UNIT 5:

CONTINUOUS PARAMETER MARKOV CHAIN

The Birth and Death process (MM/1, M/M/c, M/M/1/N)

5.1 Basics of Queueing Processes

A queue is a waiting line; queues are formed whenever the demand for service exceeds the service availability. A queuing system is composed of customers arriving for service, waiting in a queue for the service if necessary, and after being served, leaving the system. The term customer is generic and does not imply a human customer necessarily; any unit which needs a form of service is considered a customer.

A Queueing system is usually described by five basic characteristics of queueing processes: (1) arrival pattern of customers, (2) service pattern of customers, (3) queue discipline, (4) system capacity, and (5) number of service channels.

5.2 M/M/1 Queue

In Queueing theory, a discipline within the mathematical theory of probability, an **M/M/1 queue** represents the queue length in a system having a single server, where arrivals are determined by a Poisson process and job service times have an exponential distribution. The model name is written in Kendall's notation. The model is the most elementary of queueing models and an attractive object of study as closed-form expressions can be obtained for many metrics of interest in this model.

An extension of this model with more than one server is the M/M/c queue.

5.3 An M/M/1 Queueing Node

a.Model Definition

An M/M/1 queue is a stochastic process whose state space is the set {0,1,2,3,...} where the value corresponds to the number of customers in the system, including any currently in service.

- Arrivals occur at rate λ according to a Poisson process and move the process from state i to $i + 1$.

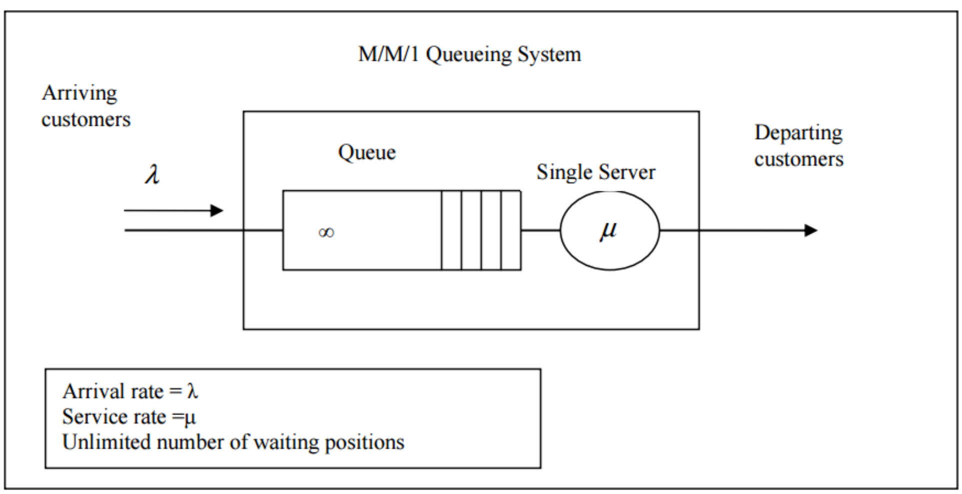
- Service times have an exponential distribution with rate parameter μ in the M/M/1 queue, where 1/μ is the mean service time.
- A single server serves customers one at a time from the front of the queue, according to a first-come, first-served discipline. When the service is complete the customer leaves the queue and the number of customers in the system reduces by one.
- The buffer is of infinite size, so there is no limit on the number of customers it can contain.

  The state space diagram for this chain is as below.

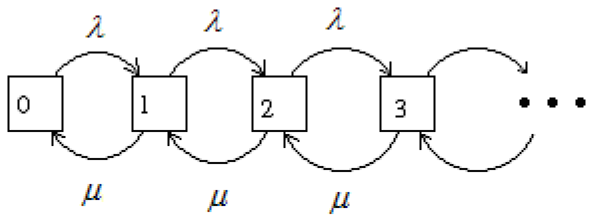
M/M/1 Queueing System is a single-server queueing system with Poisson input, exponential service times and unlimited number of waiting positions.

