Introduction to Queueing Theory

Supervisor: Dr. Shaista Rais
Department of Computer Science
University of Karachi

Queueing Systems

- **A queuing system consists of “customers” arriving at random times at some facility where they receive service of some kind and then depart.**

OR can be defined as

- model processes in which customers arrive.
- wait their turn for service.
- are serviced and then leave.

Examples

- supermarket checkouts stands.
- world series ticket booths.
- doctors waiting rooms etc..

Five components of a Queueing system:

- 1. Interarrival-time probability density function (pdf)
- 2. service-time pdf
- 3. Number of servers
- 4. queueing discipline
- 5. size of queue.

ASSUME

- an infinite number of customers (i.e. long queue does not reduce customer number).

Assumption is bad in :

- a time-sharing model.
- with finite number of customers.
- if half wait for response, input rate will be reduced.

Interarrival-time pdf

- record elapsed time since previous arrival.
- list the histogram of inter-arrival times (i.e. 10 0.1 sec, 20 0.2 sec ...).
- This is a pdf character.

Service time

- how long in the server?
- i.e. one customer has a shopping cart full the other a box of cookies.
- Need a PDF to analyze this.

Number of servers

- banks have multiserver queueing systems.
- food stores have a collection of independent single-server queues.

Queueing discipline

- order of customer process-ing.
- i.e. supermarkets are first-come-first served.
- Hospital emergency rooms use sickest first.

Finite Length Queues

- Some queues have finite length: when full customers are rejected.

ASSUME

- infinite-buffer.
- single-server system with first-come.
- first-served queues.

A/B/m notation

- A=interarrival-time pdf
- B=service-time pdf
- m=number of servers.

A,B are chosen from the set:

- M=exponential pdf (M stands for Markov)
- D= all customers have the same value (D is for deterministic)
- G=general (i.e. arbitrary pdf)

Analysibility

- M/M/1 is known.
- G/G/m is not.

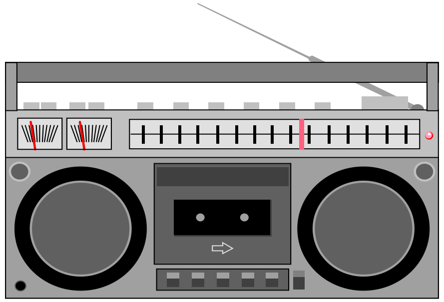
M/M/1 system

- For M/M/1 the probability of exactly n customers arriving during an interval of length t is given by the Poisson law.

Poisson’s Law

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \qquad \textbf{(1)}$$

Poisson's Law in Physics



- radio active decay
  - P[k alpha particles in t seconds]
  - λ = avg # of prtcls per second

$$\lambda$$

Poisson’s Law in Operations Research

- planning switchboard sizes
  - P[k calls in t seconds]
  - =avg number of calls per sec

$$\lambda$$

Poisson's Law in Biology

- water pollution monitoring
  - P[k coliform bacteria in 1000 CCs]
  - =avg # of coliform bacteria per cc

λ

Poisson’s Law in Transportation

- planning size of highway tolls
  - P[k autos in t minutes]
  - =avg# of autos per minute

λ

Poisson's Law in Optics

- in designing an optical recvr
  - P[k photons per sec over the surface of area A]
  - =avg# of photons per second per unit area

$$\lambda$$

Poisson's Law in Communications

- in designing a fiber optic xmit-rcvr link
  - P[k photoelectrons generated at the rcvr in one second]
  - =avg # of photoelectrons per sec.

$$\lambda$$

- Rate λ parameter

- λ =event per unit interval (time distance volume…)

Analysis

- Depend on the condition:

$$\lambda = \text{ interarrival rate } = 10 \text{ cust. per min}$$

$$n = \text{ the number of customers } = 100$$

- we should get 100 custs in 10 minutes (max prob).

