Department of Industrial Engineering and
Engineering Management
University of Sharjah

0405325: Stochastic Simulation

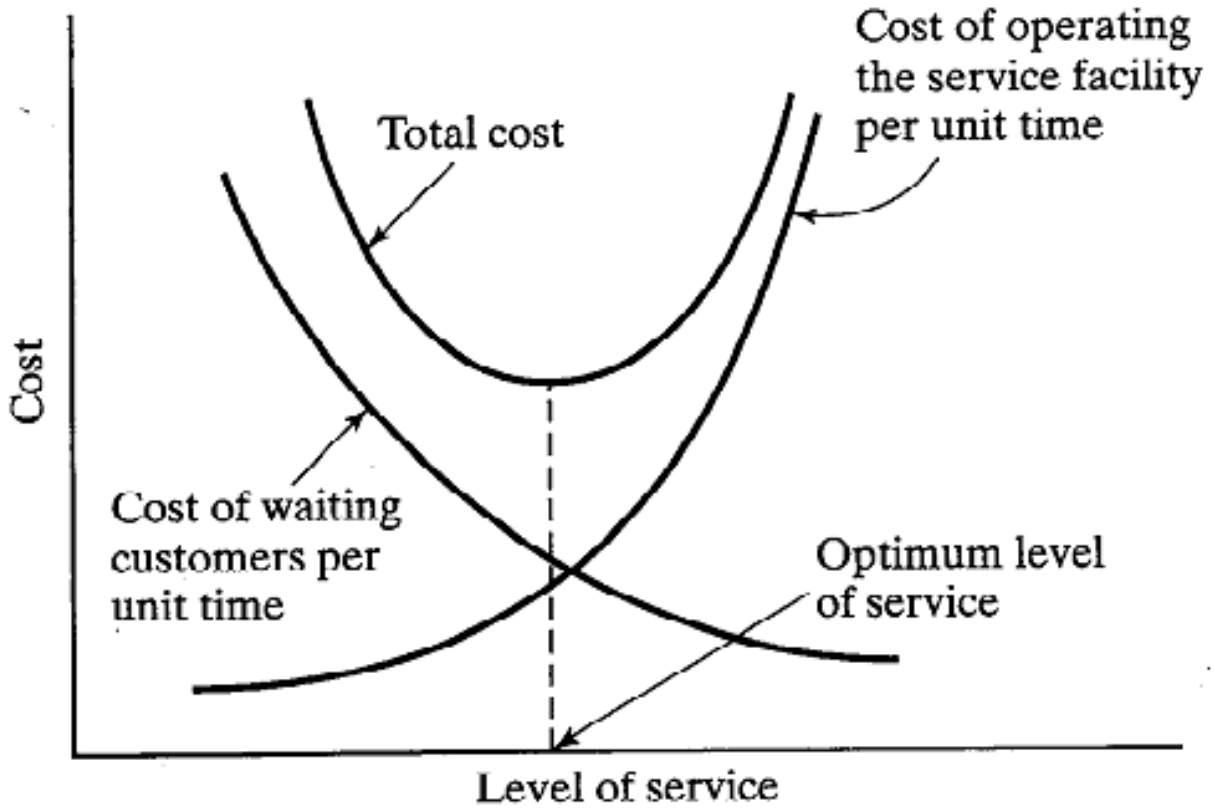
Lecture 9: Introduction to Queuing Theory

18.1 Why study queues?

- **Question:** Is queuing an optimization techniques?
- **Answer:** No, it just determines the measures of performance of waiting lines such as average waiting time in a queue.
- **Example of queues:**
  - Waiting to eat in a restaurant
  - Grocery store
  - Post offices
  - Jobs waiting to be processed on a machine
  - Planes waiting to land at an airport
  - (Software) tasks waiting to be processed by the microprocessor,…

