Queueing Theory

Units of some kind arrive at a service facility to receive some service and depart after receiving it. A queue develops when the service facility cannot cope with the number of units. In other words, arrival rate of units is more than the departure rate of units. This arrangement is queuing system.

Characteristics of a queuing system

(a) Input Process:

Let the successive arrivals to the system occur at times $t_1, t_2, \ldots$; then

$u_r = t_{r+1} - t_r \quad r = 1, 2 \ldots$

are the inter arrival times.

We assume that u_r is i.i.d r.v with a specified P.D.F. $A(u)$

$A(u)$: Inter arrival time distribution

(b) Queue Discipline:

This describes the way units are selected from the queue for service. for example, FCFS, LCFS etc.

(c) Service Mechanism:

Stated by specifying the number of servers c. $[c \geq 1]$ and the P.D.F $B(v)$ of the service times of successive customers which are also i.i.d r.v.

Notation

Queuing Systems are denoted by models

$\circ \,/\, \circ \,/\, \circ \;:\; (\circ \,/\, \circ)$

Input distrib^n or A(u) | Service distrib^n or b(v) | Number of servers (≥1) | Maximum in Queue | Queue discipline

M : Poisson / Exponential distrib^n
D : Deterministic / constant
E_k : Erlang with parameter k
G : General
GI : General Independent

Eg : M/D/S : Poisson arrival, constant service, S servers

How do we analyze a QS?

Through

(I) <u>Queue length</u> : Number of customers waiting before service.

(II) <u>Waiting time</u> : Time spent by a customer in waiting before service

(III) <u>Busy period</u> : A period of time over which the server is continuously busy.

These are studied or analysed through their P.D.F, mean, variance, etc.

_/_/_

The Queue M/M/1

Exponential arrivals, exponential service, 1 server

Number of customers arriving during a time interval (0,t] has a poisson distribⁿ

$$P[N(t)=j] = \frac{e^{-\lambda t}\cdot(\lambda t)^j}{(j)!} \quad j=0,1,2\ldots$$

Interarrival times have exp distrib. with prob density

$$a(x) = \lambda e^{-\lambda x}, \quad x>0$$

Same for service time

$$b(x) = \mu e^{-\mu x}, \quad x>0$$

we have $E[\text{Interarrival time}] = \frac{1}{\lambda} = \frac{1}{\text{arrival rate}}$

$$E[\text{service time}] = \frac{1}{\mu} = \frac{1}{\text{service rate}}$$

Consider, $P_n(t) = P[n \text{ units in QS at time } t]$

This can happen in the following ways:

(i) N units in QS & no service = $P_n(t)\cdot P[\text{No arrival/service}]$

(ii) N-1 units in QS & 1 arrival = $P_{n-1}(t)\cdot P[1 \text{ arrival}]$

(iii) N+1 units in QS & 1 service = $P_{n+1}(t)\cdot P[1 \text{ service}]$

(i) + (ii) + (iii):

$$= P_n(t)[1-(\lambda+\mu)\Delta t] + P_{n-1}(t)\cdot\lambda\Delta t + P_{n+1}(t)\cdot\mu\Delta t + o(\Delta t)$$

Create a DE

$$\lim_{\Delta t \to 0} \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \frac{d}{dt} P_n(t) = -(\lambda+\mu) P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t)$$

$n = 1, 2 \ldots$

consider the boundary case $n=0$ [No service]

$$P_0(t+\Delta t) = P_0(t) * P[\text{No arrival in } (t, t+\Delta t)) + P_1(t) * P[1 \text{ service}]$$

$$P_0(t+\Delta t) = P_0(t)\{1 - \lambda\Delta t\} + P_1(t)\mu\Delta t + o(\Delta t)$$

$$\Rightarrow \frac{dP_0}{dt} = -\lambda P_0(t) + \mu P_1(t)$$

thus, we have 2 differential difference eq^{n}s

for the steady state solution, we assume as $n \to \infty$, $\frac{dP_n}{dt} = 0$

thus

$$0 = -(\lambda+\mu)P_n + \lambda P_{n-1} + \mu P_{n+1} \quad [n \in \mathbb{N}]$$

$$0 = -\lambda P_0 + \mu P_1$$

Solve recursively as in the case of Renewal Process, we get $P_1 = \left(\frac{\lambda}{\mu}\right) P_0$

