UNIT 2
QUEUING THEORY

LESSON 23

Learning Objective:

- **Identify the classification of queuing models.**

- **Apply formulae to find solution that will predict the behaviour of the single server model I.**

Hello students,
In this lesson you are going to learn the models of queuing theory.

CLASSIFICATION OF QUEUING MODELS

There is a standard notation system to classify queuing systems as A/B/C/D/E, where:

- A represents the probability distribution for the arrival process
- B represents the probability distribution for the service process
- C represents the number of channels (servers)
- D represents the maximum number of customers allowed in the queueing system (either being served or waiting for service)
- E represents the queue discipline or service mechanism

Common options for A and B are:

- M for a Poisson arrival distribution (exponential interarrival distribution) or a exponential service time distribution
- D for a deterministic or constant value
- G for a general distribution (but with a known mean and variance)

If D is not specified then it is assumed that it is infinite.

For example the M/M/1 queueing system, the simplest queueing system, has a Poisson arrival distribution, an exponential service time distribution and a single channel (one server).

Single server model

(M/ M/1) QUEUING MODEL

The M/M/1 queuing model is a queuing model where the arrivals follow a Poisson process, service times are exponentially distributed and there is one server.

The **assumption of M/M/1 queuing model** are as follows :

1. The number of customer arriving in a time interval t follows a poison process with parameter λ.
2. The interval between any two successive arrival is exponentially distributed with parameters λ.
3. The time taken to complete a single service is exponentially distributed with parameter μ.
4. The number of server is one.
5. Although not explicitly stated both the population and the queue size can be infinity.
6. The order of service is assumed to be FCFS.

If $\frac{\lambda}{\mu} < 1$ the steady state probabilities exist and P_n the number of customers in the system follows a geometric distribution with parameter $\frac{\lambda}{\mu}$ (**also known as traffic intensity).**

The probabilities are :

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

$$= (\lambda/\mu)^n (1 - \lambda/\mu) ; \qquad n = 1, 2, \ldots\ldots$$

$$P_o = 1 - \lambda/\mu$$

The time spent by a customer in the system taking into account both waiting and service time is an exponential distribution with parameter $\mu - \lambda$.

The probability distribution of the waiting time before device can also be served in an identical manner.
The expected number of customers in the system is given by --

$$Ls = \sum_{n=1}^{\infty} nP_n = \sum_{n=1}^{\infty} n \,(1 - \lambda/\mu)\,(\lambda/\mu)^n$$

The expected number of customers in the queue is given by—

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

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

Average waiting time of a customer in the system

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

Average waiting time of a customer in the queue

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

LIMITATIONS OF SINGLE SERVER QUEUING MODEL

The single server queueing model is the most simple model which is based on the above mentioned assumptions. However, in reality, there are several limitations of this model in its applications. One obvious limitation is the possibility that the waiting space may in fact be limited. Another 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 periods and slack periods during which the arrival rate is higher and lower respectively than the overall average. These could occur at particular times during a day or a week or particular weeks during a year. There is not a great deal one can do to account for stationary without complicating the mathematics enormously. The population of customers served may be finite, the queue discipline may not be first come first serve. In general, the validity of these models depends on stringent assumptions that are often unrealistic in practice.

Even when the model assumptions are realistic, there is another limitation of queuing theory that is often overlooked. Queuing models give steady state saluting, that is, the models tell us what will happen after queuing system has 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 queuing system never reaches a steady state, so the model solution is of little value.

APPLICABILITY OF QUEUING MODEL TO INVENTORY PROBLEMS

Queues are common feature in inventory problems. We are confronted with queue-like situations in stores for spare parts in which machines wait for components and spare parts in service station. We can also look at the flow of materials as inventory queues in which demands wait in lines, conversely materials also wait in queues for demands to be served. If there is a waiting line of demands, inventory state tends to be higher than necessary. Also, if there is a negative state of inventories, then demands form a queue and remain unfulfilled.

Thus, the management is faced with the problem of choosing a combination of controllable quantities that minimize losses resulting from the delay of some units in the queue and the occasional waste of service capacity in idleness. An increase in the potential service capacity will reduce the intensity of congestion. But at the same time, it will also increase the expense due to idle facilities in periods of “No demand’. Therefore, the ultimate goal is to achieve an economic balance between the cost of service and the cost associated with the waiting of that service. Queuing theory contributes vital information required for such a decision by predicting various characteristics of the waiting line, such as average queue length. Based on probability theory, it attempts to minimize the extent and duration of queue with minimum of investment in inventory and service facilities. Further, it gives the estimated average time and intervals under sampling methods, and helps in decision of optimum capacity so that the cost of investment is minimum keeping the amount of queue tolerance limits.

