**JOURNAL OF ALGEBRAIC STATISTICS**
Volume 13, No. 2, 2022, p.3782-3788
https://publishoa.com
ISSN: 1309-3452

# Critical Study of Queuing Theory

**Bratati Ghatak, Dr. Annapurna Ramakrishna Sinde**
Department of Mathematics, Dr. A.P.J. Abdul Kalam University, Indore (M.P.), India

**Received** 2022 March 15; **Revised** 2022 April 20; **Accepted** 2022 May 10.

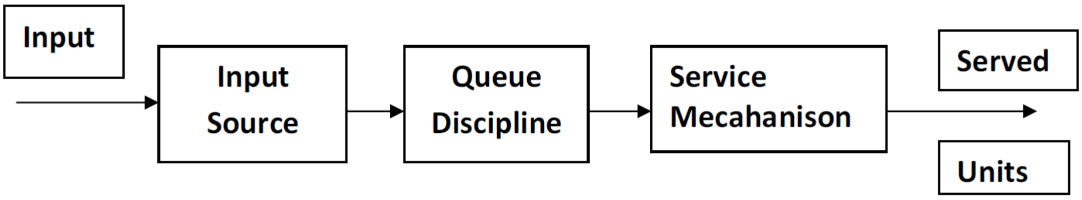
**Abstract:**
A large number of new techniques are being devised to expand the scope of scientific knowledge to unrestricted boundaries of its applications. Such techniques have brought an effective change and can be estimated as controlling services in different fields of life and for this; we set up and use mathematical models. The medical organizations can be considered as queuing systems because in these, the patients arrive then wait for the service and after obtaining service, they depart. The patient must go through the process of treatment which consists of a set of actions and procedures with the purpose of receiving the required treatment. This paper reflects critical study of Queuing Theory.

**Keywords:** queuing theory, technological, difficult, number, important

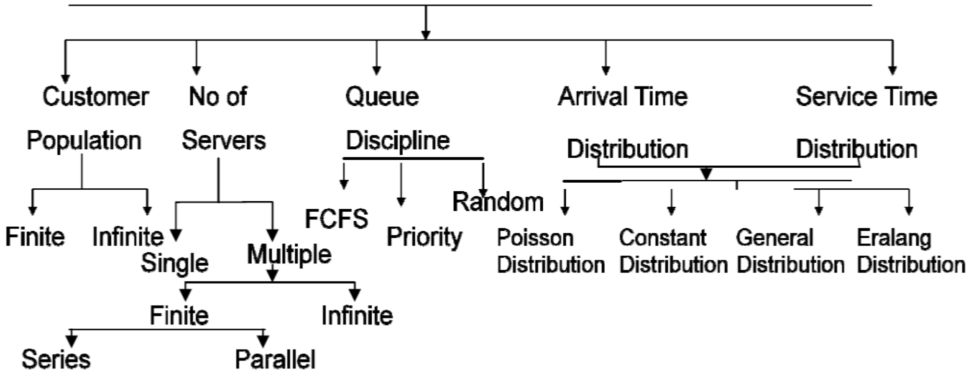
## I Introduction:

Queues are a part of everyday life. We all wait in queues to buy a movie ticket, to make bank deposit, pay for groceries, mail a package, obtain a food in a cafeteria, to have ride in an amusement park and have become adjustment to wait but still get annoyed by unusually long waits.

The Queuing models are very helpful for determining how to operate a queuing system in the most effective way if too much service capacity to operate the system involves excessive costs. The models enable finding an appropriate balance between the cost of service and the amount of waiting.



Information necessary to solve the queuing problem are shown below in diagram form:



It's not novel to wonder how to design and operate a successful queuing system. For decades, it has been the focus of both academic and applied research, and for good reason: Any service-based organization's ability to effectively manage and construct queues will have a significant impact on how well it can serve customers. While ineffective queue management wastes time and aggravates customers, an effective queueing system allows customers to leave with a favourable experience.

The traditional solutions to this issue included physical check-in at the counter, sitting in a waiting area, receiving a ticket, hanging velvet cords around a space to contain line-ups, and other methods. A queue management system is the answer for tech-savvy firms today. A virtual queue management system is a potent solution to queuing issues in the digital age for many reasons, even though the problems it solves may not be novel. A queuing system has many benefits in addition to controlling client flow. Discover the advantages of efficient queue management for your clients, staff, and bottom line.