let $\rho = \frac{\lambda}{\mu}$, this is known as traffic intensity

$$P_n = \rho^n P_0$$

as this probability is normalized, we have

$$\sum_{n=0}^{\infty} P_n = 1$$

$P_0 + \rho P_0 + \rho^2 P_0 \quad \ldots \quad + \rho^n P_0 = 1$

$P_0 (1 + \rho + \rho^2 + \rho^3 \ldots) = 1$

this is convergent only when $\rho < 1$

thus $P_0 \cdot \frac{1}{1-\rho} = 1$

$$P_0 = 1 - \rho$$

we have $P_n = \rho^n \cdot P_0$

$$P_n = (1-\rho) \cdot \rho^n$$

which is distributed geometrically

Calculate mean /exp value = L

$$L = \sum_{n=0}^{\infty} n P_n$$

$$= \sum \{ n (1-\rho) \rho^n \}$$

$$= (1-\rho)(\rho + 2\rho^2 + 3\rho^3 \ldots)$$

consider $S = \rho + 2\rho^2 + 3\rho^3 \ldots \quad n\rho^n + (n+1)\rho^{n+1} \ldots \infty$

$\rho S = \quad \rho^2 + 2\rho^3 \quad \ldots \quad + n\rho^{n+1} \ldots \infty$

$S(1-\rho) = \frac{\rho}{1-\rho} \Rightarrow S = \frac{\rho}{(1-\rho)^2}$

$$\boxed{L = \frac{\rho (1-\rho)}{(1-\rho)^2} = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}}$$

__/__/__

(i) The mean num in queue (waiting for service) given by L_q is:

$$L_q = \sum_{n=1}^{\infty} (n-1) P_n$$

$$= \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

(ii) Waiting Time

Assume the queue discipline is FIFO

Let

W_q = Waiting time in queue

W = Waiting time in system

for W_q:

customers found in queue at any time, must have arrived before the first person in the queue was served.

Thus, mean number of customers in Q is mean arrivals during mean waiting time

i.e. $L = \lambda W_q$

or $W_q = \frac{L}{\lambda} = \frac{\lambda}{\mu\{\mu-\lambda\}}$

and similarily, $L = \lambda W$

$W = \frac{L}{\lambda} = \frac{1}{(\mu-\lambda)}$

next, we find distribution of W (W_s).

Suppose there were n units in the Q and then one arrival occured for W_s in $(w, w+\Delta w)$ all the $(n+1)$ customers including the new arrival should complete service with $w+\Delta w$.

$$f(w)dw = P\{w < W_s < w + dw\}$$

$$= \sum_{n=0}^{\infty} P_n \cdot P(n \text{ service in } w) \cdot P(1 \text{ service in } dw)$$

$$= \sum (1-\rho)\rho^n \cdot \frac{(\mu w)^n}{n!} e^{-\mu w} \cdot \mu dw$$

[as service is poisson distrib]

$$= (1-\rho)\mu e^{-\mu w} dw \sum_{n=0}^{\infty} (\mu\rho w)^n / n!$$

$$= (1-\rho) \cdot \mu e^{-\mu w} \cdot dw \cdot e^{\mu\rho w} \quad \left[\sum_{i=1}^{n} \frac{x^i}{i!} = e^n\right]$$

$$W_s \quad = (1-\rho)\mu e^{-(1-\rho)\mu w} dw$$

thus, time spent in system is distributed exponentially with mean $= \frac{1}{(1-\rho)\mu} = \frac{1}{\mu - \lambda}$

[because mean of $\lambda e^{-\lambda x}$ is $1/\lambda$]

Similarily for waiting time in Queue,

$$P(W_q = 0) = P_0 = 1-\rho$$

and $f_q(w)dw = P\{w < W_q < w + dw\}$

$$f_q(w) = \rho\mu(1-\rho)e^{-(1-\rho)\mu w} \quad \text{(similarily from above)}$$

$$\text{Mean} = 0 \cdot (1-\rho) + \lim_{h \to 0^+} \int_h^{\infty} w \cdot \rho\mu(1-\rho) e^{-(1-\rho)\mu w} dw$$

__/__/__

$= \frac{\rho}{\mu(1-\rho)} = \frac{\lambda}{\mu(\mu-\lambda)}$ which matches our previous result.