18.1 Why study queues?

- Queuing analysis and cost optimization models



Difficulty:
Estimation of
waiting cost

18.2 Elements of a queuing model

- Customer and server

- Customers generated from a **source**
- On arrival to a service facility they can be served immediately or wait in a **queue** if the facility is **busy**
- When the facility completes a service, it **pulls** a customer, if any, from the queue
- If the queue is empty, the facility (server) becomes **idle** until a new customer arrives

18.2 Elements of a queuing model

When analyzing queues, we consider:

- **Interarrival time** between successive customers
- **Service time** per customer
  - They can be probabilistic (in general) or deterministic (ex. Arrival of job applicants for an interview).
- **Queue size**: finite or infinite
- **Queue discipline:** order in which the customers are pulled from the queue
  - First Come, First Served (FCFS)
  - Last Come, First Served (LCFS), ex. In some computer applications
  - Service In Random Order (SIRO), ex. A multi-media play list.
  - **Order of priority**. Customers can be given priority orders

18.3 Role of Exponential Distribution

- Customers arrive in a totally random fashion, that is:
  - the occurrence of an event is not influenced by the length of time that has elapsed since the occurrence of the last event.
- Random interarrival and service times are described quantitatively by the **exponential distribution**

pdf $$f(t) = \lambda e^{-\lambda t}, t > 0$$

Mean $$E\{t\} = \frac{1}{\lambda}$$

CDF $$P\{t \leq T\} = \int_0^T \lambda e^{-\lambda t}\, dt$$

$$= 1 - e^{-\lambda T}$$

18.5. Generalized Poisson Queuing Model

- Transient behavior of a queuing system

- Steady state behavior of a queuing system

18.5. Generalized Poisson Queuing Model

- Arrival and departure rates are **state dependent**

- Define:

n = Number of customers in the system (in-queue plus in-service)

λ_n = Arrival rate given n customers in the system

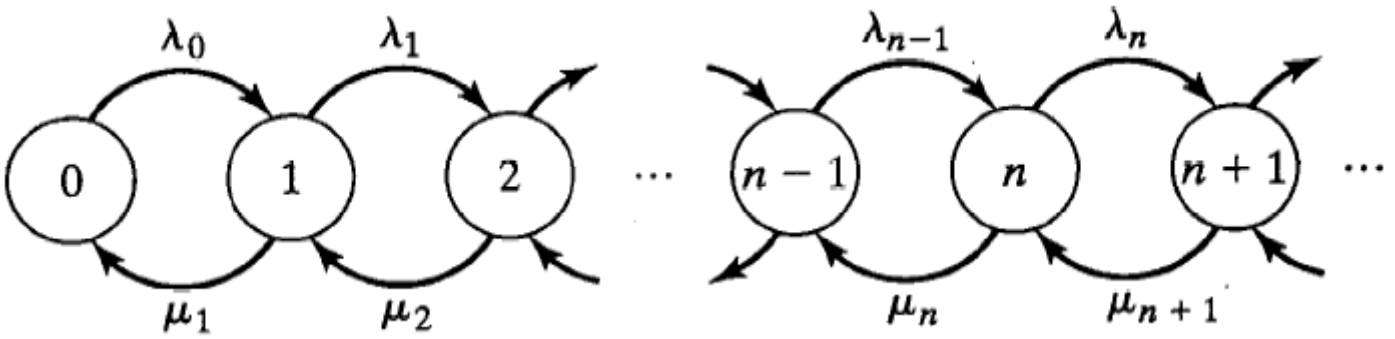
μ_n = Departure rate given n customers in the system

