$,
$t = 1/2\mu$	for $k = 2$,
- - - - - - - -	
$t = (k-1)/k\mu$	for general k,
$t = 1/\mu$	for $k \to \infty$

(4) As $k \to \infty$, $V(t) \to 0$, [since $V(t) = 1/k\mu^2$]

(5) For constant service time, $k \to \infty$ [Note.]

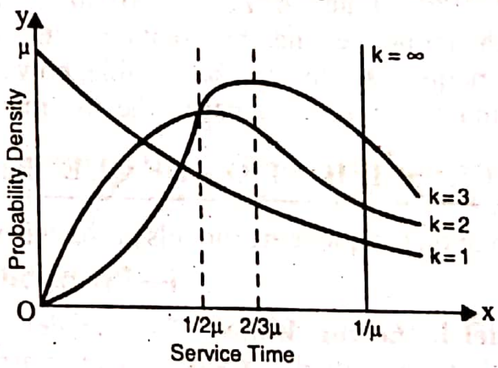


Fig. 23.13. The Earlang family of service time distribution.

23.8. SOME QUALITATIVE ASSUMPTIONS

It is essential to pay much attention to physical significance of following three qualitative assumptions for further discussion of queueing system.

(a) Stationary waiting line. The probability that n customers arrive in a time interval $(T, T + t)$ is independent of T and is a function of the variables n and t both.

(b) Absence of after effects. This means that the probability of n customers arriving during a time interval $(T, T + t)$ does not depend on the number of customers arriving before T.

(c) The orderlines of the waiting line. It expresses the practical impossibility of two or more customers arriving at the same instant of time.

A waiting line (queue) satisfying the above three conditions is usually called a *Simple Queue*.

Q. 1. (*a*) Explain the basic queueing process. What are the important random variates in queueing system to be investigated ?

(*b*) What do you understand by (*i*) queue discipline (*ii*) input, and (*iii*) holding time ?

2. Explain : (*i*) the constituents of a queueing model, and (*ii*) the characteristics to be analysed.

3. State some of the important distributions of arrival intervals and service times. [Delhi MA/M.Sc (OR) 93]
4. Give the essential characteristics of the queueing process.
5. State some of the important inter-arrival and service time distributions.
6. What do you understand by a queue ? Give some important applications of queueing theory.
7. Find the mode of the k-Erlang distribution with parameter μ.
8. What do you understand by a optimum service rate ? Show how some important queueing formulae are used in determining the optimum service rate and the number of channels.
9. Prove that, for the Erlang distribution with parameters μ and k, the mode is at $(1 - 1/k)\, 1/\mu$, the mean is $1/\mu$, and the variance is $1/k\mu^2$. [Kanpur 96; Garhwal M.Sc. (Stat) 92; Meerut (M.Sc. Maths) 90]
10. Write a note on *Erlangian* distribution. [Meerut (OR) 2003]

23.9. KENDALL'S NOTATION FOR REPRESENTING QUEUEING MODELS

Generally queueing model may be completely specified in the following symbolic form : $(a \mid b \mid c) : (d \mid e)$, where

$a \equiv$ probability law for the arrival (or inter-arrival) time.
$b \equiv$ probability law according to which the customers are being served,
$c \equiv$ number of channels (or service stations),
$d \equiv$ capacity of the system, *i.e.*, the maximum number allowed in the system (in service and waiting),
$e \equiv$ queue discipline.

It is important to note that first three characteristics $(a \mid b \mid c)$ in the above notation were introduced by *D. Kendall* (1953). Later, *A. Lee* (1966) added the *fourth* (d) and the *fifth* (e) characteristics to the notation. Although, it is noticed that this notation is not suitable for describing complex models such as queues in series, or network queues. This will be suitable, however, for the purpose of material presented here, and the reader should find it helpful in comparing the different models.

23.10. CLASSIFICATION OF QUEUEING MODELS

For simplicity, the queueing models presented here are classified as follows :

I—Probabilistic Queueing Models

Model I. (*Erlang Model*). This model is symbolically represented by $(M \mid M \mid 1) : (\infty \mid FCFS)$. This denotes *Poisson* arrival (exponential inter-arrival), *Poisson* departure (exponential service time), single server, infinite capacity and *"First Come, First Served"* service discipline.

Note. Since the *'Poisson'* and the *'Exponential'* distributions are related to each other (see **sec. 23.7-3**), both of them are denoted by the same letter 'M'. Letter 'M' is used due to *Markovian property* of exponential process.

Model II. (*General Erlang Model*). Although this model is also represented by $(M \mid M \mid 1) : (\infty \mid FCFS)$, but this is a general queueing model in which the rate of arrival and the service depend on the length n of the line.

Model III. This model is represented by $(M \mid M \mid 1) : (N \mid FCFS)$. In this model, capacity of the system is limited (finite), say N. Obviously, the number of arrivals will not exceed the number N in any case.

Model IV. This model is represented by $(M \mid M \mid s) : (\infty \mid FCFS)$, in which the number of service stations is s in parallel.

Model V. This model is represented by $(M \mid E_k \mid 1) : (\infty \mid FCFS)$, that is, *Poisson* arrivals, *Erlangian* service time for k phases (see **sec. 23.7-8**), and a single server.

Model VI. (*Machine Repairing Model*). This model is represented by $(M \mid M \mid R) : (K \mid GD)$, $K > R$, that is, *Poisson* arrivals, *Exponential* service time, R repairmen, and K machines in the system, and general queue discipline.