Thus, an M/M/1 system has the following characteristics:

1. There is a single server with exponential service times and the service rate μ customers
   per time unit
2. Customers arriving according a Poisson process with the arrival rate λ customers per time unit
3. Number of waiting positions = ∞



Rate transition diagram for an M/M/1 Queueing System



5.4 Notations & Description

N	Average number of customers in the system , $N = N_q$ +
N_q	Average number of customers in the queue
N_s	Average number of customers in the service facilities
$\tilde{x}$	Random variable which describes time spent in the service
x	Average service time for a customer, $x = E(\tilde{x})$
$\tilde{w}$	Random variable which describes time spent in the waiting
W	Average waiting time spent in the queue by a customer $W = E(\tilde{w})$
$\tilde{s}$	Random variable which describes time spent in the system by a customer; $\tilde{s} = \tilde{x} + \tilde{w}$
T	Average time spent in the system by a customer $T = E(\tilde{s})$,
λ	Arrival rate
λ_{eff}	The effective arrival rate
μ	Service rate
ρ	$\rho = \frac{\lambda}{\mu}$, offered load (offered traffic)
p_k	Stationary probabilities; p_k is the probability that there are k

5.5 Performance Measures of the M/M/1 Model

Some formulas for M/M/1 Queueing System

$$N = N_q + N_s$$

$$T = W + \bar{x}$$

$$\bar{x} = \frac{1}{\mu}$$

$$p_k = p_0 \cdot \rho^k$$

$$p_0 = 1 - \rho$$

$$N = \frac{\rho}{1-\rho},$$

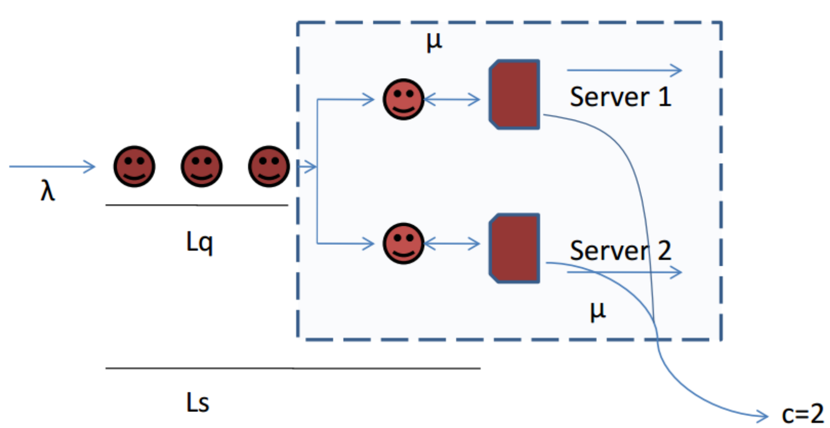
$$T = \frac{1}{\mu - \lambda}$$

$$F_{\tilde{s}}(t) = P(\tilde{s} \le t) = 1 - e^{-(\mu-\lambda)t}$$

5.6 M/M/C Queue

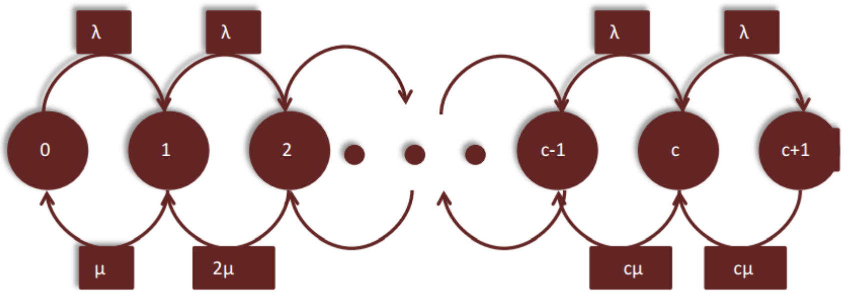
In a M|M|c queue, there are c parallel servers, each serving customers, (c>1). The arrival process and service process follow Poisson distribution. All arriving customers after entering the service system join a single queue. If all c servers are already busy in serving customers, the first customer in the queue will be served by any of the c servers as soon as any server will be free from serving previous customer. The service rate in the case will be (μc). Hence the utilization factor for the M|M|c service system will be

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



In M|M|c queues, the arrival rate remains same as M|M|1 queues but the service rate will depend on the number of servers. The service rate will be nμ for n<=c. As soon as the number of customers exceeds c, the service rate becomes μc.

