Theory notes on Markov chain and Queuing theory

Absorbing State: A state 'a_i' of a Markov chain is called an Absorbing state, if the system remains in the state a_i once it enters there.

The state a_i is absorbing then ith row of the Transition matrix has '1' on the main diagonal and zero's '0' everywhere.

Recurring State: A state 'a_i' of a Markov chain is said to be a recurring state iff if starting from the state a_i, the process eventually returns to the state a_i with probability one.

Transient state: A state 'a_i' of a Markov chain is said to be Transient iff if there is a probability that the process will not return to this state.

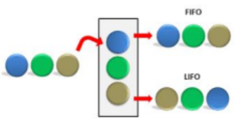
Periodic State: A state 'a_i' of a Markov chain is said to be periodic with period T, if its return to the same state is possible only at instants T, 2T, 3T, ...

QUEUING THEORY

- Queueing theory is the mathematical study of waiting lines, or queues.
- A queueing model is constructed so that queue lengths and waiting time can be predicted.
- It is not an optimizing technique

Characteristics of Queuing Theory

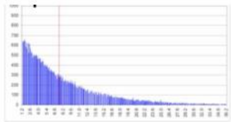
- **The Input or Arrival pattern:** Probability distribution of the number of arrivals per unit of time. It can be
  - Poisson's or Exponential distribution (Markovian) M
  - General Distribution G
- **The Service Pattern:** Probability distribution of the number of customers serviced in one period. It can be
  - Poisson's or Exponential distribution (Markovian) M
  - General Distribution G
- **The Queue discipline:** It can be
  - FIFO/FCFS – First In First Out
  - LIFO/LCFS – Last in First Out
  - SIRO – Selection in Random Order
  - PIR – Priority in selection



- **The system Capacity:** The maximum number of customers in the queuing system. It can be finite or infinite
  - Finite
  - Infinite







Customer's Behaviour

Balking : A customer leaves the queue because the queue is too long & has no time to wait or no sufficient waiting space

Reneging : This occurs when a waiting customer leaves the queue before getting served due to impatience

Jockeying : Customers behaviour of shifting from one queue to another to get immediate service

Priorities : Customers being served irrespective of their arrival time

Some definitions

Customer: A unit coming for service to the service station

Eg: person, Machine, flights etc

Waiting line: A line formed by customers waiting to receive service

Arrival rate: Total no of arrivals by the Total units of time. Denoted by λ

Service rate: Average no of customers being served per unit time. Denoted by μ.

Traffic intensity: Ratio of mean arrival rate to mean service rate. Denoted by ρ

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

Note: $\frac{1}{\mu} \rightarrow$ avg service time or time gap b/w 2 servers

$\frac{1}{\lambda} \rightarrow$ avg arrival time or time gap b/w 2 arrivals.

Idle rate: Phenomenon where server is ready to serve but there is no customer in system. Then server will be idle. The period during which server is idle is called idle time of server

Kendals Notation of Queuing System

Queuing System
$(a/b/c): (d/e)$

- $a \rightarrow$ Input/Inter arrival distribution and
- $b \rightarrow$ Output/Departure/Interservice distribution
    - a & b could be $M \rightarrow$ Markovian(Poisson or negative exponential distributions
- $c \rightarrow$ Service channels or Number of servers
- $d \rightarrow$ Maximum number of customers allowed in the system. It could be finite of infinite.
- $e \rightarrow$ Queue/Service Discipline. It could be
    - FIFO/FCFS – First Come First Out
    - LIFO/LCLS – Last Come First Out
    - SIRO – Service in Random Order
    - PIR –Priority in Selection

Queuing Models

A few important queuing models

Model 1: $(M/M/1: \infty/FIFO)$

Single server infinite capacity

Model 2: $(M/M/1: k/FIFO)$

Single server finite capacity

Model 3: $(M/M/s: \infty/FIFO)$

Multiple server infinite capacity

Terminology

- $\lambda \rightarrow$ Mean arrival rate
- $\mu \rightarrow$ Mean service rate
- $n \rightarrow$ Number of Customers(units) in the system
- $\rho = \boldsymbol{\lambda/\mu} \rightarrow$ Utilization factor(Always <1)
- $P_n(t) \rightarrow$ The probability that exactly n customers(units) in the system at time t
- $P_n \rightarrow$ The steady state probability that exactly n customers(units) in the system
- $L_s \rightarrow$ The expected *number of customers* in the system
- $L_q \rightarrow$ The expected *number of customers* in the queue
- $W_s \rightarrow$ The expected *waiting time of the customer* in the system
- $W_q \rightarrow$ The expected *waiting time of the customer* in the queue
- $L_w \rightarrow$ The expected number of the customer in a non-empty queue
- $f_s(w) \rightarrow$ The p.d.f of waiting time in system
- $f_q(w) \rightarrow$ The p.d.f of waiting time in queue