Single server model I

{(M/ M/1) : (∞ / FCFS)}-- Exponential service; Unlimited queue

Practical formulae involved in single server model I

• Arrival rate per hour	$= \lambda$
• Service rate per hour	$= \mu$
• Average utilization rate (or utilization factor), ρ	$= \lambda / \mu$
• Average waiting time in the system, (waiting and servicing time) W_s	$= 1 / (\mu-\lambda)$
• Average waiting time in the queue, Wq	$= \dfrac{\lambda}{\mu (\mu -\lambda)}$
• Average number of customers (including the one who is being served) in the system, L_s	$= \dfrac{\lambda}{(\mu - \lambda)}$
• Average number of customers(excluding the one who is being served) in the queue L_q	$= \dfrac{\lambda^2}{\mu(\mu-\lambda)}$
• Average number of customers in non-empty queue that forms time to time	$= \mu / (\mu-\lambda)$
• Probability of no customer in the system, or, system is idle or idle men in the factor P_o	$= 1 -(\lambda / \mu)$ or $1 - \rho$ = 1- utilization factor
• Probability of no customer in queue and a customer is being served P1	$= (1 - \lambda / \mu) (\lambda / \mu)$
• Probability of having 'n' customers in the system	$=(1 - \lambda / \mu) (\lambda / \mu)^n$
• Probability of having 'n' customers in the queue	$= (1 - \lambda / \mu) (\lambda / \mu)^{n+1}$

• Probability of having more than 'n' customers in the system	$= (\lambda / \mu)^{n+1}$
• Probability of having less than n customers in the system or probability that an arrival will not have to wait outside the indicated space	$= 1 — (\lambda / \mu)^{n}$
• Probability of having n or more customers in the system or , probability that an arrival will have to wait outside the indicated space	$=(\lambda / \mu)^{n}$
• Probability that a customer will wait for more than 't' hours in the queue	$=\rho \times e^{-t/ws}$ $= \frac{\lambda}{\mu} \times e^{-t(\mu - \lambda)}$
• Total costs associated with system	= Average no. of customers in system x opportunity cost of customer + cost of serving department.
• Total costs associated with queue	=Average no. of customers in queue x opportunity cost of customer + cost of serving department.

Example
The Toolroom problem

The J.C. Nickel Company toolroom is staffed by one clerk who can serve 12 production employees, on the average, each hour. The production employees arrive at the toolroom every six minutes, on the average. Find the measures of performance.

Toolroom Problem Solution

- Assume this is an M/M/1 queue
- Arrival rate is one employee (emp) every six minutes

  – Arrival times between employees follow a negative exponential distribution with mean $1/\lambda = 6$ (minutes/emp)

  –Thus, $\lambda = 1/6$ (emp/minute) = 10 (emp/hour)

- Mean service rate is 12 employees per hour

  –Thus $\mu = 12$ (emp/hour)

It is necessary first to change the time dimensions of λ and μ to a common denominator. λ is given in minutes, μ in hours. We will use hours as the common denominator.

1. Average waiting time in the system (toolroom)

$$W = \frac{1}{\mu - \lambda} = \frac{1}{12 - 10} = 0.5$$

hours, per employee

2. The average waiting time in the line

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

hours, per employee

3. The average number of employees in the system (toolroom area)

$$L = \frac{\lambda}{\mu - \lambda} = \frac{10}{12 - 10} = 5$$

employees

4. The average number of employees in the line.

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{100}{12(12-10)} = 4.17$$

employees

5. The probability that the toolroom clerk will be idle

$$P_0 = 1 - \frac{\lambda}{\mu} = 1 - \frac{10}{12} = 0.167$$

6. The probability of finding the system busy

$$\rho = \left(\frac{\lambda}{\mu}\right) = \frac{10}{12} = 0.833$$

7. The chance of waiting longer than 1/2 hour in the system. That is T = 1/2

$$P(t > T) = e^{(10-12)(1/2)} = \frac{1}{e} = 0.368$$

8. The probability of finding four employees in the system, n = 4

$$P_{n=4} = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) = \left(\frac{10}{12}\right)^4 \left(\frac{2}{12}\right) = 0.0804$$

9. The probability of finding more than three employees in the system.

$$P_{n>3} = \left(\frac{\lambda}{\mu}\right)^{N+1} = \left(\frac{10}{12}\right)^4 = 0.488$$

Case of sales counter

Customers arrive at a sales counter manned by a single person according to a poisson process with a mean rate of 20 per hour. The time required to serve a customer has an exponential distribution with a mean of 100 seconds. Find the average waiting time of a customer.

