BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Operations research

Course Code: 23ST01UEC

Unit-V

Queueing Theory

Dr. T. Jai Sankar

Associate Professor and Head

Department of Statistics

Ms. E. Devi

Guest Faculty

Department of Statistics

INTRODUCTION

Queueing or waiting in a line is a common situation occurring in everyday life. We wait in queues in ticket booths, bus stops, post offices, banks, traffic lights and so on.We can think of many more situations. In general, a queue is formed when there are customers who require some sort of services and the queuing problem is identified by the presence of a group of customers who arrive randomly to receive some service.

The customer may be a person, machine, vehicle or anything else which requires service. The objective of a queuing model is to find out the optimum service rate and the number of servers so that the average cost of being in queuing system and the cost of service are minimized.

The objective of a queueing system is to find ways of reducing the time spent in waiting by the customer and at the same time optimizing the cost to the service provider. Sometimes Queueing Theory, also reference as the waiting line theory. It was developed in 1909 when A.K. Erlang made an effort to analyses telephone traffic congestion. It can be applied to wide variety of operational situations whenever customer's expectations do not match with objectives of servers due to one's inability to predict accurately the arrival and service time of customers. The purpose of queueing analysis is to provide information to evaluate an acceptable level of service and service capacity since providing too much service capacity is costly (because of employees or equipment). In fact, providing too little service capacity is also costly.

In this Chapter, we discuss the concepts of stochastic processes, Poisson process and birth -death process. We explain the basic component, fundamental structure and operating characteristics of a queueing system. Also, we describe the M/M/1, queueing models.

Objectives:

The objective of these contents is to provide some important concepts/results to the reader like:

- Poisson and birth-death process.
- Fundamental structure and operating characteristics of a queueing system.
- M/M/1 Queueing Model.

BASIC CONCEPTS OF QUEUEING THEORY

For studying queueing systems, one should be familiar with the probability theory, specially the concept of random variable and its probability distribution such as Poisson distribution, Exponential distribution.

Let us have a glance on some basic definitions.

Consider an experiment whose outcome is not uniquely determined. In such a situation an observed outcome of the experiment is one from a set of possible outcomes.

Sample piont: An outcome of an experiment is called a sample point.

Sample Space: The set of all possible outcomes of a random experiment is called sample space.

Events: Subsets of the sample space are called events.

Random Variable: A random variable is a function that associates a point of the sample space with a real number.

Random Process/Stochastic Process: A random process or stochastic process is a family (or collection) of random variables.

For example, if a die with six faces numbered 1, 2... 6 is thrown, the set S is a sample space where the set of all possible outcomes is S= {1, 2, 3, 4, 5, 6}. A function X that assigns to an outcome or sample point, the number written on it, is the random variable. The event that an odd number is observed corresponds to the set {1, 3, 5} which is a subset of S.

Let us take more examples of stochastic process.

I. If $n \geq 1$, Suppose Xn is the outcome of the nth throw. Then $\{Xn, n \geq 1\}$ is a collection of random variables, such that for a distinct value of n, one gets a distinct random variable Xn.

II. The sequence $\{Xn, n \geq 1\}$ constitutes a random (stochastic) process called the Bernoulli process.

III. Let Xn represents the number of sixes in the first n throws. For each value of n=1, 2, ..., we get a distinct Binomial random variable.

IV. The sequence Xn:{ Xn, n≥1}, which gives a collection of random variables, is a random or stochastic process. Suppose a telephone call is received at switchboard.

V. Let Xt be the random variable, which represents the number of incoming calls in an interval (0, t). Then Xt is a random variable and the family {Xt, t ε T} constitutes a stochastic process, where T is the interval 0≤t≤∞.

