BASIC CHARACTERIZATION OF QUEUEING THEORY

Mrs.R.SOWMYA

Assistant Professor, Department of Mathematics,
Dr.G.R.Damodaran College of Science, Coimbatore

ABSTRACT

In this paper, the concept of queueing theory and mathematical analysis of several related processes, including arrival at the queue, waiting in the queue and being served by the server(s) are analysed.

Key words: *Queueing system, single vacation, Bernoulli schedule*

INTRODUCTION

Whenever there is competition for limited resources, queueing is likely to occur. In order to reduce its inconvenience to bearable levels queueing theories has increasingly occupied the attention of the researchers. Queueing theory is a branch of applied probability theory. This theory enables mathematical analysis of several related processes, including arrival at the queue, waiting in the queue and being served by the server(s). The theory also permits the derivation and calculation of several performance measures including the average waiting time in the queue or in the system, the expected number of customers waiting or receiving service. Queueing theory has a wide range of its applications.

1.1 Description of the queueing problem

Queueing theory is a study or analysis of system of service where a customer arriving for service has to wait in queue, receive service from one or more servers and then leave. The three main concepts in queueing theory are customers, queues and servers (service mechanisms). The meaning of these terms is reasonably self- evident. In general in a queueing system, customers of the queueing system are considered as an input source. The customers arrival and their inter arrival times are considered to follow certain statistical distribution. The customers join a queue at various times and are served by the server (service mechanism). The basis of which the customers are selected is called the queue discipline.

Basic Characteristics

A queueing system can be characterized by the following factor as

i) The input process
ii) The pattern of service
iii) Design of service facility
iv) The queue discipline
v) The capacity of the system.

i) The input process: The input pattern means the manner in which the arrivals occur. It also indicates whether the arrival occurs singly or in groups or batches (bulk).

If the customer decides not to enter the queue because of its huge length he is said to have **balked**. On the other hand, a customer may enter the queue, but after some time loses patience and decides to leave. In this case he is said to have **reneged**. Sometimes customers may move from one queue to another for his personal economic gains, is known as **jockey** for position.

ii) The pattern of service: By the pattern of service, we mean the manner in which the service is rendered. It is specified by the time taken to complete a service. The time may be constant (deterministic) or it may be stochastic.

iii) Design of service facility: A system may have a single server or more number of servers, when all servers offer the same service then the facility is said to have parallel servers. In this case arriving customers may choose at random any one of them for receiving service.

On the other hand, the facility may comprise a number of services stations through which the customer must pass before service is completed (for example processing of a product on a sequence of machines). The resulting situation is known as queue in series or **tandem queues**.

The most general design of a service facility includes both series and parallel processing station. This is called network queues.

iii) The queue discipline: The queue discipline is the method by which customers are selected from the queue for processing by the service mechanisms (also called servers). The queue discipline is normally first come first served (FCFS), where the customers are processed in the order in which they arrived in the queue, such that the head of the queue is always processed next. Most queueing models assume FCFS, as the queue discipline or first in first out (FIFO) some queueing models also assume last come first served (LCFS) and service in random order (SIRO). It is also possible that customer arriving at a facility may be put in priority queues such that those with a higher priority will receive preference to start service first. The specific selection of customers from each priority queue may, however, follow any queue discipline.

iv) The Capacity of the system: When the waiting line reaches a certain length, no further customers are allowed to enter until space becomes available. Such types of situation are referred to as finite source queues. A system may have an infinite capacity in that case the queue may grow to any length. With the help of characteristics discussed, we can study the nature of queueing models.

Kendall's notations

There is a standard notation for classifying queueing systems into different types. This was proposed by D.G. Kendall. According to the notation, system are described by

where

A	-	Distribution of inter arrival times of customers
B	-	Distribution of service times
C	-	Number of servers
D	-	Maximum number of customers which can be accommodated in System.
E	-	Calling Population size

Further,

A and B can take any of following distribution types:

M	-	Exponential distribution (Markovain)
D	-	Degenerate (or Deterministic) Distribution
E_k	-	Erlang Distribution (K = Shape Parameter)
G	-	General Distribution (arbitrary distribution)

1.2 Variations on queueing models

This section explains some variations on queueing models.