p_n = The steady state probability of n customers in the system

Objective 1: Derive p_n as a function of λ_n and μ_n

Objective 2: Derive measures of performance

18.5. Generalized Poisson Queuing Model

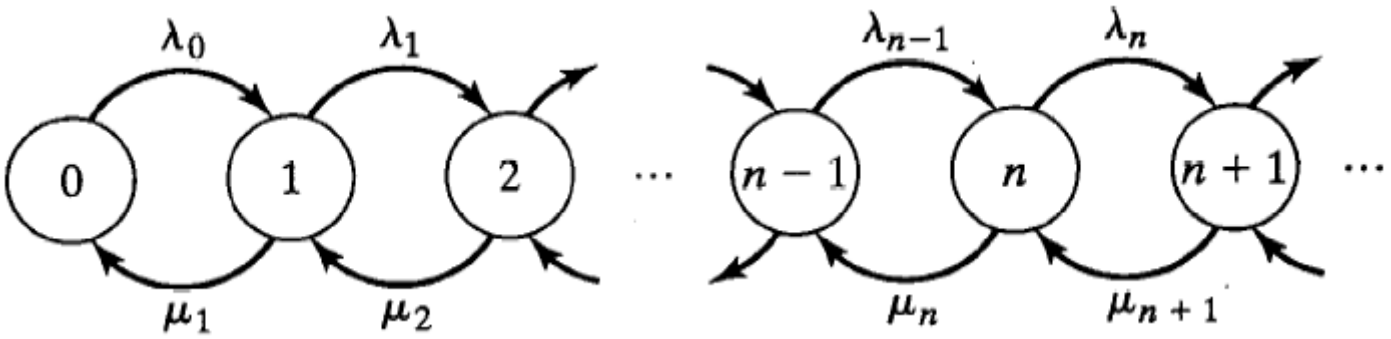


Transition-rate diagram

$$\begin{pmatrix}\text{Expected rate of}\\ \text{flow into state } n\end{pmatrix} = \lambda_{n-1}p_{n-1}+\mu_{n+1}p_{n+1}$$

$$\begin{pmatrix}\text{Expected rate of}\\ \text{flow out of state } n\end{pmatrix} = (\lambda_n+\mu_n)p_n$$

18.5. Generalized Poisson Queuing Model



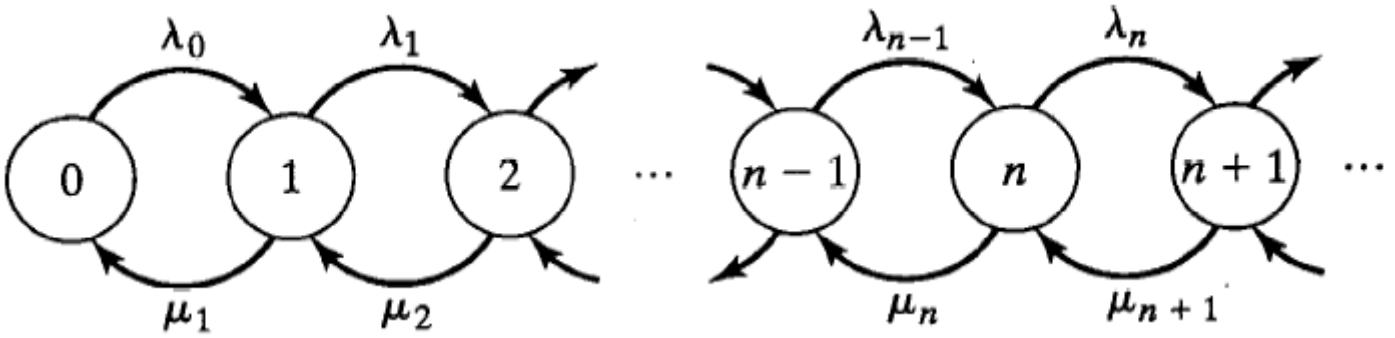
Transition-rate diagram

- Steady state conditions → expected flow rates into and out of state n should be equal → balance equation:

$$\lambda_{n-1}p_{n-1}+\mu_{n+1}p_{n+1} =(\lambda_n+\mu_n)p_n, \text{ for } n = 1,2,\ldots$$

$$\lambda_0 p_0=\mu_1 p_1, \text{ for } n = 0$$

18.5. Generalized Poisson Queuing Model



Transition-rate diagram

- Solve the balance equations recursively (in function of p_0):
  - For n=0 $$p_1 = \left(\frac{\lambda_0}{\mu_1}\right) p_0$$
  - For n = 1 $$\lambda_0 p_0 + \mu_2 p_2 = (\lambda_1 + \mu_1) p_1 \rightarrow p_2 = \left(\frac{\lambda_1 \lambda_0}{\mu_2 \mu_1}\right) p_0$$

