QUEUING THEORY

immediate

April 9, 2025

Contents

INTRODUCTION

Queuing theory is a branch of mathematics that studies and models the act of waiting in lines. This paper will take a brief look at the formulation of queuing theory along with examples of models and applications of their use.

The first paper on queuing theory, "The Theory of Probabilities and Telephone Conversations" was published in 1909 by A.K. Erlang, now considered the father of the field. His work with the Copenhagen Telephone Company was what prompted his initial foray into the field. He pondered the problem of determining how many telephone circuits were necessary to provide telephone service that would prevent customers from waiting too long for an available circuit. In developing a solution to this problem, he began to realize that the problem of minimizing waiting time was applicable to many fields, and began developing the theory further.

There are many situations in daily life when a queue is formed. for example, machines waiting to be repaired, patients waiting in a Doctor's room, cars waiting at a traffic signal and passengers waiting to buy tickets in counter form queues.

Queues may be formed if the server required by the cutomer is not immediately available,that is, if the current demand for a particular service exceeds the capacity to provide the services.

Queues may be incresed in size or prevented from forming by providing additional service facilities which results in a drop in the profit. On the other hand, excessively long queues may result in lost sales and lost customers. Hence the problem of interest is how to achieve a balance between the cost associated with the prevention of waiting and the cost associated with long waiting in order to maximise the profits.

As queuing theory provides an answer to this problem, it has become a topic of interest.

PRELIMINARIES

Poisson Probability Distribution

A Poisson queue is a queuing model in which the number of arrivals per unit time and the number of completions per unit time, when there are customers waiting, both have Poisson distribution. It is good to use if the arrivals are all randoms and independent of each other. For the Poisson distribution, the probability that there are exactly 'n' arrivals during 't' amount of time,

$$prob(n) = \frac{(\lambda t)^n e^-(\lambda t)}{n!} \tag{0.0.1}$$

where 't' is the duration of time λ is the expected (average) number of arrivals per hour, or day or whatever unit t is measured in λt is therefore the expected number of arrivals during the t time period, n is the possible number of customers arriving.

Exponential Distribution

Exponential distribution is defined as $f(t) = \lambda e^-\lambda t$; $t > 0$. Then equivalently, its probability density function is given by $f(t) = \lambda e^-\lambda t$; $t > 0$.

It expresses the probability of the time between events. The first is the fact that the exponential function is strictly decreasing function of t.

Another important property of the exponential distribution is what is known as the no-memory property.

Lack of memory property

Another important property is forgetfulness or lack of memory property, it suggests that the time until the next arrival will never depend on how much time has already passed.

Let the exponential distribution f(t), represents the time t,between successive events. If s is the interval then,

$$P(t > T + s/t > s) = P(t > T) \tag{0.0.2}$$

This is a relation between exponential and Poisson distribution.

The Poisson distribution is need to determine the probability of certain number of arrivals occurring in a given time period.
The Poisson distribution with parameter λ is given by,

$$P_n(t) = \frac{(e)^{\lambda t}(\lambda t)^n}{n!} \tag{0.0.3}$$

If we set n = 0, the Poisson distribution given us $e^{-\lambda t}$ which is equal to P$(T > 0)$ from the exponential distribution.

Chapter 1

BASIC QUEUING MODELS

1.1 Queuing systems

A queuing system consists of one or more servers, an arrival pattern of customers, service pattern, queue discipline, the order in which the service is provided and customer behaviour. The word 'queue' is sometimes used to describe the whole system, but it is part of the system that holds the excess customers who cannot be immediately served. Hence, the number of customers in the system at any given time will equal to the number of customers in the queue plus the number customers in service. Of course, these numbers will vary with time due to customers arrival and departure, so they are formally random process.

1.1.1 The input (or arrival) pattern

The input describes the manner in which the customers arrive and join the queuing system. It is not possible to observe and control the actual moment of arrival of a customer for service. Hence, the number of arrival in one-time period or the the interval between successive arrivals is not treated as a constant, but a random variable. So the mode of arrival of customers is expressed by means of the probability distribution of the number of arrivals per unit time or of the inter-arrival time.

We shall deal with only those queuing systems in which the number of arrivals per unit time has a Poisson distribution with mean λ.

In this case, the between consecutive arrivals has an exponential distribution with mean $\frac{1}{\lambda}$.

Further, the input process should specify the number of queues that are permitted to form, the maximum queue length, and the maximum number of customers requiring services, viz. the nature of source (finite or infinite) from which the customers emanate.

1.1.2 The service pattern

The mode of service represented by means of probability distribution of the number of customers serviced per unit time or of the inter-service time.

We shall deal with only those queuing systems in which the customers serviced per unit of time has a Poisson distribution with mean μ or equivalently the inter-service time (viz. The time to complete the service for a customer) has an exponential distribution with mean $\frac{1}{\mu}$.

Further, the service process should specify the number of services and the arrangement of servers (in parallel, in series, etc.), as the behavior of the queueing system also depends on them.

The following figure represents the frame word of queuing system in which only one queue is permitted to form:

1.1.3 The queue discipline

The queue discipline specifies the manner in which customers form the queue or equivalently the manner in which they are selected for service, when a queue has been formed.

The most common discipline is the ***FCFS*** (First Come First Serve) or ***FIFO*** (First In First Out) according to which the customers are served in the strict order of their arrival.

If the last arrival in the system is served first, we have ***LCFS*** or ***LIFO*** (Last In First Out) discipline.

If the service is given in random order, we have the ***SIRO*** discipline.In the queuing system which we deal with we shall assume that service is provided on the ***FCFS*** basis.