Solution

Arrival rate = λ =20 customers per hour

Service rate = μ = $\frac{3600}{100}$ =36 customers per hour

The average waiting time of a customer in the queue

$$= \frac{\lambda}{\mu(\mu - \lambda)} = \frac{20}{36(36-20)}$$

$$= \frac{5}{36 \times 4} \text{ hours} = \frac{5}{36 \times 4} \times 60 \times 60 = 125 \text{ seconds}$$

The average waiting time of a customer in the system

$$= \frac{1}{(\mu-\lambda)} = \frac{1}{36-20)}$$

$$= \frac{1}{16} \text{ hours} = \frac{1}{16} \times 60 \times 60 \times 60 = 225\text{seconds}$$

Case of cafetaria

Self service at a university cafeteria, at an average rate of 7 minutes per customer, is slower than attendant service, which has a rate of 6 minutes per student. The manager of the cafeteria wishes to calculate the average time each student spends waiting for service. Assume that customers arrive randomly at each time, at the rate of 5 per hour. Calculate the appropriating statistics for this cafeteria.

Solution

		Self service line	Attended line
A	Arrival rate (λ)	5	5
B	Service rate (μ)	8.571	10
C	Expected number of Students in cafeteria $\frac{\lambda}{\mu-\lambda}$	$\frac{5}{8.571-5}$ =1.40	$\frac{5}{10-5}$ =1
D	Expected number of students waiting for service $= \frac{\lambda^2}{\mu(\mu-\lambda)}$	$\frac{25}{8.57\,(8.571-5)}$ =0.82	$\frac{25}{10(10-5)}$ =0.5
E	Average time in system $\frac{1}{\mu-\lambda}$	$\frac{1}{(8.571-5)}$ = .28 hour or 16.8 min.	$\frac{1}{10-5}$ = .20 hour or 12 min.
F	Average time in queue $\frac{1}{\mu-\lambda} \times \frac{\lambda}{\mu}$	$\frac{1}{8.571-5} \times \frac{5}{8.571}$ = 0.1633 hour or 9.8 min	$\frac{1}{10-5} \times \frac{5}{10}$ = .10 hour or 6 min.

Case of airport

The mean rate of arrival of planes at an airport during the peak period is 20 per hour. And the actual number of arrivals in any hour follows a poisson distribution. The airport can land 60 planes per hour on an average in good weather and 30 planes per hour in bad weather, but the actual number landed in any hour follows a poisson distribution with these respective averages. When there is congestion, the planes that arrived earlier,

(i) How many planes would be flying over the field in the stack on an average in good weather and in bad weather ?

(ii) How long a plane would be in the stack and in the process of landing in good and in bad weather ?

(iii) How long a plane would be in the process of landing in good and bad weather after stack awaiting ?

Solution

		Good weather	Bad weather
A	Arrival rate (λ)	20	20
B	Service rate (μ)	60	30
C	Average number of the planes in queue $\frac{\lambda^2}{\mu} \times \frac{1}{\mu - \lambda}$	$\frac{400}{60} \times \frac{1}{40} = \frac{1}{6}$	$\frac{400}{30} \times \frac{1}{10} = \frac{4}{3}$
D	Average waiting time in the system $\frac{1}{\mu - \lambda}$	$\frac{1}{60-20} = \frac{1}{40}$ hour = 1.5 min	$\frac{1}{(30-20)} = \frac{1}{10}$ hour or 6 min

E	Average service time (i) Average waiting time in system (ii) Average waiting time in queue $\frac{\lambda}{\mu} \times \frac{1}{\mu - \lambda}$	1.5 minutes 0.5 minutes	6 minutes 4minutes

Now Practice some problems yourself,

Unsolved Queuing Problems

Q-1 Mike Moore is a small engine repairman. Engines arrive for repair according to a Poisson distribution at an average rate of 4 per day. He services the engines according to an exponential distribution averaging 4.6 repairs per day. Determine

(a) the average time an engine is out of service for repair,
(b) the average number engines waiting for repair, and
(c) the percent of time Moore is busy making repairs.