18.5. Generalized Poisson Queuing Model

$$p_n = \left(\frac{\lambda_{n-1}\lambda_{n-2} \cdots \lambda_0}{\mu_n \mu_{n-1} \cdots \mu_1} \right) p_0, n = 1, 2, \ldots$$

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

Example

B&K Groceries operates with three check-out counters. The manager uses the following schedule to determine the number of counters in operation, depending on the number of customers in store:

No. of customers in store	No. of counters in operation
1 to 3	1
4 to 6	2
More than 6	3

Customers arrive in the counters area according to a Poisson distribution with a mean rate of 10 customers per hour. The average check-out time per customer is exponential with mean 12 minutes. Determine the steady-state probability p_n of n customers in the check-out area.

- Determine the probability that only one counter is open
- Determine the average number of idle counters.

Example

B&K Groceries operates with three check-out counters. The manager uses the following schedule to determine the number of counters in operation, depending on the number of customers in store:

No. of customers in store	No. of counters in operation
1 to 3	1
4 to 6	2
More than 6	3

Customers arrive in the counters area according to a Poisson distribution with a mean rate of 10 customers per hour. The average check-out time per customer is exponential with mean 12 minutes. Determine the steady-state probability p_n of n customers in the check-out area.

$\lambda_n = \lambda = 10$ customers/h

$\mu_n =$ 60/12 = 5 cust/h, n = 0, 1, 2, 3
2 * 5 = 10 cust/h, n = 4, 5, 6
3 * 5 = 15 cust/h, n = 7, 8, …

$$p_1 = \left(\frac{\lambda_0}{\mu_1}\right)p_0 = \frac{10}{5}p_0 \qquad p_3 = (\frac{10}{5})^3 p_0 = 8p_0 \qquad p_5 = (\frac{10}{5})^3(\frac{10}{10})^2 p_0 = 8p_0$$

$$p_2 = \left(\frac{\lambda_1}{\mu_2}\frac{\lambda_0}{\mu_1}\right)p_0 = (\frac{10}{5})^2 p_0 \qquad p_4 = (\frac{10}{5})^3(\frac{10}{10})p_0 = 8p_0 \qquad p_6 = (\frac{10}{5})^3(\frac{10}{10})^3 p_0 = 8p_0$$

$$p_{n\geq 7} = (\frac{10}{5})^3(\frac{10}{10})^3(\frac{10}{15})^{n-6} p_0 = 8(\frac{2}{3})^{n-6} p_0$$

Example

B&K Groceries operates with three check-out counters. The manager uses the following schedule to determine the number of counters in operation, depending on the number of customers in store:

No. of customers in store	No. of counters in operation
1 to 3	1
4 to 6	2
More than 6	3

Customers arrive in the counters area according to a Poisson distribution with a mean rate of 10 customers per hour. The average check-out time per customer is exponential with mean 12 minutes. Determine the steady-state probability p_n of n customers in the check-out area.

$$\sum p_n = 1$$

$$p_0 + p_0(2 + 4 + 8 + 8 + 8 + 8 + 8(\frac{2}{3}) + 8(\frac{2}{3})^2 + 8(\frac{2}{3})^3 + ...) = 1$$

$$p_0(31 + 8(1 + \frac{2}{3} + (\frac{2}{3})^2 + ...)) = 1$$

Using the geometric sum series: $\sum x^i = \frac{1}{1-x}$, $|x| < 1$

$$p_0\left(31 + 8\left(\frac{1}{1-\frac{2}{3}}\right)\right) = 1 \quad \rightarrow \; p_0 = 1/55$$

Example

B&K Groceries operates with three check-out counters. The manager uses the following schedule to determine the number of counters in operation, depending on the number of customers in store:

No. of customers in store	No. of counters in operation
1 to 3	1
4 to 6	2
More than 6	3

Customers arrive in the counters area according to a Poisson distribution with a mean rate of 10 customers per hour. The average check-out time per customer is exponential with mean 12 minutes. Determine the steady-state probability p_n of n customers in the check-out area.