Chapter 2

QUEUING MODELS

2.1 Symbolic Representation of a queuing model

Usually, a queuing model is specified and represented symbolically in the form (a/d/c):(d/e), where 'a' denotes the type of distribution of the number of arrivals per unit time, 'b' denotes the type distribution of the service time, 'c' the number of servers, 'd' the capacity of the system viz., the maximum queue size and 'e' the queue discipline.

Accordingly the first four models are denoted by the symbols,
$(M/M/1) : (\infty/FIFO)$,
$(M/M/S) : (\infty/FIFO)$,
$(M/M/1) : (K/FIFO)$,
$(M/M/S) : (K/FIFO)$,

where M stands for 'Markov', indicating that the number of arrivals in time t and the number of completed services in time t follows Poisson process which is a continuous time Markov chain.

Kendall-Lee notations

Since describing all of the characteristics of a queue inevitably becomes very wordy, a much simpler notation (known as Kendall-Lee notation) can be used to describe a system.

2.2 Difference equations related to Poisson queue systems

If the characteristics of a system are independent of the time or equivalently if the behaviour of the system is independent of time, the system is said to be in steady-state. Otherwise it is said to in transient state.

Let $P_n(t)$ be the probability that there are n customers in the system at time t $(n > 0)$. Let us first derive the differential equation satisfied by $P_n(t)$ and then deduce the difference equation satisfied by p_n in the steady state.

Let λ_n be the average arrival rate when there are n customers in the system and let μ_n be the average service rate when there are n customers in the system.

Now, $P_n(t + \Delta t)$ is the probability of n customers at time $t + \Delta t$. The presence of n customers in the system at time $t + \Delta t$ can happen in any one of the following four mutually exclusive ways:

i) Presence of n customers at t and no arrival or departure during Δt time.

ii) Presence of (n-1) customers at t and 1 arrival and no departure during Δt time.

iii) Presence of (n+1) customers at t and no arrivals and 1 departure during Δt time.

iv) Presence of n customers at t and 1 arrival and 1 departure during Δt time (since, more than 1 arrival or departure during Δt is ruled out).

$$P_n(t) = P_n(t)(1-\lambda_n\Delta t)(1-\mu_n\Delta t) + P_{n-1}(t)(\lambda_{n-1}\Delta t)(1-\mu_{n-1}\Delta t) + P_{n+1}(t)(1-\lambda_{n+1}\Delta t)(\mu_{n+1}\Delta t) + P_n(t)(\lambda_n\Delta t)(\mu_n\Delta t)$$

(Since P(an arrival occurs during (Δt) time) $= \lambda\Delta t$ etc.)

$$P_n(t + \Delta t) = P_n(t) - (\lambda_n + \mu_n)P_n(t)\Delta t + \lambda_{n-1}P_{n-1}(t)\Delta t + \mu_{n+1}P_{n+1}(t), \tag{2.2.1}$$

On omitting terms containing $(\Delta t)^2$ which is negligibly small.

$$\therefore \frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = \lambda_{n-1}P_{n-1} - (\lambda_n + \mu_n)P_n(t) + \mu_{n+1}P_{n+1} \tag{2.2.2}$$

Taking limit on both sides of equation (2.2.1) as $\Delta t \longrightarrow 0$, we have,

$$P_n^{;}(t) = \lambda_{n-1}P_{n-1} - (\lambda_n + \mu_n)P_n(t) + \mu_{n+1}P_{n+1} \tag{2.2.3}$$

Equation (2.2.2) does not hold for n = 0, as $P_{n-1}(t)$ does not exist.

Hence we derive the differential equation satisfied by $P_0(t)$ independently proceeding as before,

$$P_0(t + \Delta t) = P_0(t)(1 - \lambda_0\Delta t) + P_1(t)(1 - \lambda_1\Delta t)\mu_1\Delta t, \tag{2.2.4}$$
$$\tag{2.2.5}$$

$$\therefore \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda_0 P_0(t) + \mu_1 P_1(t) \tag{2.2.6}$$

Taking limit as $\Delta t \longrightarrow 0$,

$$P_0^{;} = -\lambda_0 P_0(t) + \mu_1 P_1(t) \tag{2.2.7}$$

Now in the steady-state, $P_n(t)$ and $P_0(t)$ are independent of time and hence, $P_n^{;}(t)$ and $P_0^{;}(t)$ becomes zero.

Hence the differential equation (2.2.2) and (2.2.4) reduces to the difference equations

$$\lambda_{n-1}P_{n-1} - (\lambda_n + \mu_n)P_n + \mu_{n+1}P_{n+1} = 0 \tag{2.2.8}$$

and

$$\lambda_0 P_0 + \mu_1 P_1 = 0 \tag{2.2.9}$$

2.2.1 Values of P_0 and P_1 for Poisson systems

From equation (2.2.9), derived above we have,

$$P_1 = \frac{\lambda_0}{\mu_1} P_0 \tag{2.2.10}$$

Putting n = 1 in equation (2.2.8) and using equation (2.2.10) we have,

$$\mu_2 P_2 = (\lambda_1 + \mu_1)P_1 - \lambda_0 P_0 \tag{2.2.11}$$

$$= (\lambda_1 + \mu_1)\frac{\lambda_0}{\mu_1}P_0 - \lambda_0 P_0 \tag{2.2.12}$$

$$= \frac{\lambda_0\lambda_1}{\mu_1}P_0 \tag{2.2.13}$$

$$P_2 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}P_0 \tag{2.2.14}$$

Putting n=2,3,4,... in equation(2.2.8) and proceeding.
Similarly, we get,

$$P_3 = \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3}P_0 \tag{2.2.15}$$

Finally,

$$P_n = \frac{\lambda_0\lambda_1...\lambda_{n-1}}{\mu_1\mu_2...\mu_n}P_0 \tag{2.2.16}$$

for n=1,2,3...