Poisson Process

Suppose X(t) represent the maximum temperature at a particular place. Here we deal with discrete state, continuous time stochastic processes. The Poisson process is one of the representatives of this type of stochastic processes. The Poisson process may be explain as follows:

Let E be a random event such as (i) incoming telephone calls at a switchboard, (ii) arrival of patients for treatment at a clinic, (iii) occurrence of accidents at a certain place. We consider the total number N(t) of occurrences of an event E in an interval of time 't'. Suppose Pn(t) is the probability that the random variable N(t) assumes the value n, i.e.,

$$Pn(t) = P[N(t) = n] \rightarrow (1)\,, \quad \text{For n} = 0, 1, 2, 3.\dots\dots\dots$$

We have,

$$\sum_{n=}^{\infty} p(t) = 1 \ \rightarrow (2) \qquad F\text{or each fixed t...}$$

We can thus say from equation (2) that Pn(t) is a probability mass function of the random variable N(t) and the family of random variables [Nt, t> 0] is a stochastic process. From our earlier discussion, you may understand that this family is continuous parameter (in this case, time) stochastic process with a discrete state space. This is called a Poisson process. Under certain conditions, N(t) follows a Poisson distribution with mean λt (λ is being constant). This holds for most practical situations.

Assumptions in Poisson Process

- Events must be independent, in other words the number of the customers which is arrive in disjoint time intervals are statistically independent e.g. the number of goals scored by a team should not make the number of goals scored by another team more or less likely and the mean number of goals scored is assumed to be the same for all teams.

- The probability of two events occurring between time t and t + Δtis 0(Δ(t)), i.e. negligible. Thus,

$$P0(\Delta t) + P1(\Delta t) + 0(\Delta t) = 1$$

- The probability that event E occurs between time t and Δt is equal to λΔt+ 0(Δ)t). Thus,

$$P1(\Delta t) \;=\; \lambda\Delta t + 0(\Delta t),$$

Under the assumptions stated above, N(t) follows a Poisson distribution with mean λt, i.e. Pn(t) is given by

$$pn(t) = \frac{e-\lambda t(\lambda t)n}{n!} \qquad n = 0,1,2,3, \dots \dots ....$$

To formulate a queueing model, we have to specify the assumed form of the probability distributions of both inter arrival times and service e times. You have learnt that inter- arrival times follow the Poisson distribution. Similarly, most of the times, the service times in a queueing system follow the exponential distribution. Hence, the exponential distribution is the most important distribution in queueing theory

Markov Process

Around the beginning of the 20th century, the methodology discussed here was developed by the Russian mathematician. A.A. Markov. The Markov Process forms a sub-class of the set of all random processes. It is a sub class with enough simplifying assumptions to make them easy to handle.

A stochastic (or random) process is called a Markov process if the occurrence of a future state depends on the immediately preceding state and only on it. If, for all tn, tn-1, ...,t_0, satisfying tn>tn-1 >.....$t_0$and non- negative integer j, the family of random variables {X(t), t≥ 0} is said to be a Markov process if it satisfy the Markovian property:

$$P\left[X(tn) = \frac{jn}{X(tn-1)} = jn - 1, \dots \dots .., X(t0) = j0\right]$$

$$P[X(tn) = jn/X(tn-1) = jn - 1],$$

The process has the Markov property and is called a continuous time Markov process.

Birth and Death Process

The arrival process assumes that the customers arrive at the queueing system and never leave it. Such a system is known as pure birth process. On the other hand, the departure process assumes that no customer joins the system while service is continued for those who are already in the system. If at time t=0, N is greater than or equal to one i.e. at starting time number of customers is N. Since service is being provided at the rate of μ, therefore customers leave the system at the rate of μ being serviced. This type of process is known as pure death process.

Let the state at time t would correspond to the number of arrivals by time t. Pn(t) denotes the probability that there are n customers in the system at any time t, both waiting and being served. Arrivals can be considered as births. If the system is in state En and arrival occurs, the state is changed to En+1. Similarly, a departure can be looked upon as death. A departure occurring while the system is in state En changes the system to the state En−1.This type of process is generally referred to as a birth – death process.

Assumptions in the Birth-Death Process

The assumptions of the birth-death process are as follows:

- If the system is in state En, the current probability distribution of the time t until the next arrival (birth) is exponential with parameterλn, where n = 0, 1, 2, ...
- If the system is in stateEn, the current probability distribution of the time t until the next departure or service completion (death) is exponential with parameter μn,where n = 0, 1, 2,.... Only one birth or death can occur in a small interval of time, i.e.,Δt.

FUNDAMENTAL STRUCTURE OF A QUEUING SYSTEM

Queuing theory is related with the mathematical study of queues or waiting lines, a queue is formed when there are customers who require some sort of services and the current demand for a service exceeds the current capacity to provide the service.