Server's vacation

There are queueing models in which the server on completion of service to the existing customers continues to stay in the empty system awaiting for new arrivals. Kella (2)and Lee and Srinivasan (3)first provided detail discussions concerning N-policy $M|G|1$ and $M^{[X]}|G|1$ queueuing system with vacations. Lee et al (4) analysed in detail the batch arrival $M^{[X]}|G|1$ queueuing system under N-policy with a single vacation and repeated vacations. Their results significantly confirmed the stochastic the decomposition property given by Furhmann and Cooper (1). From practical considerations, it may not always be worthwhile to keep servers unnecessarily idle. In such situations, the server may utilize his idle time in a useful and optimal way to perform additional jobs or preventive maintenance work and it is termed as **"Servers vacation".**

Single vacation and multiple vacation

After completing service, if the server finds an empty queue, he leaves the system for a vacation. On returning from vacation, if the server finds that the number of customers waiting in the queue is not to the required level for service, he may decide to stay back in the system itself waiting for the customers to start the next service. The server will take next vacation only after performing atleast one service, this type of server's vacation is called **"Single Vacation".** In some cases, after returning from a vacation, if the server does not find sufficient customers, then the server leaves for another vacation and repeat his vacations until he finds the sufficient number of customers. This type of vacation is called **"Multiple or repeated vacation".**

Set up time:

The server may not immediately serve the waiting customers rather he spends a random amount of time called **"setup time"** to perform certain pre service work or auxiliary task.

N-Policy:

Yadhini and Naor (4) were first to introduce the concept of N-policy. In M/M/1 queueing system for reliable servers so called N-policy means that the server is turned on only when N ($N \geq 1$) or more units are present in the system and turned off when the system is empty.

Bernoulli Vacation:

The Second stage service (SSS) of a unit is completed, the server may go for a vacation of random length V with probability $P(0 \leq P \leq 1)$ or may continue to serve the next unit if any, with probability $(1-P)$ otherwise, it remains in the system itself. It is also assumed that the vacation time V of the server follows a general probability law with distribution function $V(t)$, Laplace stieltjes transform $V^*(\theta)$ and the finite K^{th} moments $E(V^k)$, K=1, 2 are independent of the service time S_i and the arrival process. Further it is also assumed that, if after returning from a vacation, the server does not find any unit in the system, then he joins the system without taking any further vacations and this policy is termed as **Single Vacation (SV) Policy with Bernoulli Schedule (BS).**

Bulk Queues:

It refers to a situation where service can be effected in a batch of upto C customers (i.e.) All waiting customers upto a fixed capacity C are taken for service in a batch. Bailey (1) introduced the concept of bulk queues in the year 1954.

Markovian Queueing Model:

Queueing models in which the inter arrival time and the service time follow the exponential distribution which satisfies the Markovian property [The future behavior depends only on the present and not of the past] are called Markovian queueing models.

Non-Markovian queueing model:

Queueing models in which either inter arrival time or service time or both do not follow the exponential distribution are called Non-Markovian models.

Transient and steady state queueing system:

A queueing system is said to be in transient state, where its operating characteristic (like input, output, mean queue length etc) are dependent upon time.

If the characteristic of the queueing system becomes independent time then the queueing system is said to be in steady state conditions.