Since number of customers in the system can be 0 or 1 or 2 or etc., which events are mutually exclusive and exhaustive. We have, $\sum_{n=0}^{\infty} P_n = 1$
ie.,

$$P_0 + \sum_{n=1}^{\infty}\left[\frac{\lambda_0\lambda_1...\lambda_{n-1}}{\mu_1\mu_2...\mu_n}\right]P_0 = 1 \tag{2.2.17}$$

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1...\lambda_{n-1}}{\mu_1\mu_2...\mu_n}P_0} \tag{2.2.18}$$

Equations (2.2.16) and (2.2.18) will be used to derive the important characteristics of four queuing models.

2.3 Characteristics of queuing models

2.3.1 Infinite capacity, Single server, Poisson queue

Model 1 (M/M/1):(∞/FIFO), [When $\lambda_n = \lambda$ and $\mu_n = \mu(\lambda < \mu)$]

1. *Average number L_s of customers in the system*:
   Let N denote the number of customers in the queuing system (ie., those in queue and the one who is being served) N is a discrete random variable, which can take values 0,1,.......∞ such that P(N=n)=$p_n = (\frac{\lambda}{\mu}^n p_0)$, (from (2.2.16) of previous section)

   From equation (2.2.18) we have,

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty}(\frac{\lambda}{\mu})^n} = 1 - \frac{\lambda}{\mu} \tag{2.3.1}$$

$$\therefore p_n = (\frac{\lambda}{\mu})^n(1 - \frac{\lambda}{\mu}) \tag{2.3.2}$$

   Now,

$$L_s = E(N) = \sum_{n=1}^{\infty} n \times p_n \tag{2.3.3}$$

$$= \left[\frac{\lambda}{\mu}\right]\left[1 - \frac{\lambda}{\mu}\right]\left[1 - \frac{\lambda}{\mu}\right]^{-2} \tag{2.3.4}$$

$$= \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} \tag{2.3.5}$$

$$= \frac{\lambda}{\mu - \lambda} \tag{2.3.6}$$

$$\therefore L_s = \frac{\lambda}{\mu - \lambda} \tag{2.3.7}$$

2. *Average number L_q of customers in the queue or*
   Average length of the queue :
   If N is the number of customers in the system, then the number of customers in the queue is (N-1).

$$\therefore L_q = E(N-1) = \sum_{n=1}^{\infty}(n-1)p_n \tag{2.3.8}$$

$$= \left[1 - \frac{\lambda}{\mu}\right]\sum_{n=1}^{\infty}(n-1)\left[\frac{\lambda}{\mu}\right]^n \tag{2.3.9}$$

$$= \left[\frac{\lambda}{\mu}\right]^2 \left[1 - \frac{\lambda}{\mu}\right] \sum_{n=2}^{\infty} (n-1) \left[\frac{\lambda}{\mu}\right]^{n-2} \tag{2.3.10}$$

$$= \frac{\left[\frac{\lambda}{\mu}\right]^2}{1 - \frac{\lambda}{\mu}} \tag{2.3.11}$$

$$= \frac{\lambda^2}{\mu(\mu - \lambda)} \tag{2.3.12}$$

$$\therefore L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} \tag{2.3.13}$$

3. *Average number L_w of customers in non-empty queues*:

$$L_w = E[(N-1)(N-1)] > 0, \tag{2.3.14}$$

since the queue in non-empty.

$$L_w = \frac{E(N-1)}{P(N-1>0)} \tag{2.3.15}$$

$$= \frac{\lambda^2}{\mu - \lambda} \times \frac{1}{\sum_{n=2}^{\infty} P_n} \tag{2.3.16}$$

$$= \frac{\lambda^2}{\mu(\mu - \lambda)} \times \frac{1}{\sum_{n=2}^{\infty} (\frac{\lambda}{\mu})^n (1 - \frac{\lambda}{\mu})} \tag{2.3.17}$$

$$= \frac{\mu}{\mu - \lambda} \times \frac{1}{(1 - \frac{\lambda}{\mu})(1 - \frac{\lambda}{\mu})^{-1}} \tag{2.3.18}$$

$$= \frac{\mu}{\mu - \lambda} \tag{2.3.19}$$

$$\therefore L_w = \frac{\mu}{\mu - \lambda} \tag{2.3.20}$$

4. *Probability that the number of customers in the system exceeds k*:

$$P(N > k) = \sum_{n=k+1}^{\infty} P_n \tag{2.3.21}$$

$$= \sum_{n=k+1}^{\infty} \tag{2.3.22}$$

5. *Probability density function of the waiting time in the system*:

Let W_s be the continuous random variable that represent the waiting time of a customer in the system viz., the time between arrival and completion of service.

Let its probability density function (denoted as pdf) be $f(w)$ and let $f(w/n)$ be the density function of W_s subject to the condition that there are n customers in the queuing system when the customer arrives then,

f(w) = $\sum_{n=0}^{\infty} f(w/n)P_n$

Now, $f(w/n)$ = pdf of sum of (n+1) service time(one part-service time of the customers being served + complete served time).