(iii) Busy period

Entire server time = Busy period + Idle period

as $t \to \infty$, we can assume number of busy & idle periods are equal.

Idle period is the time from where the server becomes idle until the next arrival. as we are studying M/M/1 model, the inter arrival times follow poisson. Hence, the mean duration of an idle period is $1/\lambda$.
Now, probability that server is idle
$= P[\text{No units in } Q]$
$= P_0 = 1-\rho$.

Over time T, server is expected to be idle for $TP_0 = (1-\rho) \cdot T$ duration.

Num of periods? Multiply by λ

$= T\lambda(1-\rho)$ total idle periods

Server's busy time = ~~$T\rho$~~ $T - (T(1-\rho)) = T\rho$
Num of periods = same as idle (assumption)

Mean busy period $= \frac{T\rho}{T\lambda(1-\rho)} = \frac{1}{\mu-\lambda}$

Mean customers served $= \underset{\text{service rate}}{\mu} \times$ Mean Busy P $= \frac{\mu}{\mu-\lambda}$

—/—/—

State dependent parameters

Here, we consider λ and μ to be dependent upon the state of n. Thus, we denote as λ_n & μ_n respectively.

for steady state

$$p_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} P_0 \quad \text{—(1)}$$

normalizing this probability

$$P_0 \left[1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \cdots + \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \right] = 1 \quad \text{—(2)}$$

for solution to exist, the inf series should be convergent. Let S is the sum of the series

$$P_0 = S^{-1} \quad (3) \quad \text{and} \quad P_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1} * S^{-1}}{\mu_1 \mu_2 \cdots \mu_n} \quad (4)$$

M/M/1/N

No more than N units are allowed in the queue.

$$\lambda_n = \begin{cases} \lambda & (0 \le n < N) \\ 0 & (n \ge N) \end{cases} \quad \text{—(5)}$$

$$\mu_n = \mu \quad (n = 1, 2, \ldots)$$

using (2)

$$S = 1 + \rho + \rho^2 \cdots + \rho^N$$

$$= \frac{1 - \rho^{N+1}}{1 - \rho} \quad \text{—(6)} \quad \& \ \rho = \frac{\lambda}{\mu}$$

_/_/_

Since ρ is finite, we conclude that the steady state exists in all cases, i.e., for all values of ρ. Substituting for S in (4), we have

$$P_n = \frac{(1-\rho)\rho^n}{(1-\rho^{N+1})}, \rho \neq 1, n = 0(1)N \quad \text{---} (7)$$

when $\rho = 1$, $S = N+1$ ($\lambda = \mu$ case)

$$P_n = \frac{1}{N+1} \quad \text{---} (8)$$

i) Mean Number in system

$$L = \sum_{n=0}^{N} n P_n$$

$$(\rho \neq 1) = \rho[1-(N+1)\rho^N + N\rho^{N+1}] / [(1-\rho)(1-\rho^{N+1})]$$

$$\text{if } \rho \to 1 \quad = \frac{N}{2}$$

ii) Mean number in ~~system~~ queue

$$L_q = \sum_{n=1}^{N} (n-1)P_N$$

$$(\rho \neq 1) = \rho^2[1 - N\rho^{N-1} + (N-1)\rho^N] / [(1-\rho)(1-\rho^{N+1})]$$

$$(\rho = 1) = \frac{N(N-1)}{2(N+1)}$$

The QS $M/M/C/\infty$

There are c servers, $a(x) \sim$ Poisson(λ) & $b(x)$ is poisson(μ)

$$\lambda_n = \lambda \quad \forall n$$

$$\mu_n = n\mu \quad \forall n < c$$

$$= c\mu \quad n \geq c$$

_/_/_

When n servers are busy, prob. of a departure during $(t, t+\Delta t)$ is the prob. that any one of the n servers completes service, which is $n\mu\Delta t$