Generally, the customer’s arrival and their service time are not known in advance or can’t be predicted accurately. Since arrival-departure process are random. So, queuing models developed to reduce waiting time/excessive costs and work for maintaining balance between service capacity and waiting time.

A simple queuing system can be described as follows:

- Input or arrival process of customers
- Service mechanism (or process)
- Queue discipline

The fundamental structure of a queueing system shown in the figure given below

Now we explain all the components of queueing system.

1. Input or Arrival Process of Customers

The rate at which the customers arrive at the service facility is determined by the arrival process. An input/arrival process can be defined completely by its size, the arrival time distribution, and the attitude of the customers. We describe these, in brief.

- **Size:** It may be finite or infinite according as the arrival rate is affected or not affected by the number of customers in the service system.
- **The arrival time distribution:** Mostly, the arrival time distribution is approximated by Poisson distribution.
- **Customer or arrivals behavior:** A customer who stay in the system until served no matter how much he has to wait for service. Such a customer is called Patient Customer.
- The customer who wait for a certain time in the queue and leave the system without getting service. This kind of customer is known as impatient or reneging behavior.
- If a customer before joining the system get discouraged by seeing the number of customers already in the queue is too large and does not join the queue.
- This behavior of the customer is called Balking behavior. Customers who move from one queue to another because they think that their queue is moving slower, the behavior of the customer is known as queue jockeying.

Remark: Generally, it is assumed that the customers arrive into the system one at a time. But, sometimes, customers may arrive in groups and such arrival is called Bulk arrival.

2. Service Mechanism (or Process)

Service time distributions are generally exponential distributions. It may be any one of the following types:

- Single channel facility
- One queue-several station facilities
- Several queues-one service station
- Multi-channel facility and
- Multistage channel facility

3. **Queue Discipline**

Queue discipline is the order or the manner in which the service station selects the next customer from the waiting line to be served. There are many ways in which a customer to be selected for service. Some of these are described as follows:

- FIRST IN, FIRST OUT (FIFO) or in other words First Come First Served (FCFS);
- Last in First Out (LIFO); and
- Service in Random Order (SIRO)

Throughout the chapter, we consider the FCFS queue discipline.

OPERATING CHARACTERISTICS OF A QUEUEING SYSTEM

For studying queuing system, we have to set up a set of equations. We have to solve these equations to determine the operating characteristics (or performance measures). There are two types of solutions of these equations**: (i) Transient (ii)Steady state.**

Transient solutions: The time dependent solutions are known as transient solutions.

Steady state solutions: These solutions are independent of time and represent the probability of the system being in a particular state in the long run.

Note: Here we shall be analysing the system under the steady state condition.

Operating characteristics/ Performance Measures

Performance Measures of a queuing system are determined by two statistical properties, namely, the probability distribution of inter-arrival times and the probability distribution of service times.

Some of the operating characteristics/performance measures of any queueing system that are of general interest for analysing the system are listed below:

- **Ls**: The average number of customers in the queuing system (those waiting to be served and those being served).
- **Ws**:The average time each customer spends in the queuing system from entry into the queue to completion of the service (the time spent waiting in the queue and during the service).
- **Lq**: The average number of customers in the queue waiting to get service (this excludes customers undergoing service).
- **Wq**: The average time each customer spends in the queue waiting to get service (this excludes customers time spent during the service).
- **Service idle time:** The relative frequency with which the service system is idle.

Now we describe the equations for each of the operating characteristics listed above under steady state solutions. By definition of expectation, we have

$Ls = E(n) = \sum_{n=1}^{\infty} np_n$ (Expected number of customers in the system) (1)

$Lq = \sum_{n=c+1}^{\infty} (n-c)p_n$ (Expected number of customers in queue) (2)

where there are c parallel servers so that c customers can be served simultaneously with λ arrival rates for those who join the system, we have

$$Ls = \lambda W_s \rightarrow (3)\ and$$

$$Lq = \lambda W_q \rightarrow (4)$$

If μ is the service rate, the expected service time is $\frac{1}{\mu}$ and we have

$$Ws = Wq + \frac{1}{\mu} \rightarrow (5)$$

Where, W_s denotes expected waiting time in system and W_q represents expected waiting time in queue Multiplying both sides of equation (5) with λ, we get