A robust employee dashboard provided by a queue management system will do away with the uncertainty and confusion that uncontrolled line-ups frequently bring. Employees should be able to quickly determine the status of the line thanks to the system's provision of all the most crucial line metrics. This is particularly crucial when there are several lines active at once. Your staff members must be able to see who is arriving next, when they anticipate receiving service, and what they require. The goal is to maximise the effectiveness and efficiency of every client engagement.

Employees will wait less for their next set of activities if they are well-informed on what has to be done to prepare for a certain customer. The customer gains as a result of your staff being prepared to assist them with all they require for the meeting rather than wasting time on routine research. Ultimately, this results in less workload fluctuation and more effective workers.

**Characteristics of the queuing system:**

(a) Input source
(b) Queue discipline
(c) Service mechanism

The modern technological progress is moved with the growth of scientific techniques, although existing methods have been faced the problems arising from growth of people and society. A large number of new techniques are being devised to expand the scope of scientific knowledge to unrestricted boundaries of its applications. Such techniques have brought an effective change and can be estimated as controlling services in different fields of life and for this; we set up and use mathematical models. Any model of a real-life situation must make simpler it greatly by giving preference to those factors which we think important for our purpose, and neglecting the rest. Otherwise, it becomes difficult to calculate, or predict the problem. We must think about whether we have made an accurate selection for it. To place our model in mathematical terms we have to formulate our ideas. Once we have a model in mathematical terms, we can organize it effectively using summarizing mathematical language. Also, calculations are more enthusiastically set up using a computer.

Unfortunately, emergencies are part of our lives. It is critical to efficiently plan and assign emergency response services and deliver effectively them to people most in need for timely relief. In a hospital emergency room, there are various patients who may be associated with the same medical incident.

The medical organizations can be considered as queuing systems because in these, the patients arrive then wait for the service and after obtaining service, they depart. The patient must go through the process of treatment which consists of a set of actions and procedures with the purpose of receiving the required treatment. The resources or servers in these queuing systems are the qualified personals and specific equipment to complete the actions and procedures required.

With the aim of characterizing the workflow in a hospital, it is essential to recognize how the various departments of the hospital work and how records flow.

## II Queuing Theory

Queuing theory is the oldest and finest developed analytical techniques which are used in daily waiting lines. It was born in the early 1900’s; when A.K.Erlang first addressed the questions raised by telephone engineers who were trying to understand the effects of the randomness of telephone traffic. Queuing theory is motivating because it is based on an easy model of a straightforward reality i.e., customer arrives at random, waits if necessary for the server until he becomes vacant then hold the server for an indiscriminating duration of time and after getting service leaves the system. This real situation exposes some amazing behaviours.

It is exciting to remind that only after the World War Π, it was realized that the queuing theory has extensive applications in a variety of fields, such as, industrial problems as manufacturing development and repairs, in road traffic overcrowding, scheduling of an traffic at the airport, waiting in a hospital outpatient department, register systematize the machines coming up to be served by repairmen, substantial processes, epidemic processes in biology, quick transportation systems, call-centres, networks, telecommunications, mainframe computer queuing of telecommunications terminals, advanced telecommunications systems etc.

The queue may be contained the customers physically as in the line in a bank or a supermarket, or it may not be physical as in calls to a telephone call centres on requests for emergency services such as for calling fire service, medical facilities as ambulance or calling police. When the queue physically contains the customers, the customers are often capable to directly calculate approximate waiting time. On the other hand, if the queue does not contain physically the customers, it may be difficult for customers to execute the estimation of the waiting time.

**2.1 Queue-characteristics:**

The queuing system is explained by the following five basic characteristics:

1. The input or arrival pattern.
2. The output or service pattern.
3. Service mechanism and the number of servers.
4. System capacity.
5. Queue discipline.

**Blocking:**

If a queuing system places a limit on length of queue blocking may occur. For example, a patient arrives in an outpatient clinic and finds that its waiting room is full, he may sidetrack.

In a hospital, the limited number of beds may put off a unit from accepting patients where inpatients can wait only in a bed. If the system has blocked then the congestion not only increases patient waiting time but also reduces the throughput of the system.

**Bottlenecks:**

There are several nodes at which point services are dispersed in a queuing system. A patient may have to go through several nodes and thus several queues in order to obtain the essential service. In the case of the appointment systems, the nodes can be expected where the ratio of require to available service capacity is comparatively high to become bottlenecks. Such bottlenecks would have high utilization and increases overall patients waiting time even through other nodes may have low utilization.