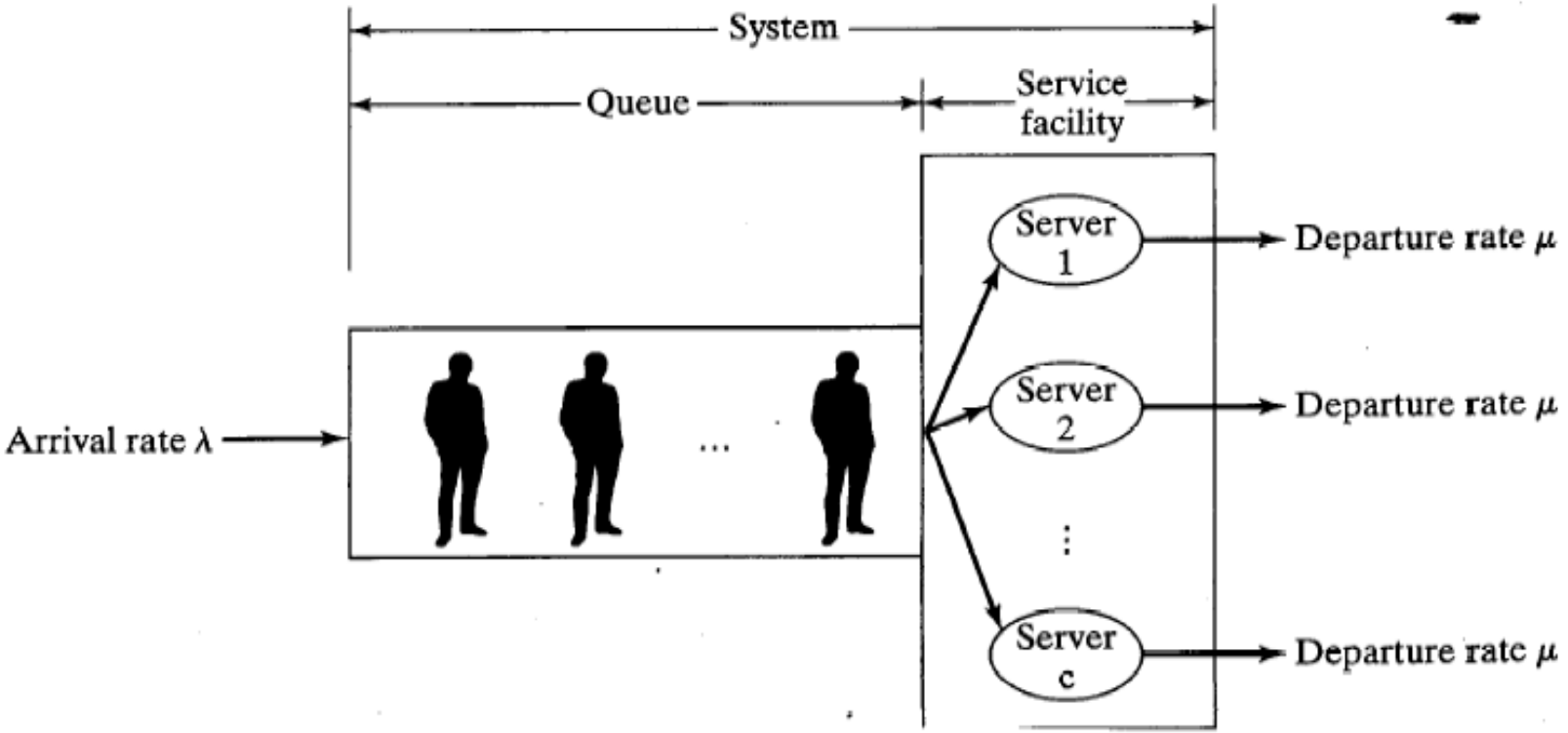
- Determine the probability that only one counter is open

$$p_1+p_2+p_3 = (2+4+8)*(1/55) = 0.255$$

- Determine the average number of idle counters.