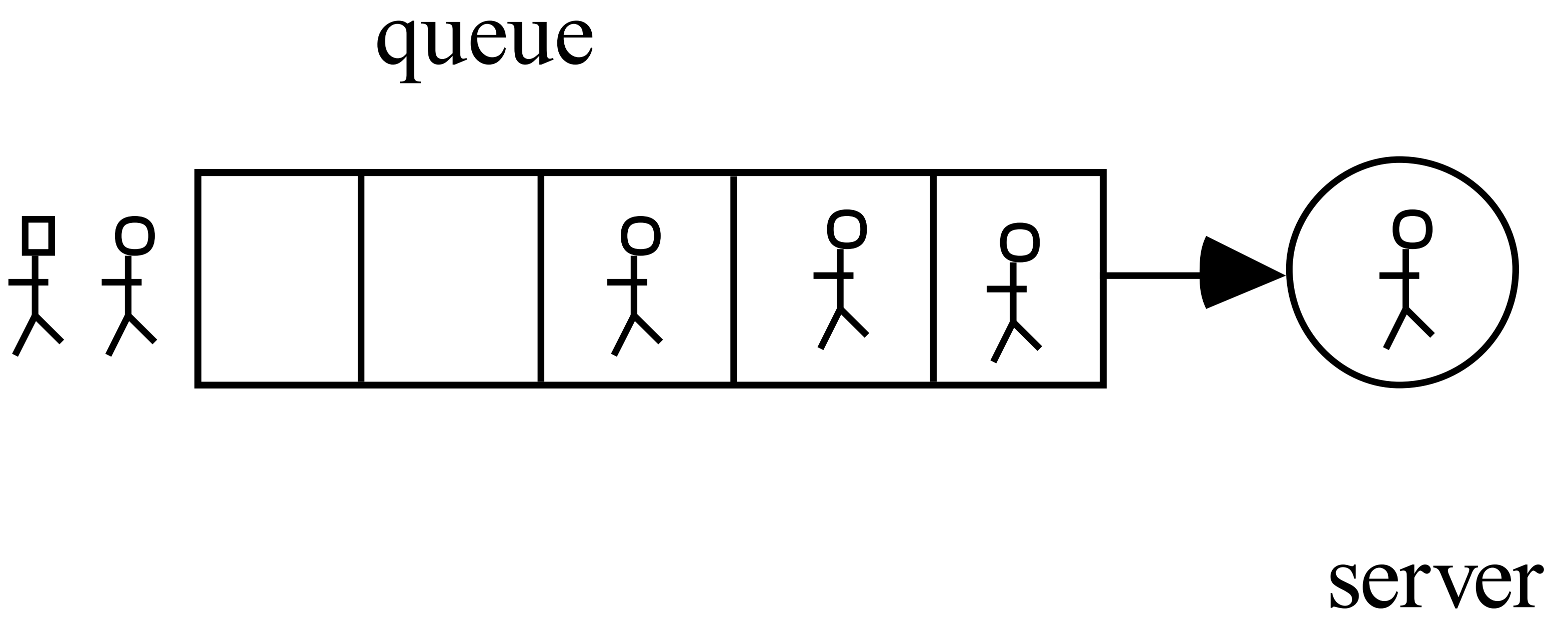
To obtain numbers with a Poisson pdf, you can write a program:

Acceptance Rejection Method

Prove:

- Poisson arrivals gene-rate an exponential interarrival pdf.

The M/M/1 queue in equilibrium



State of the system:

- There are 4 people in the system.
- 3 in the queue.
- 1 in the server.

Memory of M/M/1:

- The amount of time the person in the server has already spent being served is independent of the probability of the remaining service time.