$f(w/n)$ = pdf of sum of (n+1) independent random variable, each of which is exponentially distributed with parameter μ.

ie., $f(w/n) = \frac{\mu^{n+1}}{n!} e^{-\mu w} w^n; w > 0$ which is the pdf of Erlang distribution.

$$\therefore f(w) = \sum_{n=0}^{\infty} \frac{\mu^{n+1}}{n!} e^{-\mu w} w^n \left[\frac{\lambda}{\mu}\right]^n \left[1 - \frac{\lambda}{\mu}\right] \tag{2.3.23}$$

$$= \mu e^{-\mu w} \left[1 - \frac{\lambda}{\mu}\right] \sum_{n=0}^{\infty} \frac{1}{n!} (\lambda w)^n \tag{2.3.24}$$

$$= \mu \left[1 - \frac{\lambda}{\mu}\right] e^{-\mu w} e^{\lambda w} \tag{2.3.25}$$

$$= (\mu - \lambda) e^{-(\mu - \lambda) w} \tag{2.3.26}$$

which is the pdf of an exponential distribution with parameter $(\mu - \lambda.)$

6. *Average waiting time of a customer in the system*:

W_s follows an exponential distribution with parameter $(\mu - \lambda\ )$

$$\therefore E(W_s) = \frac{1}{\mu - \lambda} \tag{2.3.27}$$

7. *Probability that the waiting time of a customer in the system exceeds* t:

$$P(W_s > t) = \int_t^\infty f(w)dw \tag{2.3.28}$$
$$= \int_t^\infty (\mu - \lambda)e^{-(\mu-\lambda)w}dw \tag{2.3.29}$$
$$= \left[-e^{-(\mu-\lambda)w}\right]_t^\infty \tag{2.3.30}$$
$$= e^{-(\mu-\lambda)t} \tag{2.3.31}$$

8. *Probability density function of the waiting time* W_q *in the queue*:

   W_q represents the time between arrival and reach of service point.

   Let the pdf of W_q be $g(w)$ and let $g(w/n)$ be the density function of W_q subject to the condition that there are n customers in the system or there are (n - 1) customers in the queue apart from one customer receiving service.

   $g(w/n)$ = pdf of sum of n service time [one residual service time + (n - 1) full services times]

$$g(w/n) = \frac{\mu^n}{(n-1)!}e^{-\mu w}w^{n-1}; w > 0 \tag{2.3.32}$$
$$= \sum_{n=1}^{\infty} \frac{\mu^n}{(n-1)!}e^{-\mu w}w^{n-1}\left[\frac{\lambda}{\mu}\right]^n\left[1 - \frac{\lambda}{\mu}\right] \tag{2.3.33}$$
$$= \frac{\lambda}{\mu}(\mu - \lambda)e^{-\mu w}e^{\lambda w} \tag{2.3.34}$$
$$= \frac{\lambda}{\mu}(\mu - \lambda)e^{-(\mu-\lambda)w}; w > 0 \tag{2.3.35}$$
$$= 1 - \frac{\lambda}{\mu}; w = 0 \tag{2.3.36}$$

9. *Average waiting time of a customer in the queue*:

$$E(W_q) = \frac{\lambda}{\mu}(\mu - \lambda)\int_0^\infty we^{-(\mu-\lambda)w} \tag{2.3.37}$$

$$\therefore E(W_q) = \frac{\lambda}{\mu(\mu - \lambda)} \tag{2.3.38}$$

10. *Average waiting time of a customer in the queue, if he has to wait*:

$$E(W_q/W_q > 0) = \frac{E(W_q)}{P(W_q > 0)} \tag{2.3.39}$$

$$= \frac{E(W_q)}{1 - P(W_q = 0)} \tag{2.3.40}$$

$$= \frac{\lambda}{\mu(\mu - \lambda)} \frac{\mu}{\lambda} \tag{2.3.41}$$

$$= \frac{1}{\mu - \lambda} \tag{2.3.42}$$

2.3.2 Infinite capacity, Multiple server, Poisson queue

Model II (M/M/S):(∞/FIFO) model , [When $\lambda_n = \lambda$ for all n ($\lambda < s\mu$)]

1. *Values of P_0 and P_n*:

For the Poisson queue system P_n is given by,

$$P_n = \frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 ... \mu_n} P_0, n \geq 0 \tag{2.3.43}$$

where,

$$P_0 = \left[1 + \sum_{n=1}^{\infty} \left[\frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 ... \mu_n} \right] \right]^{-1} \tag{2.3.44}$$

If there is a single server $\mu_n = \mu, \forall$ n. But there are s servers working independently of each other.

If there is less than s customers ie., if n<s, only n of the s servers will be busy and the others idle and hence the mean service rate will be nμ.

If n $\geq$ s, all the servers will be busy and hence the mean service rate = sμ.
That is,

$$\mu_n = \begin{cases} n\mu, & \text{if } 0 \leq n < s \\ s\mu, & \text{if } n \geq s \end{cases} \tag{2.3.45}$$

Using equation (2.3.45) in (2.3.43) and (2.3.44) , we have

$$P_n = \frac{\lambda^n}{1\mu.2\mu...n\mu} P_0, \text{ if } 0 \leq n < s \tag{2.3.46}$$

$$= \frac{1}{n!}\left[\frac{\lambda}{\mu}\right]^n P_0, \text{ if } 0 \leq n < s \tag{2.3.47}$$

and

$$P_n = \frac{\lambda^n}{[1\mu.2\mu...(s-1)\mu][s\mu.s\mu...(n-s+1)\mu]} \times P_0 \tag{2.3.48}$$

$$= \frac{\lambda^n}{(s-1)!\mu^{s-1}(s\mu)^{n-s+1}} \times P_0 \tag{2.3.49}$$

$$= \frac{1}{s!s^{n-s}}\left[\frac{\lambda}{\mu}\right]^n P_0, \text{ if } n \geq s \tag{2.3.50}$$