Expected number of idle counters = $3p_0 + 2(p_1+p_2+p_3)+ 1(p_4+p_5+p_6) + 0(p_7+p_8+\ldots)$

18.6 Specialized Poisson Queues



All parallel servers are identical.
Customers in system=in queue + in service

18.6 Specialized Poisson Queues

- **Notation** $(a/b/c):(d/e/f)$

a = Arrivals distribution

b = Departures (service time) distribution

c = Number of parallel servers $(= 1, 2, \ldots, \infty)$

d = Queue discipline

e = Maximum number (finite or infinite) allowed in the system (in-queue plus in-service)

f = Size of the calling source (finite or infinite)

18.6 Specialized Poisson Queues

- **Notation** $(a/b/c):(d/e/f)$

- The possible values for a and b (the distributions):

M = Markovian (or Poisson) arrivals or departures distribution (or equivalently exponential interarrival or service time distribution)

D = Constant (deterministic) time

E_k = Erlang or gamma distribution of time (or, equivalently, the sum of independent exponential distributions)

GI = General (generic) distribution of interarrival time

G = General (generic) distribution of service time

18.6 Specialized Poisson Queues

- **Notation** $(a/b/c):(d/e/f)$
- The possible values for d (the queue discipline):

$FCFS$ = First come, first served

$LCFS$ = Last come, first served

$SIRO$ = Service in random order

GD = General discipline (i.e., any type of discipline)

18.6.1 Steady state measures of performance

- **Steady state measures of performance**

L_s = Expected number of customers in *system*

L_q = Expected number of customers in *queue*

W_s = Expected waiting time in *system*

W_q = Expected waiting time in *queue*

$\bar{c}$ = Expected number of busy servers

18.6.1 Steady state measures of performance

- **The expected number of customers in the system**

$$L_s = \sum_{n=1}^{\infty} n p_n$$

- **The expected number of customers in the queue**

$$L_q = \sum_{n=c+1}^{\infty} (n - c) p_n$$

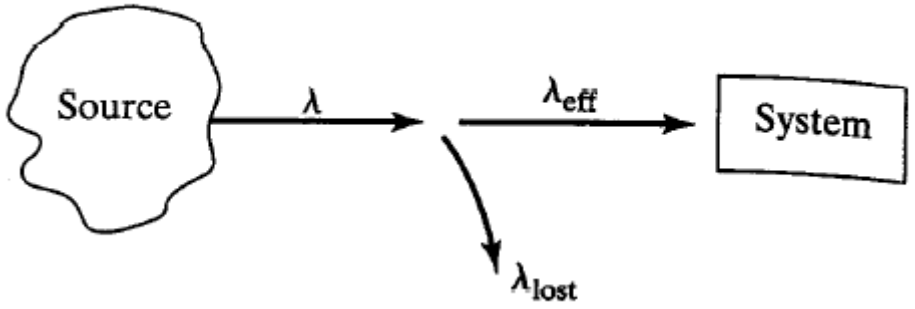
18.6.1 Steady state measures of performance

- **Little's formula: The relationship between L_s and W_s (and L_q and W_q)**

$$L_s = \lambda_{\text{eff}} W_s$$

$$L_q = \lambda_{\text{eff}} W_q$$

- λ_{eff} **is the effective arrival rate to the system**



18.6.1 Steady state measures of performance

- **Relationship between W_s and W_q**

$$\begin{pmatrix}\text{Expected waiting}\\ \text{time in system}\end{pmatrix} = \begin{pmatrix}\text{Expected waiting}\\ \text{time in queue}\end{pmatrix} + \begin{pmatrix}\text{Expected service}\\ \text{time}\end{pmatrix}$$

