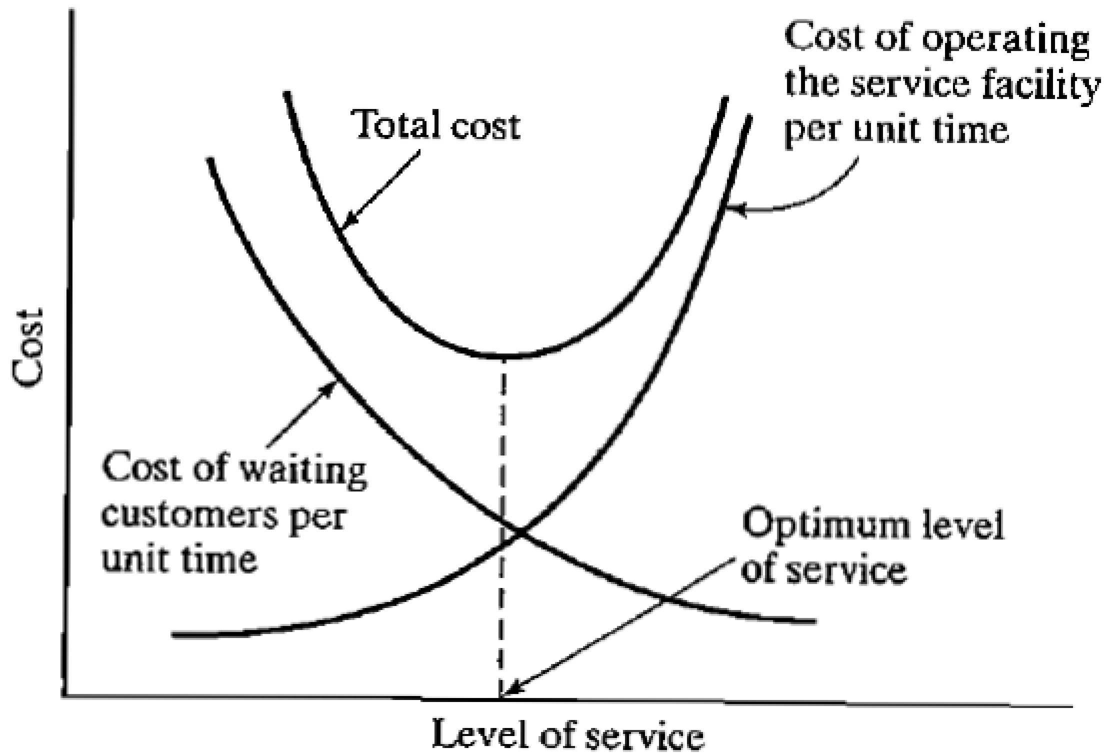
Queueing Theory

Why study queues



Elements of a queueing model

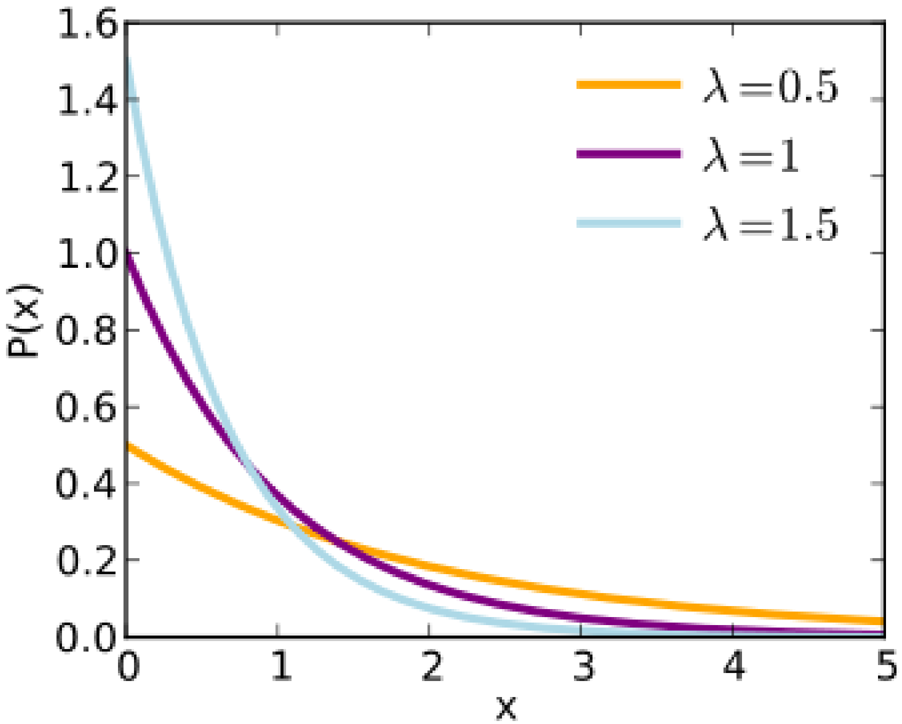
- Customer: availing service
- Server: Giving service
- Source: from which customers are generated
- Queue: waiting line before server
- Waiting customer: customer in the queue
- Queue size: No of customers that can be accommodated in the queue(N/infinity)
- Queue discipline: FCFS/LCFS/ SIRO/Priority
- Facility: place for providing service
- Idle time: server is free and there is no waiting customer
- Inter arrival time: time between two arrivals
- Service time: time taken by the server to give service to a single customer
- Jockeying: moving from one queue to other
- Balking: Not joining the queue anticipating long delay
- Renege: Leaving from queue after waiting for a long time.