Model VII. *Power-Supply Model.*

Model VIII. *Economic Cost Profit Models.*

Model IX. $(M \mid G \mid 1) : (\infty \mid GD)$, where G is the general output distribution, and GD represents a general service discipline.

II—Mixed Queueing Model

Model X. $(M \mid D \mid 1) : (\infty \mid FCFS)$, where D stands for deterministic service time.

III—Deterministic Queueing Model

Model XI. $(D \mid D \mid 1) : (K - 1 \mid FCFS)$, where
$D \rightarrow$ Deterministic arrivals, *i.e.*, inter-arrival time distribution is *constant* or *regular*.
$D \rightarrow$ Deterministic service time distribution.

Q. 1. Give a brief summery of the various types of queueing models.
2. Write a note on *Kendal* and *Lee's* notation for the identification of queues. **[Karnataka B.E. (C.S.E.) 94]**

23.11. SOLUTION OF QUEUEING MODELS AND LIMITATIONS FOR ITS APPLICATIONS

The solution of queueing models as classified in sec. **23.10** will consist of the following parts :

(*a*) ***To obtain the system of steady state equations governing the queue.***
(*b*) ***To solve these equations for finding out the probability distribution of queue length.***
(*c*) ***To obtain probability density function for waiting time distribution.***
(*d*) ***To find the busy period distribution.***
(*e*) ***To derive formula for L_s, L_q, $(L \mid L > 0)$, W_s, W_q, $(W \mid W > 0)$, and Var $\{n\}$, etc.***
(*f*) ***Also, to obtain the probability of arrival during the service time of any customer.***

The analytic procedure may be adopted for solving the steady state equations for ***Models*** I-IV. Since the analytic procedure seems to be more complicated for ***Model*** V, so we shall adopt the increasingly powerful technique of *Generating Functions.*

Limitation for Application of Queueing Model :

The single channel queueing model can be fitted in situations where the following conditions are satisfied.

(i) The number of arrivals rate is denoted by λ.
(ii) The service time has exponential distribution. The average service rate is denoted by μ.
(iii) Arrivals are from infinite population.
(iv) The queue discipline is *FCFS (i.e. FIFO), i.e.* the customers are served on a first come first served basis—
(v) There is only a single service station.
(vi) The mean arrival rate is less than the mean service rate, *i.e.* $\lambda < \mu$.
(vii)The waiting space available for customers in the queue is infinite.

The single channel queueing model is the most simple model which is based on the above mentioned assumption. But, in reality, there are several limitations of this modes in its applications. One abvious limitation is the possibility that the waiting space, in fact, be limited. Other possibility is that arrival rate is state dependent. That is, potential customers are discouraged from entering the queue if they observe a long line at the time they arrive. Another practical limitation of the model is that the arrival process is not stationary. It is quite possible that the service station would experience peak period, and slack periods during which the arrival rate is higher and lower respectively than the over all average. These could occur at particular times during a day or a week or particular weeks during a year. The population of customers served may be finite, the queue discipline may not be *first come first served.* In general, the validity of these models depends on the assumptions that are often unrealistic in practice.

Even when the assumptions are realistic, there is another limitation of queueing theory that is often overlooked. Queueing models give steady state solution, *i.e.* the model tells us what will happen after queueing system hes been in operation long enough to eliminate the effects of starting with an empty queue at the beginning of each business day. In some applications, the queueing system never reaches a steady state, so the model solutions are of little importance.

Q. 1. Mention any seven conditions that must be fulfilled by the situations if they were to be described by a queueing model. What are the limitations of this model in its applications. **[C.A. (Nov.) 93]**
2. (*a*) Describe Queueing model and its significance. What are various queue models, give in details ?
(*b*) List the factors that constitute the basic elements of queuing model. For each of those, enumerate the alternatives possible. Represent them diagramatically to cover all possible implimentations of a queueing model ? **[IGNOU 99, 98, 96]**

We now proceed to discuss each model in detail with the help of various interesting examples.

23.12. MODEL I. $(M \mid M \mid 1) : (\infty \mid FCFS)$: BIRTH AND DEATH MODEL

This model is also called the *'birth and death model'*.

I. To obtain the system of steady-state equations. [Agra 99; Meerut 93]

The probability that there will be n units $(n > 0)$ in the system at time $(t + \Delta t)$ may be expressed as the sum of three independent compound probabilities, by using the fundamental properties of probability, Poisson arrivals, and of exponential service times.

(*i*) ***The product of three probabilities*** (see **Fig. 23.14**),

(*a*) that there are n units in the system at time $t = P_n(t)$

(*b*) that there is no arrival in time $\Delta t = P_0(\Delta t) = 1 - \lambda \Delta t$ [see (23.20)]

(*c*) that there is no service in time $\Delta t = \varphi_{\Delta t}(0) = 1 - \mu \Delta t$; [see (23.47)]

is given by

$$P_n(t).(1 - \lambda \Delta t).(1 - \mu \Delta t) \cong P_n(t)\,[1 - (\lambda + \mu)\,\Delta t] + O_1(\Delta t).$$

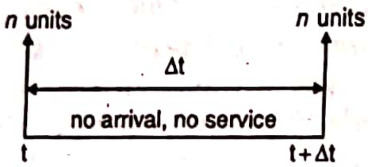


Fig. 23.14.