- **Or**

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

- **By multiplying by λ_{eff}, we can obtain:**

$$L_s = L_q + \frac{\lambda_{\text{eff}}}{\mu}$$

18.6.1 Steady state measures of performance

- **The expected number of busy servers**

$$\bar{c} = L_s - L_q = \frac{\lambda_{\text{eff}}}{\mu}$$

- **Facility utilization:**

$$\begin{pmatrix} \text{Facility} \\ \text{utilization} \end{pmatrix} = \frac{\bar{c}}{c}$$

18.6.2 Single-Server Models

- Note that:
  - The results of all these specialized models are special cases of the generalized model of Section 18.5.
  - All measures of performance presented in 18.6.1 are independent of the queue discipline (GD).

18.6.2 Single-Server Models

- First model: $(M/M/1):(GD/\infty/\infty)$

- We have:

$$\left.\begin{aligned}\lambda_n &= \lambda \\ \mu_n &= \mu\end{aligned}\right\}, n = 0, 1, 2, \ldots$$

$$\text{Also, } \lambda_{\text{eff}} = \lambda \text{ and } \lambda_{\text{lost}} = 0$$

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

- Then, from the generalized model (after simplification):

$$p_n = \rho^n p_0, n = 0, 1, 2, \ldots$$

18.6.2 Single-Server Models

- To get p_0: $p_0(1 + \rho + \rho^2 + \ldots) = 1$

  Assuming $\rho < 1$

- The sum of the geometric series is: $\left(\frac{1}{1-\rho}\right)$

- Thus $p_0 = 1 - \rho$, provided $\rho < 1$.

$$p_n = (1 - \rho)\rho^n, n = 1, 2, \ldots (\rho < 1)$$

- The measure of performance

$$L_s = \sum_{n=0}^{\infty} n p_n = \sum_{n=0}^{\infty} n(1 - \rho)\rho^n$$

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

$$= (1 - \rho)\rho \frac{d}{d\rho}\left(\frac{1}{1 - \rho}\right) = \frac{\rho}{1 - \rho}$$

What if ρ is greater than 1?

18.6.2 Single-Server Models

- Because $\lambda_{\text{eff}} = \lambda$

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu - \lambda}$$

$$W_q = W_s - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}$$

$$L_q = \lambda W_q = \frac{\rho^2}{1-\rho}$$

$$\bar{c} = L_s - L_q = \rho$$

18.7-8 Other queuing models

The literature is very rich with queuing models:

- **Multiple server models** $(M/M/c):(GD/\infty/\infty)$, $(M/M/c):(GD/N/\infty), c \leq N.$
- $(M/M/1):(FCFS/\infty/\infty)$
- $(M/M/\infty):(GD/\infty/\infty)$
- $(M/M/R):(GD/K/K), R < K$

Models with non-Poisson distributions are usually very complex:

- **Example:**

$(M/G/1):(GD/\infty/\infty)$

Other models in the literature

- Queues with priority for service, network queues, etc.

- Appendix: Example

18.6.1 Steady state measures of performance

- **Example:**

Visitors' parking at a college is limited to five spaces only. Cars making use of this space arrive according to a Poisson distribution at the rate of six cars per hour. Parking time is exponentially distributed with a mean of 30 minutes. Visitors who cannot find an empty space on arrival may temporarily wait until a parked car leaves. That temporary space can hold only three cars. Other cars that cannot park or find a temporary waiting space must go elsewhere. Determine the following:

(a) The probability, p_n, of n cars in the system.
(b) The effective arrival rate for cars that actually use the lot.
(c) The average number of cars in the lot.
(d) The average time a car waits for a parking space inside the lot.
(e) The average number of *occupied* parking spaces.
(f) The average utilization of the parking lot.

18.6.1 Steady state measures of performance

- **Example:**

Visitors' parking at a college is limited to five spaces only. Cars making use of this space arrive according to a Poisson distribution at the rate of six cars per hour. Parking time is exponentially distributed with a mean of 30 minutes. Visitors who cannot find an empty space on arrival may temporarily wait until a parked car leaves. That temporary space can hold only three cars. Other cars that cannot park or find a temporary waiting space must go elsewhere. Determine the following:

(a) The probability, p_n, of n cars in the system.