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

$$L_s = L_q + \frac{\lambda}{\mu} \rightarrow (6) \qquad \text{(By (3) and (4))}$$

Classification of Queueing Models

Generally, queueing models are described by five symbols such as a/b/c: d/e or a/b/c/d/e. The first symbol 'a' describes the arrival process. The second symbol 'b' describes the service time distribution. The third symbols 'c' stands for the number of servers. The symbols 'd' and 'e' stand for the system capacity and queue discipline respectively.

First three symbols i.e. a/b/c in the above notation were described by D. Kendall in 1953. Later, A. Lee in 1966 added the fourth (d) and fifth (e) to the Kendall notation.

Remark: The fifth symbol (e) from the notation can be omitted if the system has FCFS queue discipline and for this case it can be described as a/b/c/d. Moreover, if the system has infinite capacity with FCFS queue discipline then we can use simply a/b/c or a/b/c:∞ /FCFS for describing the queueing system.

Note: If arrivals follow Poisson distribution and departures follow exponential distribution, symbol M is used in place of 'a' and 'b' e.g. suppose there is a single server then we can use the notation M/M/1(here we consider infinite capacity and FCFS queue discipline).

M/M/1 OR M/M/1: ∞/FCFS QUEUEING MODEL

This model deals with a queueing system having a single server channel and no limit on system capacity with Poisson input and exponential distribution for services while the customers are served on a "First Come First Served" basis. For the given model, arrivals follow Poisson distribution and hence inter-arrival times have an exponential distribution. The service time for this model follows an exponential distribution. If λ and μ denote the average arrival rate and average service rate respectively then $1/\lambda$ and $1/\mu$ show mean arrival time and mean service time correspondingly. Then the ratio .$\rho=\frac{\lambda}{\mu}$ is called the traffic intensity or the utilisation factor. It is the measure of the degree to which the capacity of the service station is utilised. For example, if customer arrives at a rate of 10 per minute and the service rate 20 per minute, the utilisation of the service capacity is $\frac{10}{20}$= 50%, i.e., the service facility is kept busy 50% of the time and remains idle 50% of the time.

Arrival-Departure Equations for the M/M/1 Queuing Model (OR)

System of Differential-Difference Equations

If $n \geq 1$ the probability that the system has n customers at time $(t + \Delta t)$ can be expressed as: $p_n(t + \Delta t) =$[{Probability of n – 1 customers in the system at time t and (1 arrival, no departure in timeΔt or 2 arrivals, one departure in time Δt or 3 arrivals, two departure in time Δtor ...)} + {Probability of n – 1 customers in the system at time t and (2 arrivals, no departure in time Δt or 3 arrivals, one departure in time Δt or 4 arrivals, two departures in time Δt or ...)} + {Probability of n – 3 customers in the system at time t and (3 arrivals, no departure in time Δt or 4 arrivals, one departure in time Δt or 5 arrivals, two departures in time Δt or ...)} + ...] + [{Probability of n + 1 customers in the system at time t and (1 departure, no arrival in time Δt or 2 departures, one arrival in time Δt or 3 departures, two arrivals in time Δt or ...)} + {Probability of n + 2 customers in the system at time t and (2 departures, no arrivals in time

Δt or 3 departures, one arrival in time Δt or ...)} + {Probability of n + 3 customers in the system at time t and (3 departures, no arrival in time Δt or ...)} + ...] + [{Probability of n customers in the system at time t and (no arrival, no departure in time Δt or 1 arrival, one departure in time Δt or 2 arrivals, two departures in time Δt or ...)]