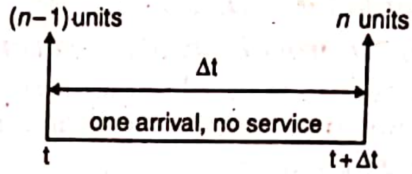


Fig. 23.15

(*ii*) ***The product of three probabilities*** (see **Fig. 23.15**),

(*a*) that there are $(n - 1)$ units in the system at time $t = P_{n-1}(t)$;

(*b*) that there is one arrival in time $\Delta t = P_1(\Delta t) = \lambda \Delta t$ [see (23.21)]

(*c*) that there is no service in $\Delta t = \varphi_{\Delta t}(0) = 1 - \mu \Delta t$;

is given by $\quad P_{n-1}(t).(\lambda \Delta t).(1 - \mu \Delta t) \cong \lambda P_{n-1}(t)\,\Delta t + O_2(\Delta t)$,

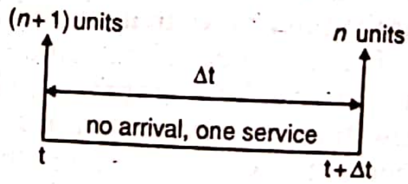
(*iii*) ***The product of probabilities*** (see **Fig. 23.16**),

(*a*) that there are $(n + 1)$ units in the system at time $t = P_{n+1}(t)$

(*b*) that there is no arrival in time Δt
$= P_0(\Delta t) = 1 - \lambda \Delta t$

(*c*) that there is one service in time
$\Delta t = \varphi_{\Delta t}(1) = \mu \Delta t$; [see (23.48)]

is given by



$$P_{n+1}(t)\,(1 - \lambda \Delta t)\,\mu \Delta t \cong P_{n+1}(t)\,\mu \Delta t + O_3(\Delta t).$$

Fig. 23.16

Note. The probabilities of more than one unit arriving and/or being served during the interval Δt are assumed to be negligible. Further, $O_1(\Delta t)$, $O_2(\Delta t)$, $O_3(\Delta t)$ are also the functions of Δt in the sense of notation '$O(\Delta t)$' as explained in sec. 23.7-2.

Now, by adding above three independent compound probabilities, we obtain the probability of n units in the system at time $(t + \Delta t)$, *i.e.*,

$$P_n(t + \Delta t) = P_n(t)\,[1 - (\lambda + \mu)\,\Delta t] + P_{n-1}(t)\,\lambda \Delta t + P_{n+1}(t)\,\mu \Delta t + O(\Delta t), \quad ...(23.51)$$

where $O(\Delta t) = O_1(\Delta t) + O_2(\Delta t) + O_3(\Delta t)$.

The equation (23.51) may be written as

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = -(\lambda + \mu)\,P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t) + \frac{O(\Delta t)}{\Delta t}$$

Now, taking limit as $\Delta t \to 0$ on both sides,

$$\lim_{\Delta t \to 0} \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \lim_{\Delta t \to 0} \left[-(\lambda+\mu) P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t) + \frac{O(\Delta t)}{\Delta t} \right].$$

or

$$\frac{dP_n(t)}{dt} = -(\lambda+\mu) P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t);\ n > 0 \left(\text{since } \lim_{\Delta t \to 0} \frac{O(\Delta t)}{\Delta t} = 0 \right) \quad ...(23.52)$$

In a similar fashion, the probability that there will be no unit (*i.e.* $n = 0$) in the system at time $(t + \Delta t)$ will be the sum of the following two independent probabilities :

(*i*) Prob. [*that there is no unit in the system at time t and no arrival in time* Δt] $= P_0(t).(1 - \lambda\Delta t)$. question of any service in time Δt does not arise because there are *no* units in the system at time *t;* and

(*ii*) Prob. [*that there is one unit in the system at time t, one unit serviced in* Δt, *and no arrival in* Δt]

$$= P_1(t).\mu\Delta t.(1 - \lambda\Delta t) \cong P_1(t)\mu\Delta t + O(\Delta t)$$

Now, adding these two probabilities, we get

$$P_0(t+\Delta t) = P_0(t)\,[1 - \lambda\Delta t] + P_1(t)\,\mu\Delta t + O(\Delta t) \quad ...(23.53)$$

or

$$\frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\lambda P_0(t) + \mu P_1(t) + \frac{O(\Delta t)}{\Delta t}.$$

Now, taking limit on both sides as $\Delta t \to 0$

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t), \text{ for } n = 0 \quad ...(23.54)$$

Since only the steady state probabilities are considered here (see **sec. 23.4**),

$$\lim_{t \to \infty} \frac{d\,[P_n(t)]}{dt} = 0, \text{ for } n \geq 0 \text{ and } \lim_{t \to \infty} P_n(t) = P_n \text{ (which is independent of } t)$$

Consequently, the equations (23.52) and (23.54) can be written as :

$$0 = -(\lambda+\mu) P_n + \lambda P_{n-1} + \mu P_{n+1}, \quad \text{if } n > 0 \quad ...(23.52a)$$

$$0 = -\lambda P_0 + \mu P_1, \quad \text{if } n = 0 \quad ...(23.54a)$$

In this way, the equations (23.52a) and (23.54a) constitute the *system of steady state difference equations* for this model.