Q-2 To support National Heart Week, the Heart Association plans to install a free blood pressure testing booth in El Con Mall for the week. Previous experience indicates that, on the average, 10 persons per hour request a test. Assume arrivals are Poisson from an infinite population. Blood pressure measurements can be made at a constant time of five minutes each. Determine

(a) what the average number of persons in line will be,
(b) the average number of persons in the system,
(c) the average amount of time a person can expect to spend in line,
(d) on average, how much time will it take to measure a person's blood pressure, including waiting time.

Additionally, on weekends, the arrival rate can be expected to increase to nearly 12 per hour.

(e) What effect will this have on the number in the waiting line?

Q-3 Trucks enter an inspection station at the rate of one every four minutes. Inspectors can inspect about 18 trucks per hour. Assume Poisson arrivals and exponential service times. Determine

(a) how many trucks would be in the system,
(b) how long it would take for a truck to get through the inspection station,
(c) the utilization of the person staffing the station,
(d) the probability that there are more than three trucks in the system.

Q-4 At a toll station only one tollbooth was open and cars were arriving at the rate of 750 per hour. The toll collector took an average of four seconds to collect the fee. Determine

(a) the percent of time the operator was idle,
(b) how much time you would expect it to take to arrive, pay your toll and move on,
(c) how many cars would be in the system,
(d) the probability that there would be more than four cars in the system.
If during a holiday weekend the arrival rate increased to 1200 per hour and a second tollbooth were opened with a toll collector of equal capability;
(e) how many cars would you expect to see in the system?

So, now let us summarize today's discussion:

Summary

We have discussed in details about

- Classification of queuing models
- Limitations of single server queueing model
- Applicability of queueing model to inventory problems
- Practical formulae involved in single server model I
- Evaluation of performance measures of single server model I

Slide 1

Kendall-Lee Notation

- Kendall (1951) defined the following notation to describe a queuing system.
- There are 6 characteristics
  - 1/2/3/4/5/6
- The first specifies the arrivals process
  - M = inter-arrivals are i.i.d exponential.
  - D = inter-arrivals are i.i.d and deterministic.
  - E_k = inter-arrivals are i.i.d Erlang(k).
  - GI = inter-arrivals are i.i.d and have some general distribution.
- The second specifies the service process
  - The notation is the same as for arrivals.
- The third specifies the number of servers.

Slide 2

Kendall-Lee Notation

- The fourth specifies the service order
  - FCFS = First Come First Served
  - LCFS = Last Come First Served
  - SIRO = Service in Random Order
  - GD = General Queuing Discipline
- The fifth specifies the maximum allowable number of customers in the system.
- The sixth specifies the size of the population from which arrivals are drawn.
- The first queue we will examine is the $M/M/1/FCFS/\infty/\infty$.
- What does this mean?
- This notation is often shortened to $M/M/1$.

Slide 3

M/M/1 queuing model Summary

- Distribution of number in system
- Expected number in system
- Expected number in queue
- Expected time in system
- Expected time in queue
- Probabilities concerning time in queue

$$\rho = \frac{\lambda}{\mu}$$

$$\pi_n = P(N = n) = \rho(1-\rho)$$

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

$$L = \lambda W = \frac{\lambda}{\mu - \lambda}$$

$$W_q = W - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)}$$

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

Slide 4

M/M/1 queuing model
Needed for steady state

- The arrival rate must be less than the service rate
  - Otherwise, the queue would eventually grow without bound

Slide 5

Example

- An average of 10 cars arrive each hour to a single server drive in teller.
- The average service time is 4 minutes.
  - What is the probability that the teller is idle?
  - What is the average number of cars waiting in the line for the teller?
  - What is the average amount of time a customers spends waiting to be served?
  - On average, how many customers will be served in an hour?

Slide 6

Example

- Car owners fill up their tanks when they are empty.
- Suppose a gas station with a single pump has 7.5 customers per hour on average.
- On average, a customers takes 4 minutes to complete service.
  - What are the average queue length and waiting time?
- During a gas shortage, customers fill up their tanks when they are half full.
  - What are the average queue length and waiting time now?