$$= P_{n-1}(t)\left[\{\lambda\Delta t + O(\Delta t)\}\{1-\mu\Delta t + O(\Delta t)\} + \text{The terms equal to } O(\Delta t)\right]$$

$$+P_{n-2}(t)\left[\{O(\Delta t)\}\{1-\mu\Delta t + O(\Delta t)\} + \text{The terms equal to } O(\Delta t)\right]$$

$$+P_{n-3}(t)O(\Delta t) + \text{The terms equal to } O(\Delta t)$$

$$+P_{n+1}(t)\left[\{(\mu\Delta t + O(\Delta t))\}\{1-\lambda\Delta t + O(\Delta t)\} + \text{The terms equal to } O(\Delta t)\right]$$

$$+P_{n+2}(t)\left[\{O(\Delta t)\}\{1-\lambda\Delta t + O(\Delta t)\} + \text{The terms equal to } \mathrm{O}(\Delta t)\right]$$

$$+P_{n+3}(t)O(\Delta t) + \text{The terms equalto } O(\Delta t)$$

$$+P_{n}(t)\left[\{1-\mu\Delta t + O(\Delta t)\}\{1-\lambda\Delta t + O(\Delta t)\}\right.$$

$$\left.+\{\mu\Delta t + O(\Delta t)\}\{\lambda\Delta t + O(\Delta t)\} + \text{The terms which ultimately become equal to } O(\Delta t)\right]$$

Notice that $\Delta t \cdot O(\Delta t), O(\Delta t)\cdot O(\Delta t)$, $O(\Delta t)+O(\Delta t), O(\Delta t)-O(\Delta t)$, $\Delta t \cdot \Delta t$ are very small and may be taken as equal to $O(\Delta t)$

Notice that $\Delta t \cdot O(\Delta t), O(\Delta t)\cdot O(\Delta t)$, $O(\Delta t)+O(\Delta t), O(\Delta t)-O(\Delta t)$, $\Delta t \cdot \Delta t$ are very small and may be taken as equal to $O(\Delta t)$

We deal with the case $n = 0$ separately because there cannot be any possibility of less than zero customers. However, in the above equation, the case of less than n customers is included.

The probability of 0(Zero)customers in the system at time $t+\Delta t$ = [{Probability of 1 customer in the system at time t and (1 departure, no customer/arrival in time Δt or 2 departure, 1 arrivals in time Δt or 3 departures, two arrivals in time Δt)} + {Probability of 2 customers in the system at time t and (2 departures, no arrival in time Δt or 3 departures, one arrival in time Δt or 4 departures, 2 arrivals in time Δt)} + {Probability of 3 customers in the system at time t and (3 departures, no arrival in time Δt or 4 departures, one arrival in time Δt or 5 departures, two arrivals in time Δt)} + …] + [{Probability of O customers in the system at time t and (no arrival, no departure in time Δt or 1 arrival, one departure in time Δt or 2 arrivals, 2 departures in time Δt)}

$$= P_1(t)\left[\{\mu\Delta t + O(\Delta t)\}\{1-\lambda\Delta t + O(\Delta t)\} + \text{The terms which ultimately become equal to } O(\Delta t)\right]$$

$$+P_2(t)\left[\{0(\Delta t)\}\{1-\lambda\Delta t + O(\Delta t)\} + \text{The terms which ultimately become equal to } O(\Delta t)\right]$$

$$+P_3(t)O(\Delta t) + \text{The terms which ultimately equal to } O(\Delta t)]$$

$P_O(t)$ [{Probability of no departure which 1 as there is no customer in the system and hence no chance of departure}

$\{1-\lambda\Delta t+O(\Delta t)\}\{\{\mu\Delta t+O(\Delta t)\}\{\lambda\Delta t+O(\Delta t)\}$ + The terms which ultimately become equal to $O(\Delta t)]$

Therefore, the arrival-departure equations for the M/M/1 Queuing model are:

$$P_n(t+\Delta t)=P_{n-1}(t)\left[\lambda\Delta t+O(\Delta t)\right]+P_{n+1}(t)\left[\mu\Delta t+O(\Delta t)\right]$$
$$+P_n(t)\left[\lambda\Delta t+O(\Delta t)\right]\left[\mu\Delta t+O(\Delta t)\right]$$
$$+P_n(t)[1-\lambda\Delta t][1-\mu\Delta t]+O(\Delta t),\, n\ge 1$$

$$P_O(t+\Delta t)=P_1(t)\left[\mu\Delta t+O(\Delta t)\right]+P_O(t)\left[\lambda\Delta t+O(\Delta t)\right]$$
$$+P_O(t)[1-\lambda\Delta t]\cdot 1+O(\Delta t),\, n=0 \tag{1}$$
$$P_n(t+\Delta t)=P_{n-1}(t)\lambda\Delta t+P_{n+1}(t)\mu\Delta t$$
$$+P_n(t)[1-\lambda\Delta t-\mu\Delta t]+O(\Delta t),\, n\ge 1$$
$$P_O(t+\Delta t)=P_1(t)\mu\Delta t+P_0(t)[1-\lambda\Delta t]+O(\Delta t) \tag{2}$$