Q. In a single server, Poisson arrival and exponential service time queueing system, show that probability P_n of n customers in steady state satisfy the following equations :

$$\lambda P_0 = \mu P_1,\ (\lambda+\mu)P_1 = \mu P_2 + \lambda P_0, \text{ and } (\lambda+\mu)P_n = \mu P_{n+1} + \lambda P_{n-1}, \text{ for } n \geq 2.$$

[Meerut 97 P]

II. To solve the system of difference equations.

By the technique of successive substitution. we solve the difference equations :

$$0 = -(\lambda+\mu) P_n + \lambda P_{n-1} + \mu P_{n+1}, \quad \text{if } n > 0,$$

$$0 = -\lambda P_0 + \mu P_1, \quad \text{if } n = 0,$$

Since $P_0 = P_0$

$$P_1 = \frac{\lambda}{\mu} P_0 \quad \text{[from the equation (23.54a) for } n = 0]$$

$$P_2 = \frac{\lambda}{\mu} P_1 = \left(\frac{\lambda}{\mu}\right)^2 P_0 \quad \text{[from letting } n = 1 \text{ in the eqn. (23.52a) for } n > 0 \text{ and substituting for } P_1]$$

$$P_3 = \frac{\lambda}{\mu} P_2 = \left(\frac{\lambda}{\mu}\right)^3 P_0 \quad \text{[from letting } n = 2 \text{ in eqn. (23.52a) for } n > 0 \text{ and substituting for } P_2]$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$P_n = \frac{\lambda}{\mu} P_{n-1} = \left(\frac{\lambda}{\mu}\right)^n P_0 \quad [\text{for } n \geq 0] \quad ...(23.55)$$

Now using the fact that $\sum_{n=0}^{\infty} P_n = 1$,

$$P_0 + \frac{\lambda}{\mu} P_0 + \left(\frac{\lambda}{\mu}\right)^2 P_0 + \ldots + \left(\frac{\lambda}{\mu}\right)^n P_0 + \ldots = 1 \quad \text{or} \quad P_0 \left[1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu}\right)^2 \cdots \right] = 1$$

or $\quad P_0\left[\dfrac{1}{1-\lambda/\mu}\right]=1 \quad$ [since $(\lambda/\mu)<1$ as explained in sec. **23.6**, sum of infinite G.P. is valid]

or $\quad P_0 = 1-(\lambda/\mu). \quad$ ...(23.56)

Now, substituting the value of P_0 from (23.56) in (23.55), we get

$$P_n = \left(\frac{\lambda}{\mu}\right)^n\left(1-\frac{\lambda}{\mu}\right)=\rho^n(1-\rho),\ \left(\rho=\frac{\lambda}{\mu}<1, n\ge 0\right) \quad \text{...(23.57)}$$

Thus, equations (23.56) and (23.57) rather give the required *probability distribution of queue length.*

Q. 1. Derive the differential-difference equations for the queueing model $(M \mid M \mid 1):(\infty \mid FCFS)$. How would you proceed to solve the model? **[Meerut (Stat.) 98; Delhi M.A/M.Sc. (OR). 90]**

2. Obtain the steady state solution of $(M \mid M \mid 1):(\infty \mid FCFS)$ system and also find expected value of queue length n. **[Meerut (Maths.) 97P; Garhwal (Stat.) 96]**

3. Explain $(M \mid M \mid 1):(\infty \mid FCFS)$ queueing model, derive and solve the difference equations in steady state, of the model. **[Agra 93; Garhwal M.Sc. (Stat.) 93]**

4. Show that for a single service station, Poisson arrivals and exponential service time, the probability that exactly n calling units are in the queueing system is $P_n=(1-\rho)\,\rho^n,\ n\ge 0$, where ρ is the traffic intensity.

Further, we may also compute a useful probability, *viz.,*

$$\text{Prob. }[\textit{queue size} \ge N] = \sum_{n=N}^{\infty} P_n = \sum_{n=0}^{\infty} P_n - \sum_{n=0}^{N-1} P_n = 1-(P_0+P_1+\ldots+P_{N-1})$$

$$=1-\left[P_0+\frac{\lambda}{\mu}P_0+\ldots+\left(\frac{\lambda}{\mu}\right)^{N-1}P_0\right]=1-P_0\left[\frac{1-(\lambda/\mu)^N}{1-\lambda/\mu}\right]$$

$$=1-\left(1-\frac{\lambda}{\mu}\right)\left[\frac{1-(\lambda/\mu)^N}{1-\lambda/\mu}\right]=\left(\frac{\lambda}{\mu}\right)^N \quad [\text{from (5.56)}, P_0 = 1-(\lambda/\mu)]$$

$$\therefore \quad \textbf{Prob. }[\textit{queue size} \ge N]=(\lambda/\mu)^N=\rho^N. \quad \text{...(23.58)}$$

Q. 1. Describe a queue model and steady state equations of $M \mid M \mid 1$ queues. What is the prob. that at least one unit is present in the system. **[Meerut (I.P.M.) 90]**

2. Explain M I M I 1 queue model in the transient state. Derive steady state solution for the M I M I 1 queue model. **[Garhwal M.Sc. (Math.) 91]**

3. If P_n represents the probability of finding n in the long run in a queueing system with Poisson arrivals having parameter λ and exponential service times with parameter μ, show that.

$$\lambda P_{n-1}-(\lambda+\mu)P_n+\mu P_{n+1}=0 \quad \text{for } n>0$$

and

$$-\lambda P_0+\mu P_1=0. \quad \text{for } n=0$$

Solve these difference equations and obtain P_n in terms of $P=\lambda/\mu$. **[I.A.S. (Maths.) 95]**

III. To obtain probability density function of waiting time (excluding service time) distribution. [Kanpur 93]

In the steady state, each customer has the same waiting time distribution. This is a continuous distribution with probability density function $\Psi(w)$, and we denote by $\Psi(w)\,dw$ the probability that a customer begins to be served in the interval $(w, w+dw)$, where w is measured from the time of his arrival. We suppose that a customer arrives at time $w=0$ and service begins in the interval $(w, w+dw)$. *Fig.* 23.17 illustrates this situation. For convenience, we label the customer as A.