**The state of the system:**

The state of the system is one of the fundamental concepts to understand queuing theory; it involves the study of a system’s behaviour over time. It may be classified as follows:

- Transient state
- Steady state:
- Some distributions:

Here are some distributions which are used in queuing theory:

- Exponential distribution
- Regular distribution
- Erlang service time distribution with ‘k’ phases:

- Poisson Process

**Kendall’s notation:**

After developing sufficient expressions to demonstrate many queuing systems, now, we describe a standard notation used in different models in queuing theory which is initially given by D.G. Kendall. Kendall’s notation for specifying a queue’s characteristics is a/b/c/d/e. The first characteristic ‘a’ represents the arrival pattern. The second characteristic ‘b’ specifies the service pattern. The following abbreviations are used for the arrival and service pattern to substitute the entries ‘a’ and ‘b’:

M = Markovian (or Exponential) inter-arrival time or service-time distribution.

D = Deterministic (or constant) inter-arrival time or service-time.

G = General distribution of service time (departure), i.e. no assumption is made about the type of distribution with mean and variance.

Ek = Erlang-k distribution for interarrival or service time with parameter k (i.e. if k=1, Erlang is equivalent to exponential and if k=∞, Erlang is equivalent to deterministic).

GI = General probability distribution – normal, uniform or any empirical distribution, for inter-arrival time.

The third characteristic ‘c’ specifies the number of service stations i.e., the number of servers.

The fourth characteristic ‘d’ describes the maximum number of customers permitted in the system both in queue and in service i.e. the capacity of the system. The fifth characteristic ‘e’ specifies the queue-discipline:

FCFS = First come, first served

LCFS = Last come, first served

SIRO = Service in random order

GD = General queue discipline.

We are living in a quickly changing society where some knowledge of Mathematics is essential. A sound knowledge of Mathematics is not just a discipline to sharpen a person’s mental power but it helps to promote scientific temper that includes the excellent qualities of objectivity, precision and accuracy. That is why the study of queuing-theory helps persons to think better, logically and sequentially.

Queuing theory originated in perspective of military operations but now it is accepted as a strong tool for planning and decision making especially in business and industries. Its approach has provided a new outlook to many conventional problems. In fact, its techniques do represent a scientific methodology of analysing the problems. They provide an improved basis for decisions. The practical benefits obtained from its many applications have attracted increasing attention to study it. With computer facilities, the significance and scope of it has grown and is still growing. Exciting discoveries of broad importance are being made in quick succession. Deep and new ideas of a rapidly growing theory very often shed new and sharp light on the most elementary topics.

## III Basic Features Of Queueing System

Queueing systems are simplified mathematical models to explain congestion. In general, a queueing system exists whenever "customers" request "service" from a facility; often, both the customers' arrival and the times for providing service are supposed to be random. When new clients arrive and all of the "servers" are full, they will typically queue up for the next server that becomes available. The arrival pattern, service mechanism, and queue discipline are the three elements that make up a simple queueing system. From a probabilistic perspective, queue characteristics are typically derived from those of the stochastic processes that are connected to them. However, figuring out the state probabilities in all but the simplest queues is really challenging. However, it is frequently possible to identify their long-term limit, also known as the equilibrium or steady-state distribution. This distribution is stationary and independent of the system's initial conditions. The limitations on the parameters under which the system would finally find equilibrium are provided by the ergodic conditions. Queueing theory mainly focuses on computing steady-state probabilities and using them to calculate other (steady-state) queue performance indicators. Little's law is a very helpful formula for systems in equilibrium when only the expected values are needed. The study of queueing theory's mathematical features and the probabilistic construction of queueing models have received much of the attention. As a result, the parameters guiding the models are mostly taken as givens. Uncertainty-introducing statistical analyses are comparatively uncommon. The creation of the requisite sample distributions can be highly complex, and analysis is frequently limited to asymptotic results, making inference in queueing systems challenging. If the statistical analysis is approached from a Bayesian viewpoint, it is easier to understand. All that is required from the data for Bayesian analyses is a likelihood function, which when combined with the prior distribution on the parameters yields the posterior distribution from which inferences are derived. This is because Bayesian analyses are insensitive to (noninformative) stopping rules. This is a crucial simplification for queue analysis because there are many distinct ways to observe the system, many of which have proportional likelihood functions but have highly different sample distributions. The previous distribution quantifies all information about the system that was already known before to the data collection. Usually, there is a lot of knowledge about the queue a priori, especially if equilibrium is expected. However, by doing a Bayesian analysis—often referred to as "objective" because the prior distribution utilised is of the "non-informative" or "objective" type—it is also feasible to maintain the parallelism with a probability analysis and prevent the insertion of additional subjective inputs. Estimates and standard errors can be calculated immediately from the posterior distribution. Additionally, it is possible to compute probabilities of immediate importance, such as the likelihood that the ergodic condition would hold. Most crucially, the approach easily takes into account constraints on the parameter space imposed by the equilibrium assumption. The system's measurements of congestion (such as the number of clients queuing up, the amount of time spent waiting in line, the number of servers that are busy, and so on) are predicted using the associated predictive distributions, which are also very helpful for system design and intervention.