Dividing equation (2) by Δt and taking the limit as Δt tends to zero, we have

$$\underset{\Delta t\to 0}{\text{limit}}\frac{P_n(t+\Delta t)-P_n(t)}{\Delta t}=\lambda P_{n-1}(t)+\mu P_{n+1}(t)-(\lambda+\mu)P_n(t)$$
$$\Rightarrow P_n^1(t)=\lambda P_{n-1}(t)+\mu P_{n+1}(t)-(\lambda+\mu)P_n(t) \tag{3}$$

and

$$P_O^1(t)=-\lambda P_O(t)+\mu P_1(t) \tag{4}$$

When steady state, i.e., the equilibrium state, is reached, $P_n^1(t)$ becomes independent of time, say p_n, and the rate of its change with respect to time becomes zero, i.e.,

$P_n^1(t)=0$.

Therefore, the **steady-state solution** is given by

$$\lambda P_{n-1}+\mu P_{n+1}-(\lambda+\mu)P_n=0,\, n\ge 1$$
$$\mu P_1=\lambda P_0 \tag{5}$$

Now, from equation (5), we have

$$\mu P_{n+1} - \lambda P_n = \mu P_n - \lambda P_{n-1} \tag{6}$$

This implies that

$\mu P_n - \lambda P_{n-1} = \mu P_{n-1} - \lambda P_{n-2}$ (changing n to $n-1$)

$\mu P_{n-1} - \lambda P_{n-2} = \mu P_{n-2} - \lambda P_{n-3}$ (again, changing n to $n-1$)

$\mu P_{n-2} - \lambda P_{n-3} = \mu P_{n-3} - \lambda P_{n-4}$ (again, changing n to $n-1$)

…

…

…

$\mu P_2 - \lambda P_1 = \mu P_1 - \lambda P_0$

But $\mu P_1 - \lambda P_0 = 0$ (from (5))

Therefore, $\mu P_{n+1} - \lambda P_n = \mu P_n - \lambda P_{n-1}$

$= \mu P_{n-1} - \lambda P_{n-2}$

$\vdots$

$= \mu P_1 - \lambda P_0 = 0$

$$\Rightarrow P_n = \frac{\lambda}{\mu} P_{n-1}$$

$$= \frac{\lambda}{\mu}\left(\frac{\lambda}{\mu} P_{n-2}\right) \quad \left(\because \mu P_{n-1} - \lambda P_{n-2} = 0 \Rightarrow P_{n-1} = \frac{\lambda}{\mu} P_{n-2}\right.$$

$$= \frac{\lambda}{\mu}\frac{\lambda}{\mu}\cdot\left(\frac{\lambda}{\mu} P_{n-3}\right)\left(\because \mu P_{n-2} - \lambda P_{n-3} = 0 \Rightarrow P_{n-2} = \frac{\lambda}{\mu} P_{n-3}\right)$$

$$= \frac{\lambda\cdot\lambda\cdot\lambda\ldots\lambda(n \text{ times})}{\mu\cdot\mu\cdot\mu\ldots\mu(n \text{ times})} P_0 = \frac{\lambda^n}{\mu^n} P_0$$

$$= \left(\frac{\lambda}{\mu}\right)^n P_0 = P^n P_0.$$

Since $\sum_{n=0}^{\infty} P_n = 1$, it follows that

$$\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} P_0 P^n = 1 \quad (\because P_n = P\ldots\ldots\ldots\ldots)$$

$$\Rightarrow P_0\left(1+HP+P^2+\ldots+P^n+\ldots\right)=1$$

$$\Rightarrow P_0\left(\frac{1}{1-P}\right)=1$$

$$\Rightarrow P_0=1-P \text{ Where } P=\frac{\lambda}{\mu}$$

Therefore, the probability of n customers (units) in the system is given by

$$P_n=(1-P)P^n,\ n\geq 0$$

Model → M/M/1 or M/M/1: ∞/FCFS

Probability of n customers in the system is given by

$$\mathrm{p_n}=(1-\rho)\rho^{\mathrm{n}},\mathrm{n}\geq 1$$

Probability of zero customers (units) in the queue/system is given by $\mathrm{p_0}=1-\rho$