Normalizing again

$$S = 1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{2!\mu^2} + \dots + \frac{\lambda^{c-1}}{(c-1)!\mu^{c-1}} + \frac{\lambda^c}{c!\mu^c}\left\{1 + \frac{\lambda}{c\mu} + \frac{\lambda^2}{(c\mu)^2}\right\}$$

$$p = \frac{\lambda}{c\mu} \Rightarrow S = \sum_{j=0}^{c-1} \frac{(cp)^j}{j!} + \left\{\frac{(cp)^c}{c!}\right\}(1-p)^{-1}$$

for $p<1$ [convergence]

$$P_n = P_0\, c^n p^n / n! \quad \forall\, n < c$$

$$= P_0\, c^c p^c / c! \quad \forall\, n \geq c$$

where $P_0 = 1/S$

$$Pr[\text{Wait}] = P[n \geq c] = P_0 (c^c/c!) \sum_{n=c}^{\infty} p^n = \frac{P_0 (cp)^c}{c!(1-p)}$$

$$= P_0/(1-p)$$

Expected people in Q : $L_q = \sum_{n=c}^{\infty} (n-c) P_n$

$$= \frac{\lambda\mu(\lambda/\mu)^c P_0}{(c-1)!\,(c\mu-\lambda)^2}$$

Expected customers Served : L_s

$$L_s = \sum_{n=0}^{c-1} n P_n + \sum_{n=c}^{\infty} c P_n$$

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

_/_/_

Mean Number in the sys:

$$L = L_q + L_s$$
$$= L_q + (\lambda/\mu)$$

Mean WT in Queue

$$W_q = L_q/\lambda = \frac{\mu(\lambda/\mu)^c P_o}{\{(c-1)!\,(c\mu-\lambda)^2\}}$$

Mean Time Spent in Q

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

Queiveing System M/M/∞

In this system, every ~~server~~ unit is able to find a server. Eg: are self service systems or e-commerce website.
In such system, no queuing is involved

$$\lambda_n = \lambda \quad \forall n$$
$$\mu_n = n\mu \quad \forall n = 0, 1, 2 \ldots$$

Therefore, sum $S = 1 + \left(\frac{\lambda}{\mu}\right) + \frac{(\lambda/\mu)^2}{2!} \ldots$

$$= e^{\lambda/\mu}$$

This series is convergent & therefore, the steady state solution exists for all values of λ/μ. We have

$$p_n = \frac{(\lambda/\mu)^n e^{-\lambda/\mu}}{n!}$$

So that the number in the system has a poisson distribution with mean (λ/μ).
Thus $L = \lambda/\mu$

and $$W = \frac{L}{\lambda} = \frac{1}{\mu}$$

Queue with Balking

Balking: If the queue is too long, the customer may not join it.

Consider the M/M/1 model with λ and μ such that when there are n units in the system, a customer joins the queue with the probability $1/(n+1)$. So, we take

$$\lambda_n = \frac{\lambda}{(n+1)}, \quad n = 0, 1, 2 \ldots$$

$$\mu_n = \mu \quad \forall n$$

then $$S = 1 + \left(\frac{\lambda}{\mu}\right) + \frac{(\lambda/\mu)^2}{2!} + \ldots = e^{\lambda/\mu}$$

Same as before. The number in the system also follows a poisson distribution. It follows that the mean number in the system is λ/μ as before. But in the present case, queuing is allowed.

Mean:
$$L_q = \sum_{n=1}^{\infty} (n-1) P_n$$
$$= \left(\frac{\lambda}{\mu}\right) + e^{-\lambda/\mu} - 1$$

M/M/C/C Loss Model

Here, queuing is not allowed as maximum number in the system is c with c channels.

Example - Telephone exchange with c lines.

λ = mean arrival

μ = mean service

$$\lambda_n = \lambda \quad \text{for } n<c$$

S is the sum

$$S = 1 + (\lambda/\mu) + \frac{(\lambda/\mu)^2}{2!} \cdots + \frac{(\lambda/\mu)^c}{c!}$$

steady state solⁿ exists for all values of λ/μ and

$$p_n = \frac{(\lambda/\mu)^n}{n!} S^{-1}$$

the prob. that all channels are busy & that incoming calls are lost is:

$$p_c = \frac{(\lambda/\mu)^c / c!}{1 + \left(\frac{\lambda}{\mu}\right) + \left(\frac{(\lambda/\mu)^2}{2!}\right) + \cdots + \left(\frac{\lambda}{\mu}\right)^c \frac{1}{c!}}$$

this is Erlang's Loss Formula