Memoryless

- M/M/1 queues are memoryless (a popular item with queueing theorists, and a feature unique to exponential pdfs).

$$P_k = \text{equilibrium prob}\ \cdot$$

$$\text{that there are k in system}$$

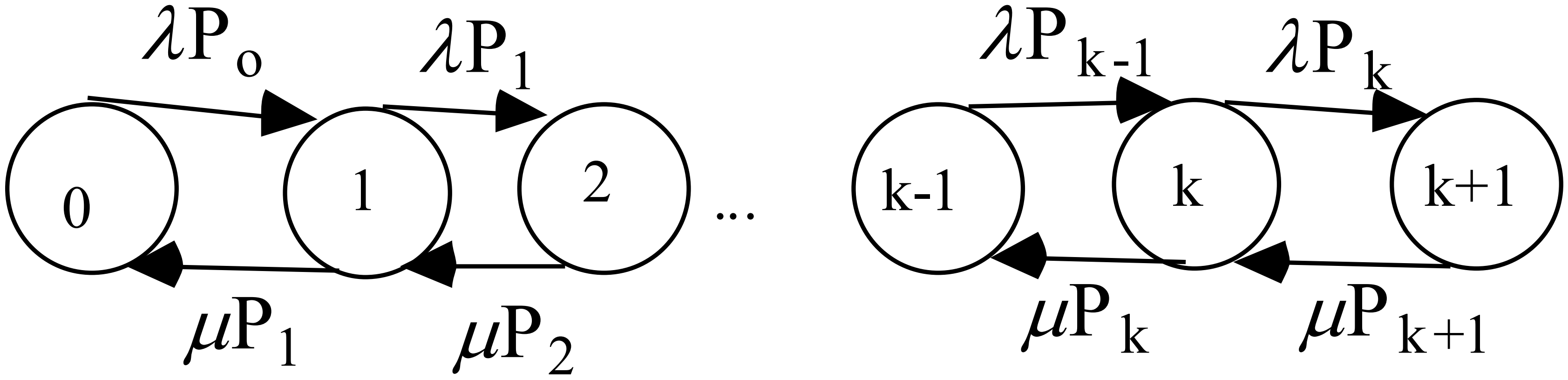
Birth-death system

- In a birth-death system once serviced a customer moves to the next state.
- This is like a nondeterminis-tic finite-state machine.

State-transition Diagram

- The following state-transition diagram is called a Markov chain model.
- Directed branches represent transitions between the states.
- Exponential pdf parameters appear on the branch label.

Single-server queueing system



Symbles:

$$\lambda = \text{mean arrival rate} \ \ (\text{cust.}/\text{sec})$$

$\lambda P_0 =$ **mean number of transitions/ sec from state 0 to 1**

$$\mu = \text{mean service rate} \ \ (\text{cust.}/\text{sec})$$

$\mu P_1 =$ **mean number of transitions/ sec from state 1 to 0**

States

- State 0 = system empty
- State 1 = cust. in server
- State 2 = cust in server, 1 cust in queue etc...

Probalility of Given State

- Prob. of a given state is invariant if system is in equilibrium.
- The prob. of k cust’s in system is constant.

Similar to AC

- This is like AC current entering a node
- is called detailed balancing
- the number leaving a node must equal the number entering

Derivation

3 $$\lambda P_0 = \mu P_1$$

3a $$P_1 = \frac{\lambda P_0}{\mu}$$

4 $$\lambda P_1 = \mu P_2$$

4a $$P_2 = \frac{\lambda P_1}{\mu}$$

by 3a

4 $$P_2 = \frac{\lambda \frac{\lambda P_0}{\mu}}{\mu} = P_2 = \frac{\lambda^2 P_0}{\mu^2}$$

since

5 $$\lambda P_k = \mu P_{k+1}$$

then:

$$\mathbf{6} \qquad \mathrm{P}_{\mathrm{k}} = \frac{\lambda^{k}\mathrm{P}_{0}}{\mu^{k}} = \rho^{k} P_{0}$$

where $\rho = \dfrac{\lambda}{\mu}$ **= traffic intensity < 1**

since all prob. sum to one

6a $$\sum_{k=0}^{\infty} \rho^k P_0 = 1 = P_0 \sum_{k=0}^{\infty} \rho^k = 1$$

Note: the sum of a geometric series is

7 $$\sum_{k=0}^{\infty} \rho^k = \frac{1}{1-\rho}$$

$$\sum_{k=0}^{\infty} \rho^k = \frac{1}{1-\rho}$$

- Suppose that it is right, cross multiply and simplify:

$$\sum_{k=0}^{\infty} \rho^k - \rho \sum_{k=0}^{\infty} \rho^k = 1$$

So $$\sum_{k=0}^{\infty} \rho^k - \sum_{k=1}^{\infty} \rho^k = \rho^0 = 1$$

Q.E.D.