Operating Characteristics or Measures of Performances of the Model M/M/1

1. The average number of customers (units) in the system is given by

$$L_s=E(n)=\sum_{n=0}^{\infty}nP_n=\sum_{n=0}^{\infty}n(1-\rho)\rho^n=(1-\rho)\sum_{n=0}^{\infty}n\rho^n \quad \ldots(*)$$

Now, consider $\sum_{n=0}^{\infty}n\rho^n=0+\rho+2\rho^2+3\rho^3+\ldots$

Let $S=\sum_{n=0}^{\infty}n\rho^n=\rho+2\rho^2+3\rho^3+\ldots$

$$\Rightarrow \rho S=\rho^2+2\rho^3+3\rho^4+\ldots$$

$$\Rightarrow S-\rho S=\left(\rho+2\rho^2+3\rho^3+\ldots\right)-\left(\rho^2+2\rho^3+3\rho^4+\ldots\right)$$

$$\Rightarrow S(1-\rho)=\rho+\rho^2+\rho^3+\ldots$$

$$\Rightarrow S(1-\rho)=\frac{\rho}{1-\rho}$$

$$\Rightarrow S=\frac{\rho}{(1-\rho)^2}$$

Thus from (*), we have $L_s=(1-\rho)S=\frac{(1-\rho)\rho}{(1-\rho)^2}=\frac{\rho}{1-\rho}$

Now, ρ is the traffic intensity or utilisation factor, where $\rho = \frac{\lambda}{\mu}$

$$\text{Thus, } L_s = \frac{\rho}{1-\rho} = \frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}} = \frac{\lambda}{\mu-\lambda} \qquad (1)$$

Where, λ denotes the average arrival rate and μ denote the average service rate.

2. The average queue length (Expected number of customers) is L_q

And $L_q = L_s -$ Traffic intensity

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

3. The average time an arrival spends in the system (Expected waiting time in system) is $W_s = \frac{L_s}{\lambda} = \frac{\lambda}{\lambda(\mu-\lambda)} = \frac{1}{\mu-\lambda}$ (from (1))

$$\text{Thus } W_s = \frac{1}{\mu-\lambda} \qquad (3)$$

4. The average waiting time of an arrival in the queue is W_q.

$$\text{And } W_q = \frac{L_q}{\lambda} = \frac{\lambda^2}{\lambda\mu(\mu-\lambda)} = \frac{\lambda}{\mu(\mu-\lambda)} \qquad (4)$$

from (4), we can also obtain the relation

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

5. The probability that the number of customers (units) waiting in the queue and the number of units being serviced is greater than k is

$$P[n > k] = \rho^{k+1} \qquad (5)$$

6. The probability of having a queue i.e. $1 - P_0$

$$\text{Thus } 1 - P_0 = 1 - (1-\rho) = \rho \qquad (6)$$

Now, we explain the M/M/1 queuing model with the help of following examples.

Example.

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

Solution: Here, we have

Arrival rate (λ) = $\frac{10}{8}$ arrivals/hr

Service rate (μ) = $\frac{60}{30}$ services/hr

∴ The probability of the repairman being idle is

$$P_0 = 1 - \rho = 1 - \frac{\lambda}{\mu} = 1 - \frac{10}{8} \times \frac{1}{2} = 1 - \frac{5}{8} = \frac{3}{8}$$

Thus, $P_0 = \frac{3}{8}$.

Hence the idle time in the 8-hour day = $\frac{3}{8} \times 8$ hours = 3 hours

Now, the average number of jobs that are ahead of the set (person) just brought in

= Average number of units in the system means who are in queue and at counter, i.e.

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\frac{10}{8}}{2 - \frac{10}{8}} = \frac{10}{6} = \frac{5}{3} = 1.666$$

Thus, $L_s = 1.67$.