Now, P_0 is given by $\sum_{n=0}^{\infty} P_n = 1$

$$\Longrightarrow \left[\sum_{n=0}^{s-1}\frac{1}{n!} + \sum_{n=s}^{\infty}\frac{1}{s!s^{n-s}}(\frac{\lambda}{\mu})^n\right] P_0 = 1$$

$$\Longrightarrow \left[\sum_{n=0}^{s-1}\frac{1}{n!} + \frac{s^s}{s!}(\frac{\lambda}{\mu})^s\frac{1}{1-\frac{\lambda}{\mu s}}\right] P_0 = 1$$

$$\Longrightarrow \left[\sum_{n=0}^{s-1}\frac{1}{n!}(\frac{\lambda}{\mu})^n + \frac{1}{s!(1-\frac{\lambda}{\mu s})}(\frac{\lambda}{\mu})^s\right] P_0 = 1$$

or,

$$P_0 = \frac{1}{\left[\sum_{n=0}^{s-1}\frac{1}{n!}(\frac{\lambda}{\mu})^n + \frac{1}{s!(1-\frac{\lambda}{\mu s})}(\frac{\lambda}{\mu})^s\right]} \tag{2.3.51}$$

2. *Average number of customers in the queue or average queue length*:

$$L_q = E(N_q)$$

$$= E(N-S)$$

$$= \sum_{x=0}^{\infty}(n-s)P_n$$

$$= \sum_{x=0}^{\infty} xP_{x+s}$$

$$
\begin{aligned}
&= \sum_{x=0}^{\infty} x \frac{1}{s!s^x} (\frac{\lambda}{\mu})^{s+x} P_0 \\
&= \frac{1}{s!} (\frac{\lambda}{\mu})^s \frac{\lambda}{\mu s} P_0 \frac{1}{(1-\frac{\lambda}{\mu s})^2} \\
&= \frac{1}{s.s!} \frac{(\frac{\lambda}{\mu})^{s+1}}{(1-\frac{\lambda}{\mu s})^2} P_0
\end{aligned}
$$

$$
L_q = \frac{1}{s.s!} \frac{(\frac{\lambda}{\mu})^{s+1}}{(1-\frac{\lambda}{\mu s})^2} P_0 \tag{2.3.52}
$$

3. *Average number of customers in the system*:

   By Little's formula,

$$
\begin{aligned}
E(N_s) &= E(N_q) + \frac{\lambda}{\mu} \\
&= \frac{1}{s.s!} \frac{(\frac{\lambda}{\mu})^{s+1}}{(1-\frac{\lambda}{\mu s})^2} P_0 + \frac{\lambda}{\mu}
\end{aligned}
$$

$$
E(N_s) = \frac{1}{s.s!} \frac{(\frac{\lambda}{\mu})^{s+1}}{(1-\frac{\lambda}{\mu s})^2} P_0 + \frac{\lambda}{\mu} \tag{2.3.53}
$$

4. *Average number of customer has to be spend in the system*:

   By Little formula,

$$
\begin{aligned}
E(W_q) &= \frac{1}{\lambda} E(N_q) \\
&= \frac{1}{\mu} + \frac{1}{\mu} \frac{1}{s.s!} \frac{(\frac{\lambda}{\mu})^{s}}{(1-\frac{\lambda}{\mu s})^2} P_0
\end{aligned}
$$

$$
E(W_q) = \frac{1}{\mu} + \frac{1}{\mu} \frac{1}{s.s!} \frac{(\frac{\lambda}{\mu})^{s}}{(1-\frac{\lambda}{\mu s})^2} P_0 \tag{2.3.54}
$$

5. *Average time a customer has to be spend in the queue*:

By Little's formula,

$$
\begin{aligned}
E(W_q) &= \frac{1}{\lambda}E(N_q) \\
&= \frac{1}{\mu}\frac{1}{s.s!}\frac{(\frac{\lambda}{\mu})^s}{(1-\frac{\lambda}{\mu s})^2}P_0
\end{aligned}
$$

$$
E(W_q) = \frac{1}{\mu}\frac{1}{s.s!}\frac{(\frac{\lambda}{\mu})^s}{(1-\frac{\lambda}{\mu s})^2}P_0 \qquad (2.3.55)
$$

6. *Probability that an arrival has to wait*:

Required probability = Probability that there are s or more customers in the system

$$
\begin{aligned}
\implies P(W_s > 0) &= P(N \geq s) \\
&= \sum_{n=s}^{\infty} P_n \\
&= \sum_{n=s}^{\infty} \frac{1}{s.s!}(\frac{\lambda}{\mu})^n P_0 \\
&= \frac{(\frac{\lambda}{\mu})^s}{s!(1-\frac{\lambda}{\mu s})^2}P_0
\end{aligned}
$$

$$
P(W > 0) = \frac{(\frac{\lambda}{\mu})^s}{s!(1-\frac{\lambda}{\mu s})^2}P_0 \qquad (2.3.56)
$$

7. *Probability that an arrival enters the service without waiting*:

Required probability = 1 - P(arrival has to wait)

$$
\therefore \text{Required probability } = 1 - \frac{(\frac{\lambda}{\mu})^s}{s!(1-\frac{\lambda}{\mu s})}P_0 \qquad (2.3.57)
$$

8. *Mean waiting time in the queue for those who actually wait*:

$$\begin{aligned}
E(W_q/W_s > 0) &= \frac{E(W_q)}{P(W_s > 0)} \\
&= \frac{1}{\mu}\frac{1}{s.s!}\frac{(\frac{\lambda}{\mu})^s}{(1-\frac{\lambda}{\mu s})^2}P_0 \times \frac{s!(1-\frac{\lambda}{\mu s})}{(\frac{\lambda}{\mu})^s P_0} \\
&= \frac{1}{\mu s(1-\frac{\lambda}{\mu s})} \\
&= \frac{1}{\mu s - \lambda}
\end{aligned}$$

$$\therefore E(W_q/W_s > 0) = \frac{1}{\mu s - \lambda} \tag{2.3.58}$$

9. *Probability that there will be someone waiting*:

$$\begin{aligned}
\text{Required probability} &= P(N \geq s+1) \\
&= \sum_{n=s+1}^{\infty} P_n \\
&= \sum_{n=s}^{\infty} P_n - P(N = s) \\
&= \frac{(\frac{\lambda}{\mu})^s}{s!(1-\frac{\lambda}{\mu s})}P_0 - \frac{(\frac{\lambda}{\mu})^s}{s!}P_0 \\
&= \frac{(\frac{\lambda}{\mu})^s}{s!}P_0 \times \frac{(\frac{\lambda}{\mu})^s P_0}{1-\frac{\lambda}{\mu s}}
\end{aligned}$$

$$\therefore \text{ Required probability} = \frac{(\frac{\lambda}{\mu})^s}{s!}P_0 \times \frac{(\frac{\lambda}{\mu})^s P_0}{1-\frac{\lambda}{\mu s}} \tag{2.3.59}$$