subst 7 into 6a

6a $$P_0 \sum_{k=0}^{\infty} \rho^k = 1$$

7a $$\frac{P_0}{1-\rho} = 1$$ **and**

7b $$P_0 = 1 - \rho$$

=prob server is empty

subst into

6 $$\mathrm{P}_{\mathrm{k}} = \frac{\lambda^k \mathrm{P}_0}{\mu^k} = \rho^k P_0$$

yields:

8 $$\mathrm{P}_{\mathrm{k}} = (1-\rho)\rho^k$$

Mean value:

- let N=mean number of cust’s in the system
- To compute the average (mean) value use:

$$E[k] = \sum_{k=0}^{\infty} kP_k \qquad \textbf{8a}$$

Subst (8) into (8a)

8 $$\mathrm{P_k} = (1-\rho)\rho^k$$

8a $$E[k] = \sum_{k=0}^{\infty} kP_k$$

we obtain

8b $$E[k] = \sum_{k=0}^{\infty} k(1-\rho)\rho^k = (1-\rho)\sum_{k=0}^{\infty} k\rho^k$$

differentiate (7) wrt k

7 $$\sum_{k=0}^{\infty} \rho^k = \frac{1}{1-\rho}$$

we get

8c $$D_k \sum_{k=0}^{\infty} \rho^k = D_k \frac{1}{1-\rho} = \sum_{k=0}^{\infty} k\rho^{k-1} = \frac{1}{(1-\rho)^2}$$