Example.

Customers arrive at a window in a bank, according to a Poisson distribution with mean 10 per hour. Service time per customer is exponential with mean 5 minutes. The space in front of the window including that for the serviced customers can accommodate a maximum of three customers. Other customers can wait outside this space.

(a) What is the probability that an arriving customer can go directly to the space in front of the window?

(b) What is the probability that an arriving customer will have to wait outside the indicated space?

(c) How long is an arriving customer expected to wait before being served?

Solution: Here, arrival rate $\lambda = 10$ per hour and service rate $(\mu) = \frac{60}{5}$ per hour = 12 per hour

Thus, traffic intensity $(\rho) = \frac{\lambda}{\mu} = \frac{10}{\frac{60}{5}} = \frac{5}{6}$

$$\Rightarrow \rho = \frac{5}{6}$$

(a) The probability that an arriving customer can go directly to the space in front of the window i.e. the probability of n customers in the system here $2 \geq n \geq 0$, therefore the required probability $= P_0 + P_1 + P_2$

$$= \left(1-\frac{5}{6}\right) + \left(1-\frac{5}{6}\right)\frac{5}{6} + \left(1-\frac{5}{6}\right)\left(\frac{5}{6}\right)^2$$

$$= \frac{1}{6} + \frac{1}{6}\cdot\frac{5}{6} + \frac{1}{6}\cdot\frac{25}{36} \approx 0.42$$

(b) The probability that an arriving customer will have to wait outside the indicated space, i.e. the required probability $= P_3 + P_4 + P_5 + \ldots$

$$= 1 - (P_0 + P_1 + P_2)$$

$$= 1 - 0.42 = 0.58$$

(c) Expected waiting time before being served

$$= W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{10}{12(12-10)} = 0.417 \text{ hours}$$

Exercises

1. Customers arrive at a box office window being served by a single individual according to a Poisson input process with a mean rate of 30 per hour. The time required to serve a customer has an exponential distribution with a mean of 90 seconds. Find the average waiting time of a customer in the queue.

Solution. $W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{30}{40(40-30)} = \frac{3}{10} hr$

$= \frac{3}{4} \times 60 \text{ minutes} = 4.5 \text{ minutes}$

2. A fertilizer company distributes its products by trucks loaded at its only loading station. Both company trucks and contractor's truck are used for this purpose. It was noticed that on an average on truck arrived every 5 minutes and the average loading time was 3 minutes. Forty

percent of the trucks belong to the contractor. Determine the expected waiting time of contractor's trucks per day.

Solution.

The expected waiting time of contractor's truck per day=8.64 hrs.

3. In a railway marshalling yard, goods train arrives at a rate of 36 trains per day. Assuming that inter-arrival time follows exponential distribution and the service time distribution is also exponential with an average of 30 minutes, calculate the following:

- The mean line length
- The probability that the queue size exceeds 10.
- If the input increases to an average of 42 per day, what will the change in (a) and (b) be?

Solution.

(i) Mean length of a system

Ls = 3 trains 1

(ii) Probability that the queue size exceeds 10= probability that the system size exceeds 11

i.e., $P(n>11) = (0.75)^{12}$

c) $L_s = \lambda' / \mu - \lambda' = 42/48-42 = 7$ trains

And the probability that the queue size exceeds 10 is $(7/8)^{12}$

CHECK YOUR PROGRESS

1. What is Queueing Theory?

2. Explain what do you mean by Poisson process.

3. Describe birth and death process.

4. Explain briefly fundamental structure of Queueing System.

5. Write a note on operating characteristics of Queueing System.

6. Explain classification of Queueing System.

7. Derive the arrival-departure equation for M/M/1 Queueing Model.

8. State operating characteristics of M/M/1 Queueing Model.

SUMMARY

In this chapter, we have introduced the concept of queueing theory. It has been observed that for understanding queueing system, we have to familiar with some concepts namely Stochastic Process, Poisson Process, and Birth-Death Process along with some distributions such as Poisson distribution and exponential distribution. We explained the basic component, fundamental structure and operating characteristics of a queueing system. Also, we discussed the M/M/1, queueing models.