(*i*) There is a finite probability that waiting time is zero (P_0 the probability that the system is empty).

(*ii*) If there are n customers already in the system when the customer A *arrives,* n must leave before the service of A

Fig. 23.17

begins. More precisely, $(n-1)$ customers must leave during the time interval $(0, w)$, and the nth customer during $(w, w+dw)$.

[If n customers left by the time w, service of A could begin before the interval $(w, w+dw)$; and if fewer than $(n-1)$ left by time w, service could only begin in $(w, w+dw)$; if there were two or more departures in that interval, the probability is $O(dt)$ which may be ignored].

The server's mean rate of service is μ in unit time, or μw in time w, and the probability of $(n-1)$ departures in time w, during which the sever is busy, is the appropriate term of the Poisson distribution $(\mu w)^{n-1} e^{-\mu w}/(n-1)\,!$.

Let there be n units in the system (see *Fig.* 23.18), then

$\Psi_n(w)\, dw =$ Prob. [$(n-1)$ *units are served at time* w] × Prob. [*one unit is served in time* dw],

or
$$\Psi_n(w)\, dw = \frac{(\mu w)^{n-1} e^{-\mu w}}{(n-1)\,!} \times \mu\, dw. \qquad \text{...(23.59)}$$

Let W be the waiting time of a unit who has to wait such that $w \le W \le w\, dw$, then the probability $\Psi(w)\, dw$ is given by

$\Psi(w)\, dw =$ Prob. $(w \le W \le w + dw)$

= (The probability of n customers in the system when customer A arrives) × [the probability that exactly $n-1$ customers leave in $(0, w)$] × [the probability that nth customer leaves in $(w, w+dw)$], summed over all n from 1 to ∞

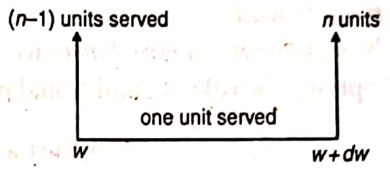


Fig. 23.18

$$= \sum_{n=1}^{\infty} P_n \Psi_n(w)\, dw = \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) \cdot \frac{(\mu w)^{n-1} e^{-\mu w}}{(n-1)\,!} \mu\, dw$$

[from (23.57) and (23.59)]

$$P_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu}\left(1 - \frac{\lambda}{\mu}\right) \mu\, e^{-\mu w} \sum_{n=1}^{\infty} \frac{[(\lambda/\mu)\,(\mu w)]^{n-1}}{(n-1)\,!} dw = \lambda\left(1 - \frac{\lambda}{\mu}\right) e^{-\mu w} \sum_{n=1}^{\infty} \frac{(\lambda w)^{n-1}}{(n-1)\,!} dw$$

$$= \lambda\left(1 - \frac{\lambda}{\mu}\right) e^{-\mu w}\left[1 + \frac{(\lambda w)}{1\,!} + \frac{(\lambda w)^2}{2\,!} + \ldots\right] dw \qquad = \lambda\left(1 - \frac{\lambda}{\mu}\right) e^{-\mu w} e^{\lambda w}\, dw$$

$$\therefore \qquad \Psi(w) = \lambda\left(1 - \frac{\lambda}{\mu}\right) e^{-(\mu - \lambda)\, w}, \quad w > 0. \qquad \text{...(23.60)}$$

This result may also be obtained by using the *Laplace transform* of service time distribution, and the properties of the waiting time distribution may be found from its *Laplace transform.*

Obviously, $\int_0^{\infty} \Psi(w)\, dw \ne 1$, because it has the value λ/μ.

It is important to note that the case for which $w = 0$ has been excluded in eqn. (23.60). Thus,

Prob. $[W = 0]$ = Prob. [*no unit in the system*] $= P_0 = 1 - (\lambda/\mu)$. [from eqn. (23.56)]

Now, the sum of all probabilities of waiting time

$$= \int_0^{\infty} \text{Prob. } [w \le W \le w + dw] + \text{Prob. } [W = 0] \qquad \text{...(23.61)}$$

$$= \int_0^{\infty} \lambda\left(1 - \frac{\lambda}{\mu}\right) e^{-(\mu - \lambda) w}\, dw + \left(1 - \frac{\lambda}{\mu}\right)$$

$$= \frac{\lambda}{\mu} + \left(1 - \frac{\lambda}{\mu}\right) = 1$$

Hence it is concluded that the complete distribution for waiting time is partly continuous and partly discrete :

(*i*) ***continuous*** for $w \le W \le w + dw$ with probability density function $\Psi(w)$ given by eqn. (23.60); and

(*ii*) ***discrete*** for $W = 0$, with Prob $(W = 0) = 1 - (\lambda/\mu)$.

The probability that waiting time exceeds w is given by

$$\int_w^\infty \Psi(w)\,dw = \int_w^\infty \lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)w}\,dw = \left(-\frac{\lambda}{\mu}e^{-(\mu-\lambda)w}\right)_w^\infty = \frac{\lambda}{\mu}e^{-(\mu-\lambda)w} = \rho e^{-(\mu-\lambda)w}$$

which does not include the service time.

Q. 1. Define cumulative probability distribution of waiting time for a customer who has to wait and show that in an $(M \mid M \mid 1) : (\infty \mid FIFO)$ queue system, it is given by $1-\rho\, e^{-\mu t(1-\rho)}$ where $\rho = \lambda/\mu$.