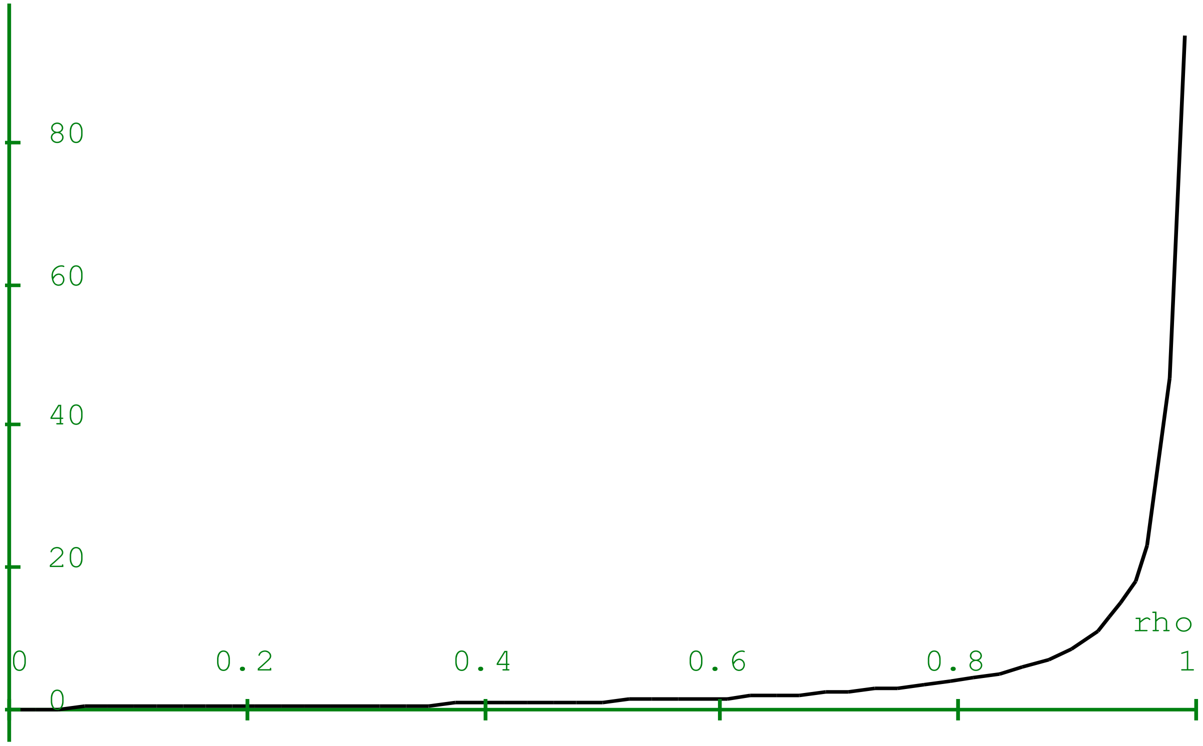
multiply both sides of (8c) by

ρ

8d $$\sum_{k=0}^{\infty} k\rho^k = \frac{\rho}{(1-\rho)^2}$$

9 $$E[k] = N = (1-\rho)\frac{\rho}{(1-\rho)^2} = \frac{\rho}{(1-\rho)}$$

Relationship of , N ρ



as r approaches 1, N grows quickly.

T and λ

- T=mean interval between cust. arrival and departure, including service.

$$\lambda = \text{mean arrival rate} \ \ (\text{cust./sec})$$

Little's result:

- In 1961 D.C. Little gave us Little's result:

$$T = \frac{N}{\lambda} = \frac{\rho/\lambda}{1-\rho} = \frac{1/\mu}{1-\rho} = \frac{1}{\mu-\lambda} \qquad \mathbf{10}$$

For example:

- A public bird bath has a mean arrival rate of 3 birds/min in Poisson distribution.
- Bath-time is exponentially distributed, the mean bath time being 10 sec/bird.

Compute how long a bird waits in the Queue (on average):

$$\lambda = 0.05 \text{ cust / sec} = 3 \text{ birds / min } * \ 1 \text{ min } / \ 60 \text{ sec}$$

= mean arrival rate

$$\mu = 0.1 \text{ bird / sec} = \frac{1 \text{ bird}}{10 \text{ sec}}$$

= mean service rate

Result:

- So the mean service-time is 10 seconds/bird =(1/ service rate)

$$T = \frac{1}{\mu - \lambda} = \frac{1}{0.1 - 0.05} = 20 \textbf{ sec}$$

for wait + service

Mean Queueing Time

- The mean queueing time is the waiting time in the system minus the time being served, 20-10=10 seconds.

M/G/1 Queueing System

- Tannenbaum says that the mean number of customers in the system for an M/G/1 queueing system is:

11 $$N = \rho + \rho^2 \frac{1 + C_b^2}{2(1-\rho)}$$

This is known as the Pollaczek-Khinchine equation.

What is c_b

$$c_b = \frac{\text{standard deviation}}{\text{mean}}$$

of the service time.

Note:

- M/G/1 means that it is valid for any service-time distribution.
- For identical service time means, the large standard deviation will give a longer service time.

Customer Behaviour in a Queue

Various customers while in queue behave differently. Their behavior pattern can fall in one of the following categories.

1. Balking

A customer may not like to wait in a que due to lack of space or otherwise. They do not join the queue at their correct position and attempt to jump the queue and reach the service centre by passing other ahead of them. This is known as balking.

2. Reneging

A customer may leave the queue due to impatience. This is called reneging.

3. Collusion

Some customers may collaborate and only one of them may join the queue. As at the cinema ticket window one person may join the queue and purchase tickets for his friends.

4. Jockeying

If there are more than one queues then one customer may leave one queue and join the other. This occurs generally in the super market or shopping malls.