In designing queueing systems, we need to aim for a balance between service to customers and economic consideration. In terms of analyzing the queueing situations, the types of questions in which typically concerned with measures of system performance and might include, what is the average length of the queue?, what is the probability of a customer having to wait longer than a given time interval before they are being served?, what is the probability that the queue will exceed a certain length? and what is the expected utilization of the server and the expected time period during which he will be fully occupied and so on. These are the questions to be answered so that the management can evaluate alternatives in an attempt to control or improve the situation.

There are two basic approaches in queueing theory such as analytic solutions (formula based) and simulation techniques (computer based). The reason for there being two approaches is that analytic methods are only available for relatively simple queueing systems whereas complex queueing systems are almost analyzed using simulation. The simple queueing systems that can be tackled via queueing theory essentially have distributions for the arrival and service processes that are well defined as standard statistical distributions such as Poisson or Erlang. To analyze the queueing system, we need information related to: arrival process, service mechanisms, queue discipline, system capacity and service channels. This information is very useful for successfully designing the queueing systems that achieve an appropriate balance between the cost of providing a service and the cost associated with waiting time of the customer for that service. The principles of management are to balance between minimizing costs due to occupation of resources and minimizing wait times of customers.

### Arrival Process

The arrival pattern means the manner in which customers arrive and join the system. Arrivals may occur in single or in groups (batch or bulk arrival). It is specified by the probability of time between successive arrivals, that is, the inter arrival time distributions. If the arrival pattern does not change with time, then it is called a stationary arrival pattern, otherwise, it is called non stationary (transition) arrival pattern. The commonly used input distributions are Poisson, Deterministic, Erlangian type with parameter k (k= 1,2,3,...), Hyper exponential and Phase type distribution. A Poisson stream of arrivals corresponds to arrivals at random. In a Poisson stream, successive customers independently arrive after some intervals are exponentially distributed. The Poisson stream is a convenient mathematical model of many real-life queueing systems and is described by a single parameter the average arrival rate. Other important arrival processes are scheduled arrivals; batch arrivals; and time dependent arrivals.

### Balking, Reneging and Jockeying

If the customer decides not to join the queue when it is too long then he is said to have balking. if the customer leaves the queue after waiting too long for service, he is said to have reneging. If the customer switches between queues as he thinks he will get served faster by so doing is known as jockeying.

**Service Mechanism**

Service mechanism is a description of the resources needed for service to begin. It explains the service time distribution, the available number of servers available, whether the servers are in series (each server has a separate queue) or in parallel (one queue for all servers), whether preemption is allowed (a server can stop processing a customer to deal with another emergency customer) etc. The common assumption is that the service times of the customers are independent of the arrival process and it is exponentially distributed with the parameter µ. Customers may also be served in batches and it is termed as bulk service queueing system. The batches may be of fixed size or of variable size. Neuts (1981) introduced the general bulk service rule. As per the rule, (i) the server starts service only if the minimum batch size 'a (quorum) number of customers are waiting in the queue and the maximum capacity is 'b' (ii) if the server finds 'm' ($a \leq m \leq b$) customers in the queue then the entire queue is taken up for service (iii) the queue length is more than 'b' then the first 'b' customers are taken for service leaving others to wait in the queue. The late arrivals are not allowed to join the ongoing service. In particular, if a=1, the above rule will be called as usual bulk service rule and if a=b=k, the rule is then called fixed size bulk service rule.