[Hint. Cum. Distribution $= \int_0^t \lambda(1-\rho)\, e^{-\mu(1-\rho)w}\,dw = (1-\rho)$.] **[IGNOU 99 (Dec.)]**

2. Define $(M \mid M \mid 1)$ system.

IV. To find prob. distribution of time spent in the system (busy period distribution). **[Kanpur M.Sc. (Math.) 93]**

In order to find the probability density function for the distribution of total time (waiting + service) an arrival spends in the system, let $\Psi(w \mid w>0) =$ probability density function for waiting time such that *a person has to wait.*

The statement *"person has to wait"* is meant that the server remains busy in the *busy period.*

Applying the rule of conditional probability,

$$\Psi(w \mid w>0)\,dw = \frac{\Psi(w)\,dw}{\text{Prob. }(w>0)} = \frac{\Psi(w)\,dw}{\int_0^\infty \Psi(w)\,dw}.$$

Substituting the value for $\Psi(w)$ from eqn. (23.60),

$$\Psi(w \mid w>0)\,dw = \frac{\lambda(1-\lambda/\mu)\,e^{-(\mu-\lambda)w}\,dw}{\int_0^\infty \lambda(1-\lambda/\mu)\,e^{-(\mu-\lambda)w}\,dw} = \frac{\lambda(1-\lambda/\mu)\,e^{-(\mu-\lambda)w}\,dw}{\lambda/\mu}$$

or
$$\Psi(w \mid w>0) = (\mu-\lambda)\,e^{-(\mu-\lambda)w}. \qquad \text{...(23.62)}$$

Here, $\int_0^\infty \Psi(w \mid w>0)\,dw = \int_0^\infty (\mu-\lambda)\,e^{-(\mu-\lambda)w}\,dw = 1.$

Hence it gives the required probability density function for the busy period.

If service time is included then,

$$P(W \ge w) = \int_w^\infty (\mu-\lambda)\,e^{-(\mu-\lambda)w}\,dw = e^{-(\mu-\lambda)w}$$

V. MEASURES OF MODEL I :

(i) To find expected (average) number of units in the system, L_s.

By definition of expected value,

$$L_s = \sum_{n=1}^\infty n P_n = \sum_{n=1}^\infty n\left(\frac{\lambda}{\mu}\right)^n\left(1-\frac{\lambda}{\mu}\right) = \left(1-\frac{\lambda}{\mu}\right)\frac{\lambda}{\mu}\sum_{n=1}^\infty n\left(\frac{\lambda}{\mu}\right)^{n-1}$$

$$= \left(1-\frac{\lambda}{\mu}\right)\frac{\lambda}{\mu}\left[1+2\left(\frac{\lambda}{\mu}\right)+3\left(\frac{\lambda}{\mu}\right)^2+\ldots\infty\right] = \left(1-\frac{\lambda}{\mu}\right)\frac{\lambda}{\mu}\left[\frac{1}{(1-\lambda/\mu)^2}\right]^* \quad \textbf{(See foot-note)}$$

---

* Let $S = 1+2\left(\frac{\lambda}{\mu}\right)+3\left(\frac{\lambda}{\mu}\right)^2+\ldots\infty$, which is Arithmatico-Geometric series,

$\therefore \quad \frac{\lambda}{\mu}S = \left(\frac{\lambda}{\mu}\right)+2\left(\frac{\lambda}{\mu}\right)^2+\ldots\infty$

On subtracting,

$\left(1-\frac{\lambda}{\mu}\right)S = 1+\left(\frac{\lambda}{\mu}\right)+\left(\frac{\lambda}{\mu}\right)^2+\ldots\infty = \frac{1}{1-\lambda/\mu}$ (sum of infinite G.P.)

$\therefore \quad S = \frac{1}{(1-\lambda/\mu)^2}.$

or $$L_s = \frac{\lambda/\mu}{(1-\lambda/\mu)} = \frac{\rho}{1-\rho}, \text{ where } \rho = \lambda/\mu < 1 \quad ...(23.63)$$

which is the required formula.

Q. 1. In a certain queueing system with one server, the arrivals obey a Poisson distribution with mean λ and the service time distribution has mean $1/\mu$. Obtain the generating function of the length of the queue which a departing customer leaves behind him.

2. Show that for a single service station, Poisson arrivals and exponential service time, the probability that exactly *an calling units are in queueing system is* $P_n = (l-\rho)\rho^n, n \geq 0$ (*ρ is the traffic intensity*). Also, find the expected line length. **[B.H.U. 93]**

3. Show that average number of units in a $M \mid M \mid 1$ system is equal to $\rho/(1-\rho)$. **[Agra 99; Raj. Univ. (M. Phil) 93]**

4. Discuss (M | M | 1) : (∞ | FCFS) queueing model and find the expected line length $E(L_s)$ in the system. **[Garhwal M.Sc. (Math.) 96]**

5. For the M | M | 1 queueing system, find :
   (a) Expected value of queue length n
   (b) Prob. distribution of waiting time w. **[Meerut M.Sc. (Math.) BP-96]**

6. Explain a system with Poisson input, exponential waiting time with single chennel. Also determining the average length of the waiting time. **[Meerut 2002]**

(*ii*) **To find expected (average) queue length, L_q :** **[IGNOU 1999; Meerut 97 P, 93; Kanpur 93]**

Since there are $(n-1)$ units in the queue excluding one being serviced,

$$L_q = \sum_{n=1}^{\infty}(n-1)P_n = \sum_{n=1}^{\infty} nP_n - \sum_{n=1}^{\infty} P_n = \sum_{n=1}^{\infty} nP_n - \left[\sum_{n=0}^{\infty} P_n - P_0\right] = L_s - [1-P_0] \quad \left(\text{since } \sum_{n=0}^{\infty} P_n = 1\right)$$

Substituting the value of P_0 from (23.56), we have

$$L_q = L_s - 1 + \left(1 - \frac{\lambda}{\mu}\right) \quad ...(23.64)$$

or $$L_q = L_s - \frac{\lambda}{\mu} = \frac{\rho^2}{1-\rho}, \text{ where } L_s = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\rho}{1-\rho}.$$

Q. 1. For $M \mid M \mid 1$ model, obtain the expected number of waiting customers if the queueing process is going on for a long time.