10. *Average number of customers (in non-empty) queues who have to actually wait*:

$$\begin{aligned}
L_w &= E(N_q/N_q \geq 1) \\
&= E(N_q)/P(N \geq s) \\
&= \frac{1}{s.s!}\frac{(\frac{\lambda}{\mu})^{s+1}}{(1-\frac{\lambda}{\mu s})^2}P_0 \times \frac{s!(1-\frac{\lambda}{\mu s})}{(\frac{\lambda}{\mu})^s P_0}
\end{aligned}$$

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

$$\therefore L_w = \frac{\frac{\lambda}{\mu s}}{1 - \frac{\lambda}{\mu s}} \tag{2.3.60}$$

2.3.3 Finite capacity, Single server, Poisson queue

Model III (M/M/I):(K/FIFO)

1. *Values of P_0 and P_n:*

   For the poisson queue system, $P_n = P(N = n)$ in the steady state is given by the difference equations,

$$\lambda_{n-1}P_{n-1} - (\lambda_n + \mu_n)P_n + \mu_{n+1}P_{n+1} = 0 \text{ ; n} > 0$$

   and

$$-\lambda_0 P_0 + \mu_1 P_1 = 0 \text{ ; n} = 0$$

   This model represents the situation in which the system can accommodate only a finite number k of arrivals. If a customer arrives and the queue is full. The customer leaves without joining the queue. For this model,

$$\mu_n = \mu, \text{ n=1,2,3,...}$$

   and

$$\lambda_n = \begin{cases} \lambda, & \text{for n=0, 1, 2,..., (k - 1)} \\ 0, & \text{for n=k, k + 1, ...} \end{cases}$$

   Using these values in the difference equations given above, we have,

$$\mu P_1 = \lambda P_0 \tag{2.3.61}$$

$$\mu P_{n+1} = (\lambda + \mu)P_n - \lambda P_{n-1}, \text{ for } 1 \leq n \leq k-1 \text{ and} \tag{2.3.62}$$

$$\mu P_k = \lambda P_{k-1}, \text{ for n=k (since, } P_{k+1} \text{ has no meaning)} \tag{2.3.63}$$

From equation (2.3.51),

$$P_1 = \frac{\lambda}{\mu} P_0$$

From equation (2.3.52),

$$\mu P_2 = (\lambda+\mu)\frac{\lambda}{\mu}P_0 - \lambda P_0$$
$$\therefore P_2 = \left[\frac{\lambda}{\mu}\right]^2 P_0$$

and so on.

In general,

$$P_n = \left[\frac{\lambda}{\mu}\right]^n P_0 \text{ true for } 0 \le n \le k-1$$

From equation (2.3.53),

$$P_k = \frac{\lambda}{\mu}\left[\frac{\lambda}{\mu}\right]^{k-1} P_0 = \left[\frac{\lambda}{\mu}\right]^k P_0$$

Now,

$$\sum_{n=0}^{k} P_n = 1$$
$$\Longrightarrow P_0 \sum_{n=0}^{k} \left[\frac{\lambda}{\mu}\right]^n = 1$$
$$\Longrightarrow \frac{P_0\left[1-(\frac{\lambda}{\mu})^{k+1}\right]}{1-(\frac{\lambda}{\mu})} = 1 \text{ which is valid even for } \lambda > \mu.$$

$$\therefore P_0 = \begin{cases} \frac{1-(\frac{\lambda}{\mu})}{1-(\frac{\lambda}{\mu})^{k+1}}, & \text{if } \lambda \neq \mu \\ \frac{1}{k+1}, & \text{if } \lambda = \mu, \because lim_{\frac{\lambda}{\mu}\to 0}\left[\frac{1-\frac{\lambda}{\mu}}{1-(\frac{\lambda}{\mu})^{k+1}}\right] = \frac{1}{k+1} \end{cases} \tag{2.3.64}$$

$$\therefore P_n = \begin{cases} \left[\frac{\lambda}{\mu}\right]^n \left[\frac{1-\frac{\lambda}{\mu}}{1-(\frac{\lambda}{\mu})^{k+1}}\right], & \text{if } \lambda \neq \mu \\ \frac{1}{k+1}, & \text{if } \lambda = \mu \end{cases} \tag{2.3.65}$$