**Queue discipline**

The method by which the customers are selected from the queue is called as queue discipline. The most common discipline that is observed in everyday life is first-in first-out (FIFO) also known as first-come first-served (FCFS) under which, the customers are served in the strict order of their arrival. Another queue discipline is last-in first-out (LIFO) which is applicable to some inventory systems where there is no obsolescence of stored units, as it is easier to reach the nearest items, which is last-in. Yet another queue discipline is service in random order (SIRO) in which the customers are served randomly irrespective of their arrivals into the system.

Priority queue discipline allows priority in service to some customers in relation to other customers waiting in the queue. It is further subdivided into two categories via preemptive priority and non-preemptive priority. In preemptive case the customer with higher priority is allowed to enter service immediately suspending even the service in progress to a customer with lower priority. In the non-preemptive case, the higher priority customer goes to the head of the queue but gets into the service only after the completion of service to the customer with lower priority. It is assumed that, once a server, who is able to provide service to a waiting customer becomes free, the customer immediately enters into service without loss of time.

**System capacity**

The number of customers in the queue and in service put together is called system capacity. A system may have a queue of finite capacity or effectively infinite capacity. A system with finite capacity can be viewed as one with forced balking of a customer arriving when the system has full capacity. If the system capacity is not mentioned, it is assumed to be infinite.

**Service Channels**

A queueing system may have one or more service channels to provide service to customers. The service channels may be arranged in parallel or in series or a combination of both depending on the nature of the service. It is generally assumed that the service mechanisms of the parallel channels operate independently of each other. A queueing system of only one server is called a single server queueing model and a system with two or more number of parallel servers is called a multi-server queueing model. In case of multi-server models the customers may form a single queue or a parallel queue in front of each server.

**References:**

[1] Adan, Ivo & Boucherie, R.J. (2014). Queues in Health. Queueing Systems. 79. 10.1007/s11134-014-9430-x.

[2] Afrane, Samuel & Appah, Alex. (2014). Queuing theory and the management of Waiting-time in Hospitals: The case of Anglo Gold Ashanti Hospital in Ghana. International Journal of Academic Research in Business and Social Sciences. Vol. 4. 34-44. 10.6007/IJARBSS/v4-i2/590.

[3] Agrawal, Sachin & Singh, B.K. (2018). Development and Analysis of Generalized Queuing Model. International Journal of Computer Sciences and Engineering. 6. 15-32. 10.26438/ijcse/v6i11.1532.

[4] Ailobhio, Titilayo & Owolabi, T & Ayoo, Peter. (2020). Application of Queuing Theory in Antenatal Clinics. IOSR Journal of Mathematics. 16. 42-47. 10.9790/5728-1606034247.

[5] Ameh, Nkeiruka & Sabo, B & Oyefabi, Moses. (2013). Application of queuing theory to patient satisfaction at a tertiary hospital in Nigeria. Nigerian medical journal : journal of the Nigeria Medical Association. 54. 64-7. 10.4103/0300-1652.108902.

[6] Ariff, Hajar & Kamardan, M & Sufahani, Suliadi & Ali, Maselan. (2018). Review on Queueing Problem in Healthcare. International Journal of Engineering & Technology. 7. 304. 10.14419/ijet.v7i4.30.22291.

[7] Aziati, A. & Hamdan, Nur Salsabilah. (2018). Application Of Queuing Theory Model and Simulation to Patient Flow at The Outpatient Department

[8] Borthakur, Partha & Borthakur, Barbie. (2021). An Introduction to Queuing Theory. 10.13140/RG.4.24284.72322.

[9] Cho, Kyoung & Kim, Seong & Chae, Young & Song, Yong. (2017). Application of Queueing Theory to the Analysis of Changes in Outpatients' Waiting Times in Hospitals Introducing EMR. Healthcare Informatics Research. 23. 35-42. 10.4258/hir.2017.23.1.35.

[10] Cochran, K. J., & Bharti, A. (2006), ‘A Multi-stage Stochastic Methodology for Whole Hospital Bed Planning Under Peak Loading’, International Journal of Industrial and Systems Engineering (1 (1/2)), pp.8-35.

[11] Gronroos, C. (1984),’A service quality model and its marketing implications’, European Journal of Marketing, 18(4), pp. 36-44.

[12] Gupta, Diwakar. (2013). Queueing Models for Healthcare Operations. 10.1007/978-1-4614-5885-2_2.

[13] Haight, F.A. (1957) Queuing with Balking. Biometrika, 44:362–369.

[14] Haight, F.A. (1959) Queuing with Reneging. Metrika, 2:186–197.