2. Describe the system of steady state equations for a queueing model ($M \mid M \mid 1$) : (*FCFS*, ∞) and obtain their solution. Obtain the mean queue length and mean number of units in the system. **[Meerut (Stat.) 95, 90]**

(*iii*) **To find mean (or expected) waiting time in the queue (excluding service time), W_q :**

Since expected time an arrival spends in the queue is given by

$$W_q = \int_0^{\infty} w\Psi(w)\,dw = \int_0^{\infty} w.\lambda\left(1-\frac{\lambda}{\mu}\right)e^{-(\mu-\lambda)w}\,dw \qquad \text{[from eqn. (23.60)]}$$

Integrating by parts,

$$= \lambda\left(1-\frac{\lambda}{\mu}\right)\left[w\cdot\frac{e^{-(\mu-\lambda)w}}{-(\mu-\lambda)} - \frac{1}{(\mu-\lambda)^2}e^{-(\mu-\lambda)w}\right]_0^{\infty} = \lambda\left(\frac{\mu-\lambda}{\mu}\right)\frac{1}{(\mu-\lambda)^2}$$

$$\therefore \quad W_q = \frac{\lambda}{\mu(\mu-\lambda)} \quad ...(23.65)$$

(*iv*) **To find expected waiting time in the system (including service time), W_s :**

Since expected waiting time in the system = Expected waiting time in queue + expected service time, i.e. $W_s = W_q + 1/\mu$ (expected service time or mean service time $= 1/\mu$). Substituting the value of W_q from eqn. (23.65), we get

$$W_s = \frac{\lambda}{\mu(\mu-\lambda)} + \frac{1}{\mu} = \frac{1}{\mu-\lambda} \quad ...(23.66)$$

(*v*) **To find expected waiting time of a customer who has to wait, $(W \mid W > 0)$.**

The expected length of the busy period is given by

$$(W \mid W>0)=\int_0^{\infty} w\Psi(w>0)\,dw=\int_0^{\infty} w\,.\,(\mu-\lambda)\,e^{-(\mu-\lambda)w}\,dw \quad \text{[from eqn. (23.62)]}$$

Integrating by parts, we get

$$(W \mid W>0)=\frac{1}{\mu-\lambda}=\frac{1}{\mu(1-\rho)} \qquad ...(23.67)$$

(vi) To find expected length of non-empty queue, $(L \mid L>0)$.

By definition of conditional probability,

$$(L \mid L>0)=L_s/\text{ Prob. (an arrival has to wait, } L>0)$$
$$=L_s/(1-P_0) \text{ (since probability of an arrival not to wait is } P_0)$$

Substituting the value of L_s and P_0 from eqns. (5.63) and (23.56), we get

$$(L \mid L>0)=\frac{(\lambda/\mu)/(1-\lambda/\mu)}{\lambda/\mu}=\frac{\mu}{\mu-\lambda}=\frac{1}{1-\rho} \qquad ...(23.68)$$

[Kanpur 93]

(vii) To find the variance of queue length.

By definition,

$$\text{Var.}\{n\}=\sum_{n=1}^{\infty} n^2 P_n-\left(\sum_{n=1}^{\infty} n P_n\right)^2=\sum_{n=1}^{\infty} n^2 P_n-[L_s]^2$$

$$=\sum_{n=1}^{\infty} n^2(1-\rho)\,\rho^n-\left(\frac{\rho}{1-\rho}\right)^2 \qquad \text{(using (23.57) and (23.63))}$$

$$=(1-\rho)\,[1^2\rho+2^2\rho^2+3^2\rho^3+...]-\rho^2/(1-\rho)^2$$

$$=(1-\rho)\,\rho\,[1+2^2\rho+3^2\rho^2+...]-\rho^2/(1-\rho)^2$$

$$=(1-\rho)\,\rho\left[\frac{1+\rho}{(1-\rho)^3}\right]^*-\frac{\rho^2}{(1-\rho)^2} \qquad \text{[see foot note]}$$

$$=\rho/(1-\rho)^2$$

Q. 1. Obtain the steady state equations for the model $(M \mid M \mid 1):(\infty \mid FCFS)$ *i.e.* single server, Poisson arrival, negative exponential service), and also find the formula for :
(*i*) variance of the queue length, (*ii*) the average waiting length, (*iii*) Prob. Queue size $\geq N$,
(*iv*) The average (mean) queue length., (*v*) the average waiting length given that it is greater than zero, **[Raj. Univ. (M. Phil) 90]**
(*vi*) The average number of customers in the system.

2. Define the concept of busy period in queueing theory and obtain its distribution for the system $M \mid M \mid 1:(\infty \mid FCFS)$.
Show that the average length of the busy period is $1/(\mu-\lambda)$.