Server = parking space ➔ c = 5 parallel servers

The maximum capacity of the system is 5 +3 = 8 cars

λ_n = 6 cars/hour with n=0, 1, …, 8

$$\mu_n = \begin{cases} n\,(60/30) = 2n \text{ cars/hour}, & n=1, 2, \ldots, 5 \\ 5\,(60/30) = 10 \text{ cars/hour}, & n=6, 7, 8 \end{cases}$$

18.6.1 Steady state measures of performance

$$p_1 = \left(\frac{\lambda_0}{\mu_1}\right) p_0 = \frac{6}{2*1} p_0 = \frac{3}{1} p_0$$

$$p_2 = \left(\frac{\lambda_1}{\mu_2}\frac{\lambda_0}{\mu_1}\right) p_0 = (\frac{3}{2})*(\frac{3}{1}) p_0$$

$$p_3 = \left(\frac{\lambda_2 * \lambda_1}{\mu_3 * \mu_2}\frac{\lambda_0}{\mu_1}\right) p_0 = (\frac{3}{3})*(\frac{3}{2})*(\frac{3}{1}) p_0 = \frac{3^3}{3!} p_0$$

$$p_4 = \frac{3^4}{4!} p_0$$

$$p_5 = \frac{3^5}{5!} p_0$$

$$p_6 = \frac{6}{5*2}\frac{3^5}{5!} p_0 = \frac{3^6}{5*5!} p_0$$

$$p_7 = \frac{6}{5*2}\frac{3^6}{5*5!} p_0 = \frac{3^7}{5^2*5!} p_0 = \frac{3^7}{5^{(7-5)}*5!} p_0$$

$$p_8 = \frac{6}{5*2}\frac{3^7}{5^2*5!} p_0 = \frac{3^8}{5^3*5!} p_0 = \frac{3^8}{5^{(8-5)}*5!} p_0$$

$$p_n = \frac{3^n}{n!} p_0, n = 1,2,3,4,5$$

$$p_n = \frac{3^n}{5!5^{n-5}} p_0, n = 6,7,8$$

18.6.1 Steady state measures of performance

(a) The probability, p_n, of n cars in the system.

$p_0+p_1+\ldots.+p_8 = 1$

$p_0 = 0.04812$

n	1	2	3	4	5	6	7	8
p_n	.14436	.21654	.21654	.16240	.09744	.05847	.03508	.02105

(b) The effective arrival rate for cars that actually use the lot.

A car will not be able to enter the parking lot if 8 cars are already in.

The portion of cars that will not be able to enter the lot is p_8.

$\lambda_{lost} = \lambda\, p_8 = 6 * .\ 02105 = .1263$ cars/hour

$\lambda_{eff} = \lambda - \lambda_{lost} = 6 - .1263 = 5.8737$ cars/hour

18.6.1 Steady state measures of performance

(c) The average number of cars in the lot.

$L_s = \Sigma\, n\, p_n$

$L_s = 0\, p_0 + 1\, p_1 + 2\, p_2 + \ldots + 8\, p_8 = 3.1286$

(d) The average time a car waits for a parking space inside the lot.

$W_q = W_s - 1/\mu$

$$W_s = \frac{L_s}{\lambda_{e\,ff}} = \frac{3.1286}{5.8737} = .53265\ hour$$

$$W_q = .53265 - \frac{1}{2} = .03265\ hour$$

(e) The average number of *occupied* parking spaces.

$$\bar{c} = L_s - L_q = \frac{\lambda_{eff}}{\mu} = \frac{5.8737}{2} = 2.9368\ spaces$$

(f) The average utilization of the parking lot.

$$Parking\ \_utilizatio\ n = \frac{\bar{c}}{c} = \frac{2.9368}{5} = .58736$$