Steady State conditions

In order to study the steady state results, it is important to know the conditions for the existence of the limit. If λ denotes the average rate of customers arriving the queueing system and the rate of servicing the customers is μ then the traffic congestions for c server systems is $\lambda\rho = C\mu.$ The average service rate of the system, we would expect. As time goes on, the queue becomes bigger and bigger, at some point customers were not allowed to join. Suppose $\rho > 1$ then the queue size never settles down, and there is no steady state. Hence in turn we find that for steady state results to exist. ρ **must be strictly less than1.**

Probability Generating Function:

Suppose that X is a random variable that assumes non- negative integral values 0,1,2,., and $\Pr[X = K] = g_k$, $K = 0,1,2,......$ such that $\sum_{k=0}^{\infty} g_k = 1$. Then the corresponding generating function P(z) = $\sum_{k=1}^{\infty} g_k z^k$ of the sequence of probability $\{g_k\}$ is known as the probability generating function of the random variable X. It is some time called as geometric transform of the random variable X The normalizing condition is P(1)=1, where the series P(z) converges and is finitely differentiable. Also the probability generating function (PGF) determines a distribution uniquely.

Mean and variance in terms of (derivatives of) PGF:

Let X be as random variable whose probability generating function is given by $P(z) = \sum_{n=0}^{\infty} P_r(X = n) z^n$ then the mean of the random variable E(x) is given by

$$E(X) = \sum_{n=0}^{\infty} nP_n = \frac{d}{dz}(P(z))_{z=1}$$

$$E(X(X-1)) = \sum_{n=1}^{\infty} n(n-1)P_n = \frac{d^2}{dz^2}\left(P(z)\right)_{z=1}$$

Laplace Transforms

It is a generalization of generating function. Laplace transforms serve as very powerful tools in many situations. It provides an effective means for the solution of many problems arising in our study. For example, the transforms are very effective for solving linear differential equations. The Laplace transformation reduces a linear differential equation to an algebraic equation. In the study of some probability distributions, this method could be used with great advantage, that it is easier to find the Laplace transform of a probability distribution rather than the distribution itself.

Definition

Let f(t) be a function of a positive real variable 't'. Then the Laplace transform of f(t) is defined by $L\left(f(t)\right) = \int_0^{\infty} \exp(-st) f(t) dt$ for the range of value of s for which the integral exists.

Laplace steiltjes transforms of a probability distribution or of a Random variable.

Let X be a non-negative random variable with distribution function $F\left(x\right) = \Pr\left\{X \le x\right\}$

The Laplace stieltjes transform (LST) $F^*(s)$ of this distribution is given by

$$F^*(s) = \int_0^{\infty} \exp(-SX) df(X) \qquad \text{for } S \ge 0$$

Thus we have $F^*(s) = E\left\{\exp(-SX)\right\}$ and $F^*(0) = 1$

Suppose that X is a continuous variable with density $f(X) = F'(X)$ then its LST is given by

$$F^*(s) = \int_0^{\infty} \exp(-SX) f(X) dX$$

$$F^*(s) = L\left\{f(X)\right\} \qquad \text{(by definition)}$$

Suppose that X is an integral – valued random variable with distribution $g_k = \Pr[X = K], K = 0,1,2,......$ and its probability function is given by $P(s) = \sum g_k z^k$

Then we have $F^*(s) = E[\exp(-sx)] = \Pr\ [\exp(-s)]$

Hence for discrete random variable assuming non-negative values 0,1,2…. the Laplace transform of the variables differs from its PGF only by a change of variable.

Conclusion

In this paper, we studied about the concept of characterization of queueing theory in operations research.

References

[1]Furhmann, S.W and Cooper, R.B (1981), "A note on the $M\left|G\right|1$ queue with server vacations", Questa 31, 13-68.

[2]. Kella, O, (1989), "The threshold policy in the $M\left|G\right|1$ queue with server vacation", Naval Res. Logist 36, Vol 1, 111-123.

[3]. Lee et al (1993), "Operating characteristics of $M^{[X]}|G|1$ queue with N – policy", QFSTA.

[4]. Lee, H.S and Srinivasan, M.M (1989), "Control policies for the $M^{[X]}|G|1$ queueing system", Management Science, 708 -721.

[5]. Yadhini, M and Naor, P (1963), "Queueing system with a removable service station", Operations Research, 14, 393-405.