Transition of states in M/M/C Queueing Model is given as follows



5.7 Performance measures of M/M/C Model

1.The number of customers in the queue, Lq

$$L_q = \left(\frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} \right) P_0$$

2.Customers waiting in the service system will be addition of Wq and service time

$$W = W_q + \frac{1}{\mu}$$

The expected number of customers in the system is

$$L_s = L_q + \frac{\lambda}{\mu}$$

Expected number of customers waiting to be served at any 't' is

$$L_w = \frac{c\mu}{c\mu - \lambda}$$

The average waiting time of an arrival is

$$W_q = \frac{L_q}{\lambda}$$

Average time an arrival spends in the system is $W_s = \frac{L_s}{\lambda}$.

Utilization factor $\rho = \frac{\lambda}{c\mu}$

5.8 M/M/1/N Queue

It is a queueing model with finite queues. In real cases, queues never become infinite, but are limited due to space, time or service operating policy. Such queuing model falls under the category of finite queues.

Examples: (1) Parking of vehicles in a supermarket is restricted to the space of the parking area. (2) Limited seating arrangement in a restaurant.

Finite queue models restrict the number of customers allowed in service system. That means if N represents the maximum number of customers allowed in the service system, then the $(N+1)^{th}$ arrival will depart without being part of the service system or seeking service.

5.9 Performance Measures of M/M/1/N Queueing System

Average number of customers in the system, Ls can be determined using probability of having finite, N, customers in the service system

$$L_s = \sum_{n=0}^{N} n \times P_N$$

The number of customers in queue can determined as given below.

$$L_q = L_s - (1 - P_0)$$

Average waiting time in the queuing system, Ws

$$W_S = \frac{L_q}{\lambda(1 - P_N)} + \frac{1}{\mu}$$

Average waiting time in the queue, Wq

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

5.10 The Pure Birth Process:

If the death rates $\mu_k=0$ for all k=1, 2… we have a **Pure birth process**. If, in addition, we impose the condition of constant birth rates—that is, $\lambda_k=\lambda (k=0,1,2\ldots.)$—then we have the familiar Poisson process. The reduced equation is given by:

$$\frac{dp_o(t)}{dt} = -\lambda p_o(t) \quad \text{k=0} \qquad \text{.......... (1)}$$

$$\frac{dP_k(t)}{dt} = \lambda P_{k-1}(t) - \lambda P_k(t), \quad \text{K>=1 .... (2)}$$

Where we have assumed that the initial state N(0)=0, so that:

$$P_o(0) = 1,$$
$$P_k(0) = 0 \quad \text{for} \quad k \geq 1 \qquad \text{............(3)}$$

One method of solving such differential equation is to use the Laplace transform, which simplifies the system of differential equation to a system of algebraic equations. The Laplace transform of $P_k(t)$ denoted by $\bar{P}_k(s)$ is defined in the usual way, namely:

$$\bar{P}_k(s) = \int_0^\infty e^{-st} P_k(t)\, dt \qquad \text{............ (4)}$$

And the Laplace transform of the derivative dP_k/dt is given by

$$s\bar{P}_k(s) - P_k(0)$$

Now, taking Laplace transforms on both sides of equations above, we get:

$$s\bar{P}_0(s) - P_0(0) = -\lambda \bar{P}_0(s)$$

$$s\bar{P}_k(s) - P_k(0) = \lambda \bar{P}_{k-1}(s) - \lambda \bar{P}_k(s) \quad \text{for} \quad k >= 1$$

Using (3) and rearranging, we get:

$$\bar{P}_0(s) = \frac{1}{s+\lambda}$$

And

$$\bar{P}_k(s) = \frac{\lambda}{s+\lambda}\bar{P}_{k-1}(s)$$

From which we have

$$\bar{P_k}(s) = \frac{\lambda^k}{(s+\lambda)^{k+1}}, \qquad k \geq 0$$

In order to invert this transform, we note that if Y is a (k+1)- stage Erlang random variable with parameter λ, then:

$$L_Y(s) = \bar{f_Y}(s) = \frac{\lambda^{k+1}}{(s+\lambda)^{k+1}}$$

It follows that:

$$P_k(t) = \frac{1}{\lambda} f_Y(t)$$

Therefore:

$$P_k(t) = P[X(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k \geq 0; t \geq 0 \quad \text{..........................(5)}$$

Thus, X (t)is Poisson distributed with parameter λt .