2. *Average number of customers in the system*:

$$
\begin{aligned}
E(N) &= \sum_{n=0}^{k} P_n \\
&= \frac{1-\frac{\lambda}{\mu}}{1-(\frac{\lambda}{\mu})^{k+1}} \sum_{n=0}^{k} n \left[\frac{\lambda}{\mu}\right]^n \\
&= \frac{\left[1-\frac{\lambda}{\mu}\right]\cdot\frac{\lambda}{\mu}}{1-\left[\frac{\lambda}{\mu}\right]^{k+1}} \sum_{n=0}^{k} \frac{d}{dx}(x^n), && \text{where } x = \frac{\lambda}{\mu} \\
&= \frac{\left[1-\frac{\lambda}{\mu}\right]\cdot\frac{\lambda}{\mu}}{1-(\frac{\lambda}{\mu})^{k+1}} \cdot \frac{d}{dx}\left[\frac{1-x^{k+1}}{1-x}\right] \\
&= \frac{x[1-(k+1)x^k + kx^{k+1}]}{(1-x)(1-x^{k+1})} \\
&= \frac{x(1-x^{k+1}) - (k+1)(1-x)x^{k+1}}{(1-x)(1-x^{k+1})} \\
&= \frac{x}{1-x} - \frac{(k+1)x^{k+1}}{1-x^{k+1}} \\
&= \frac{\lambda}{\mu-\lambda} - \frac{(k+1)(\frac{\lambda}{\mu})^{k+1}}{1-(\frac{\lambda}{\mu})^{k+1}}, && \text{if } \lambda \neq \mu
\end{aligned}
$$

$$
\therefore E(N) = \frac{\lambda}{\mu-\lambda} - \frac{(k+1)(\frac{\lambda}{\mu})^{k+1}}{1-(\frac{\lambda}{\mu})^{k+1}}, \qquad \text{if } \lambda \neq \mu \tag{2.3.66}
$$

$$
\text{and } E(N) = \sum_{n=0}^{k} \frac{n}{k+1} = \frac{k}{2}, \qquad \text{if } \lambda = \mu \tag{2.3.67}
$$

3. *Average number of customers in the queue*:

$$
\begin{aligned}
E(N_q) &= E(N-1) \\
&= \sum_{n=1}^{k} (n-1) P_n \\
&= \sum_{n=0}^{k} P_n - \sum_{n=1}^{k} P_n \\
&= E(N) - (1-P_0)
\end{aligned}
$$

$$\Longrightarrow E(N_q) = E(N) - (1 - P_0) \tag{2.3.68}$$

As per Little's formula (iv),

$$E(N_q) = E(N) - \frac{\lambda}{\mu}$$

which is true when the average arrival rate is throughout.
Now, we see that in step (2.3.56) , $(1 - P_0) \neq \frac{}{\mu}$, because the average arrival rate λ as long as there is a vacancy in the queue and it is zero when the system is full. Hence we define the overall effective arrival rate, denoted by λ' or $\lambda_{eff'}$ by using step (2.3.56) and Little's formula as

$$\frac{\lambda'}{\mu} = 1 - P_0 \text{ or } \lambda' = \mu(1 - P_0) \tag{2.3.69}$$

Thus step (2.3.56) can be written as,

$$E(N_q) = E(N) - \frac{\lambda'}{\mu'} \tag{2.3.70}$$

which is modified Little's formula for this model.

4. *Average waiting time in the system and in the queue*:
By the modified Little's formula

$$E(W_s) = \frac{1}{\lambda'} E(N) \tag{2.3.71}$$

$$\text{and } E(W_q) = \frac{1}{\lambda'} E(N_q) \tag{2.3.72}$$

where λ' is the effective arrival rate, given by step (2.3.57).

2.3.4 Finite capacity, Multiple server, Poisson queue

Model IV [(M/M/S):(K/FIFO) Model]

1. *Values of P_0 and P_n:*

   For the Poisson queue system P_n is given by,

$$P_n = \frac{\lambda_0\lambda_1\lambda_2...\lambda_{n-1}}{\mu_1\mu_2...\mu_n}P_0; n \geq 1 \tag{2.3.73}$$

$$\text{where } P_0 = \left[1 + \sum_{n=1}^{k}\frac{\lambda_0\lambda_1\lambda_2...\lambda_{n-1}}{\mu_1\mu_2...\mu_n}\right]^{-1} \tag{2.3.74}$$

   For this (M/M/S):(K/FIFO) model,

$$\lambda_n = \begin{cases} \lambda, & \text{for n= 0,1,2,...,k-1} \\ 0, & \text{for n= k,k+1,....} \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & \text{for n=0,1,2,...s-1} \\ s\mu, & \text{for n=s,s+1,...} \end{cases}$$

   Using these values of λ_n and μ_n in equation (2.3.64) and noting $1 < s < k$, we get

$$P_0^{-1} = \left[1 + \tfrac{\lambda}{1!\mu} + \tfrac{\lambda^2}{2!\mu^2} + ... + \tfrac{\lambda^{s-1}}{(s-1)!\mu^{s-1}}\right] + \left[\tfrac{\lambda^s}{(s-1)!\mu^{s-1}\mu s} + \tfrac{\lambda^{s+1}}{(s-1)!\mu^{s-1}(\mu s)^2} + ... + \tfrac{\lambda^k}{(s-1)!\mu^{s-1}(\mu s)^{k-s+1}}\right]$$

$$\therefore P_0^{-1} = \sum_{n=0}^{s-1}\frac{1}{n!}\left[\frac{\lambda}{\mu}\right]^n + \frac{1}{s!}\left[\frac{\lambda}{\mu}\right]^s\sum_{n=s}^{k}\left[\frac{\lambda}{\mu s}\right]^{n-s} \tag{2.3.75}$$

$$P_n = \begin{cases} \frac{1}{n!}(\frac{\lambda}{\mu})^n P_0, & \text{for } n \leq s \\ \frac{1}{s!s^{n-s}}(\frac{\lambda}{\mu})^n P_0, & \text{for } s \leq n \leq k \\ 0, & \text{for } n > k \end{cases} \tag{2.3.76}$$