3. Customers arrive at a sales counter in a Poisson fashion with mean arrival rate λ and exponential service times with mean service rate μ. Determine :
(*i*) Average length of non-empty queues, (*ii*) Average waiting time of an arrival.

4. For the queueing system in which there is a single channel and the inter-arrival time of units and the service time of units follow exponential distribution prove that :
(*i*) $1/(\mu-\lambda)$ is the average time an arrival spends in the system, (*ii*) $\lambda^2/\mu(\mu-\lambda)$ is average queue length.

(viii) To find the probability of arrivals during the service time of any given customer.

Since the arrivals are Poisson and service times are exponential, the probability of r arrivals during the service time of any given customer is given by

---

* Let $S=1+2^2\rho+3^2\rho^2+...$
Integrating both sides in the limit 0 to ρ

$$\int_0^{\rho} S\,d\rho=\rho+2\rho^2+...=\rho(1-\rho)^{-2}$$

[See foot-not on p. 832]

Now, differentiating w.r.t. 'ρ'

$$S=\frac{1}{(1-\rho)^2}+\frac{2\rho}{(1-\rho)^3}=\frac{(1+\rho)}{(1-\rho)^3}.$$

$$K_r = \int_0^\infty P_r(t)\, s(t)\, dt = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^r}{r!} \cdot \mu e^{-\mu t}\, dt = \frac{\lambda^r \mu}{r!} \int_0^\infty e^{-(\lambda+\mu)t}\, t^r\, dt$$

$$= \frac{\lambda^r \mu\, \Gamma(r+1)}{r!\,(\lambda+\mu)^{r+1}} \quad \left[\text{using } \int_0^\infty e^{-at}\, t^n\, dt = \frac{\Gamma(n+1)}{a^{n+1}}\right]$$

$$= \left(\frac{\lambda}{\lambda+\mu}\right)^r \cdot \frac{\mu}{\lambda+\mu} \qquad [\text{since } \Gamma(r+1) = r!]$$

23.12-1. Inter-Relationship Between L_s, L_q, W_s, W_q.

It can be proved under (rather) general conditions of arrival, departure, and service discipline that the formulae,

$$L_s = \lambda W_s, \qquad \text{...(23.69)}$$

and

$$L_q = \lambda W_q, \qquad \text{...(23.70)}$$

will hold in general. These formulae act as key points in establishing the strong relationships between W_s, W_q, L_s and L_q which can be found as follows.

By definition,
$$W_q = W_s - 1/\mu. \qquad \text{...(23.71)}$$

Thus, multiplying both sides by λ and substituting the values from (23.69) and (23.70),

$$L_q = L_s - \lambda/\mu. \qquad \text{...(23.72)}$$

This means that one of the four expected values (together with λ and μ) should immediately yield the remaining three values.

Q. In (M | M | 1) : (∞ | FCFS) model obtain p.d.f. of waiting time (excluding service time) and hence obtain $E(W_q)$, $E(W_s)$, $E(L_q)$, $E(L_s)$. **[Garhwal M.Sc. (Stat.) 93]**

23.12-2. Illustrative Examples on Model I

Example 1. *A TV repairman finds that the time spent on his jobs has an exponential distribution with mean 30 minutes. If he repairs sets in the order in which they come in, and if the arrival of sets is approximately Poisson with an average rate of 10 per 8-hour day, what is repairman's expected idle time each day ? How many jobs are ahead of the average set just brought in ?*

[JNTU (B. Tech.) 2002; Agra 98, 93; Karnataka B.E. (CSE) 93; Meerut (Maths.) 91]

Solution. Here, $\mu = 1/30$, $\lambda = 10/(8 \times 60) = 1/48$. Therefore, expected number of jobs are

$$L_s = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda} = \frac{1/48}{1/30 - 1/48} = 1\tfrac{2}{3} \text{ jobs.}$$ **Ans.**

Since the fraction of the time the repairman is busy (*i.e.* traffic intensity) is equal to λ/μ, the number of hours for which the repairman remains busy in a 8-hour day is

$$= 8\,.\,(\lambda/\mu) = 8 \times 30/48 = 5 \text{ hours.}$$

Therefore, the time for which the repairman remains idle in 8-hour day = (8 – 5) hours = 3 hours. **Ans.**

Example 2. *At what average rate must a clerk at a supermarket work in order to ensure a probability of 0.90 that the customer will not have to wait longer than 12 minutes ? It is assumed that there is only one counter to which customers arrive in a Poisson fashion at an average rate of 15 per hour. The length of service by the clerk has an exponential distribution.* **[Meerut (Maths.) 99, 96]**

Solution. Here, $\lambda = 15/60 = 1/4$ customer/minute, $\mu = ?$ Prob. [waiting time ≥ 12] = 1 – 0.90 = 0.10.

Therefore,
$$\int_{12}^\infty \lambda\left(1 - \frac{\lambda}{\mu}\right) e^{-(\mu-\lambda)w}\, dw = 0.10 \quad \text{or} \quad \lambda\left(1 - \frac{\lambda}{\mu}\right)\left[\frac{e^{-(\mu-\lambda)w}}{-(\mu-\lambda)}\right]_{12}^\infty = 0.10$$

or
$$e^{(3-12\mu)} = 0.4\,\mu \quad \text{or} \quad 1/\mu = 2.48 \text{ minute per service.}$$ **Ans.**

Example 3. *Arrivals at a telephone booth are considered to be Poisson, with an average time of 10 minutes between one arrival and the next. The length of a phone call assumed to be distributed exponentially with mean 3 minutes. Then,*