The Poisson process can be generalized to the case where the birth rate λ is varying with time. Such a process is called a **Nonhomogeneous Poisson process.** The generalized version of equation (5) .In this case is given by:

$$P_k(t) = e^{-\Lambda(t)} \frac{[\Lambda(t)]^k}{k!}, k \geq 0$$

$$\Lambda(t) = \int_0^t \lambda(x)dx$$

The non homogeneous Poisson process finds its use in reliability computations when the constant- failure- rate assumptions cannot be tolerated.

Thus, for instance, $\lambda(t) = c\alpha t^{\alpha-1}$ (α>0), then the time to failure of each component is Weibull distributed with parameters c and α , and the pmf of the number of failures N (t) in the interval (0, t) is:

$$P_k(t) = P[N(t) = k] = e^{-ct^{\alpha}} \frac{(ct^{\alpha})^k}{k!}, k \geq 0$$

5.11 Pure Death Processes

Another special case of a birth-death process occurs when the birth rates are all assumed to be zero; that is λ_k=0 for all k. The system starts in some state n > 0 at time t =0 and eventually decays to state 0.Thus, state 0 is an absorbing state. We consider two special cases of interest.

a. Death Process with a Constant Rate

Besides λ_i=0 for all *i*, we have $\mu_i=\mu$ for all i. This implies that the differential difference equation reduce to:

$$\frac{dP_n(t)}{dt} = -\mu P_n(t) \quad \text{for} \quad k = n,$$

$$\frac{dP_k(t)}{dt} = -\mu P_k(t) + \mu P_{k+1}(t) \qquad 1 \leq k \leq n-1,$$

$$\frac{dP_0(t)}{dt} = \mu P_1(t) \qquad k=0$$

Where we have assumed that the initial state N (0) = n, so that:

$$P_n(0) = 1 \quad P_k(0) = 0 \qquad 0 \leq k \leq n-1.$$

Taking Laplace transforms and rearranging, we reduce the above system of equations to:

$$\bar{P}_k(s) = \frac{1}{s+\mu} \quad , \qquad k=n$$

$$\bar{P}_k(s) = \frac{\mu}{s+\mu} \bar{P}_{k+1}(s) \qquad 1 \leq k \leq n-1.$$

$$\bar{P}_k(s) = \frac{\mu}{s}\bar{P}_1(s)$$ K=0

So: $$\bar{P}_k(s) = \frac{1}{\mu}\left(\frac{\mu}{s+\mu}\right)^{n-k+1}$$ 1≤ k≤ n

If Y is an (n-k+1) – stage Erlang random variable with parameters μ, then the Laplace transform of its pdf is known to be $\bar{f}_y(s) = \left[\mu/(s+\mu)\right]^{n+k-1}$.

It follows that:

$$P_k(t) = \frac{1}{\mu} f_Y(t)$$

$$= e^{-\mu t}\frac{(\mu t)^{n-k}}{(n-k)!}$$ 1≤k≤n.

Now, recalling that

$$\sum_{k=0}^{n} P_k(t) = 1$$ We have

$$P_0(t) = 1 - \sum_{k=1}^{n} P_k(t)$$

$$= 1 - \sum_{k=1}^{n} e^{-\mu t}\frac{(\mu t)^{n-k}}{(n-k)!}$$

$$= 1 - \sum_{k=0}^{n-1} e^{-\mu t}\frac{(\mu t)^{k}}{k!}$$

$P_0(t)$ is easily recognized to be the CDF an n-stage Erlang random variable with mean *n/μ*.