2. *Average queue length or average number of customers in the queue:*

$$\begin{aligned} E(N_q) &= E(N - S) \\ &= \sum_{n=s}^{k}(n-s)P_n \\ &= \frac{P_0}{s!}\sum_{n=s}^{k}(n-s)\frac{(\frac{\lambda}{\mu})^n}{s^{n-s}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{(\frac{\lambda}{\mu})^s P_0}{s!} \sum_{x=0}^{k-s} x. \left[\frac{\lambda}{\mu s}\right]^x \\
&= \frac{(\frac{\lambda}{\mu})^s P_0.\rho}{s!} \sum_{x=0}^{k-s} x(\rho)^{x-1}, \qquad \text{where } \rho = \frac{\lambda}{\mu s} \\
&= \frac{(\frac{\lambda}{\mu})^s P_0.\rho}{s!} \sum_{x=0}^{k-s} \frac{d}{d\rho}(\rho)^x \\
&= (\frac{\lambda}{\mu})^s \frac{P_0\rho}{s!} \frac{d}{d\rho} \left[\frac{1-\rho^{k-s+1}}{1-\rho}\right] \\
&= P_0(\frac{\lambda}{\mu})^s \frac{\rho}{s!(1-\rho)^2}[1-\rho^{k-s}-(k-s)(1-\rho)\rho^{k-s}], \qquad \text{where } \rho = \frac{\lambda}{\mu s}
\end{aligned}
$$

$$
\therefore E(N_q) = P_0(\frac{\lambda}{\mu})^s \frac{\rho}{s!(1-\rho)^2}[1-\rho^{k-s}-(k-s)(1-\rho)\rho^{k-s}], \quad \text{where } \rho = \frac{\lambda}{\mu s} \tag{2.3.77}
$$

3. *Average number of customers in the queue*:

$$
\begin{aligned}
E(N) &= \sum_{n=0}^{k} nP_n \\
&= \sum_{n=0}^{s-1} nP_n + \sum_{n=s}^{k} nP_n \\
&= \sum_{n=0}^{s-1} nP_n + \sum_{n=s}^{k} (n-s)P_n + \sum_{n=s}^{k} nP_n \\
&= \sum_{n=0}^{s-1} nP_n + E(N_q) + s\left[\sum_{n=0}^{k} P_n - \sum_{n=0}^{s-1} P_n\right] \\
&= E(N_q) + s - \sum_{n=0}^{s-1} (s-1)P_n
\end{aligned}
$$

$$
E(N) = E(N_q) + s - \sum_{n=0}^{s-1} (s-1)P_n \tag{2.3.78}
$$

Obviously, $[s - \sum_{n=0}^{s-1}(s-n)P_n] \neq \frac{\lambda}{\mu}$, so that step (2.3,68) represents Little's formula.

In order to make (2.3.68), assume the form of Little's formula, we defime the overall effective arrival ratw λ' or $\lambda_{eff'}$ as follows,

$$\frac{\lambda'}{\mu} = s - \sum_{n=0}^{s-1}(s-n)P_n$$

$$\lambda' = \mu[s - \sum_{n=0}^{s-1}(s-n)P_n]] \tag{2.3.79}$$

With this definition of λ', step (2.3.68) becomes,

$$E(N) = E(N_q) + \frac{\lambda'}{\mu} \tag{2.3.80}$$

which is the modified Little's formula for this model.

4. *Average waiting time in the system and in the queue*:
   By the modified Little's formula ,

$$E(W_s) = \frac{1}{\lambda'}E(N) \tag{2.3.81}$$

and

$$E(W_q) = \frac{1}{\lambda'}E(N_q) \tag{2.3.82}$$

where λ' is the effective arrival rate, given by step (2.3.69).

Chapter 3

APPLICATIONS

3.1 A Library Management

A library is organized collection of books and some special materials like audio or visual materials, CD and cassettes, video tapes DVDs, e-books, audio books and many other type of electronic resources.

Scope of queuing application in libraries are circulation of books, counter service and allied services like reprography. The basic tasks in library are slacks maintenance, membership management, selection of library materials and planning the acquisition of materials.

3.2 ATM s

In ATM, bank customers arrive randomly and the service time that is the time customers takes to do transaction in ATM is also random. We use queuing model to derive the arrival rate, service rate, utilization rate, waiting time in the queue and the average number of customers in the queue.

In this (M/ M /I) queuing model is used. The technique used is Poisson process. Here the service time and arrival time are exponentially distributed.Here the single queue and single server is used.

3.3 Toll Plaza

In toll plaza,M/G/I queuing model is used. Computer simulation is on of the popular approaches to the design of toll plaza.

In this model design the toll collection methods, types of vehicles and number of toll booths configuration are used. Various performance parameters are being used, some are as average waiting time, average queue length, maximum queue length and maximum waiting time. These performances parameters are combined with designing factors such as number of manual tolls, electronic toll collection rates lane selections.

Finding appropriate values of input parameters for a traffic simulation model is always a challenge to simulation model, deterministic traffic counts for a time period can be used as an input parameter into a model rather than considering a probabilistic distribution.

3.4 Call Center

In Call Center application, M/M/S queuing model is used. Here the customers are referred to as callers, servers are telephone operators. The tele-queues are consists of callers that are in waiting state to be serves by the system resources.

This figure represents that there is single queue for the customers whenever the customer enters into the queue, depending on the availability of the system resources, it is served by any one of the available telephone operators.

3.5 Banking

Today banks are one of the most important units of public. Most banks used standard queuing models. It's very useful to avoid standing in a queue for a long time or in a wrong line and to give tickets to all customers. Bank is an example of infinite queue length. Queue is used to generate a sequence of customers arrival time and to choose randomly between three different services. Open an account, transaction, and balance, with different period of time for each service.

3.6 Traffic system

The vehicular traffic flow and explore could be minimized using queuing theory in order to reduce the delay on the roads. The role of transportation in human life cannot be over-emphasized. A basic model of vehicular traffic is based on queuing theory. It will determine the best times of the red, amber, and green lights to be either on or off in order to reduce traffic congestion on the roads. Queuing also helps to reduce fuel consumption there by saving money for the Government to tackle problem of other sectors of the economy.