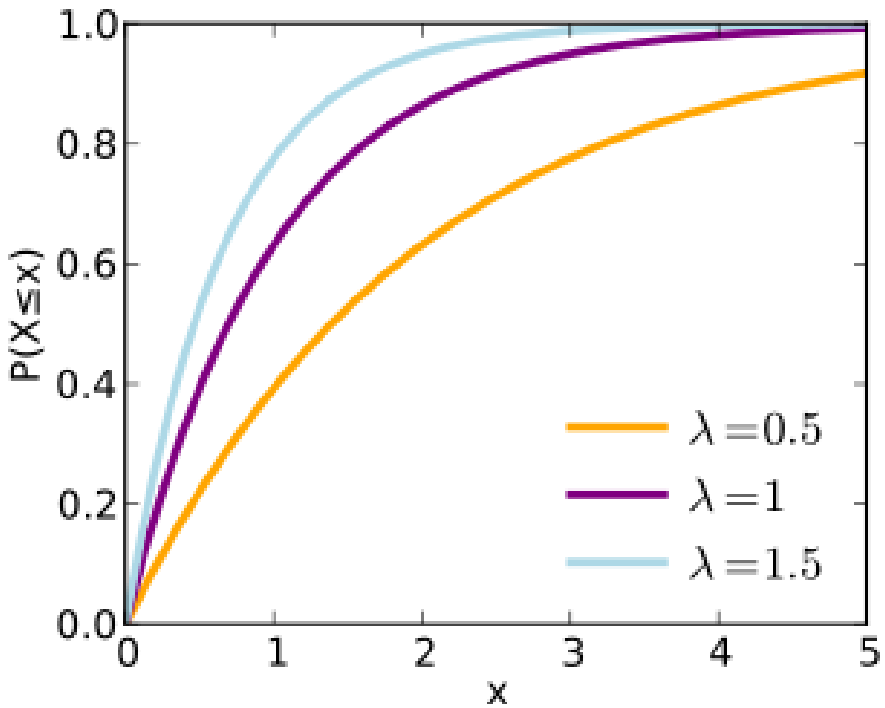
Role of exponential distribution

- Two statistical properties to determine operating characteristics of queuing system.
  - Inter arrival time
  - Service time
- In most cases occurrence of an event (arrival of a customer/completion of a service) is not affected by the time that is lapsed since the occurrence of the last event. Hence, can be approximated by exponential distribution.

Exponential Distribution

- pdf $f(t)=\lambda\, e^{-\lambda t}$
- cdf $F(t)$= P{t<=T}=1 – $e^{-\lambda T}$
- Mean = $1/\lambda$



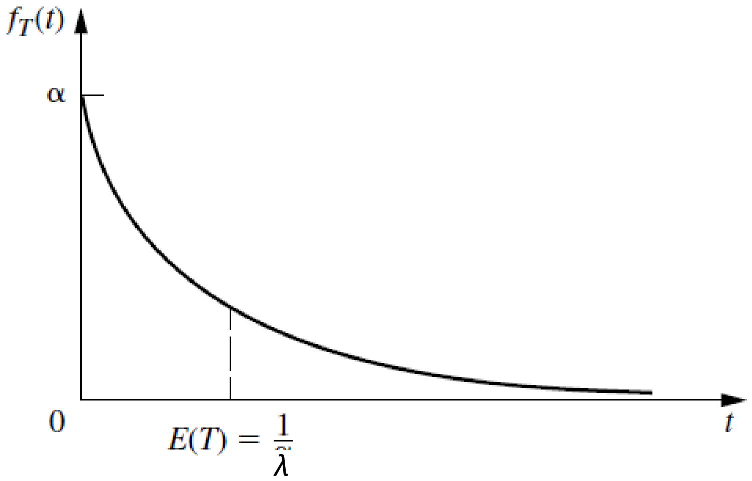


Properties of exponential distribution

Property 1: $f_T(t)$ is a strictly *decreasing* function of t ($t \geq 0$).

One consequence of Property 1 is that

$$P\{0 \leq T \leq \Delta t\} > P\{t \leq T \leq t + \Delta t\}$$



$$P\left\{0 \leq T \leq \frac{1}{2}\frac{1}{\lambda}\right\} = 0.393$$

$$P\left\{\frac{1}{2}\frac{1}{\lambda} \leq T \leq \frac{3}{2}\frac{1}{\lambda}\right\} = 0.383,$$

Implications:
Service time: Most of the time service time is around the mean. Occasionally, service time is very high.
Inter-arrivel time: Arrivals are clustered with occasional long gaps.

Properties of exponential distribution

Property 2: Lack of memory

The probability distribution of the *remaining* time until the event (arrival or service completion) occurs always is the same, regardless of how much time (t) already has passed.

$$
\begin{aligned}
P\{T > t + \Delta t \mid T > \Delta t\} &= \frac{P\{T > \Delta t, T > t + \Delta t\}}{P\{T > \Delta t\}} \\
&= \frac{P\{T > t + \Delta t\}}{P\{T > \Delta t\}} \\
&= \frac{e^{-\alpha(t+\Delta t)}}{e^{-\alpha \Delta t}} \\
&= e^{-\alpha t} \\
&= P\{T > t\}.
\end{aligned}
$$

Implications:

Inter-arrivel time: the time until the next arrival is completely uninfluenced by when the last arrival occurred

Service time: Some customers require more careful service than others.

Properties of exponential distribution

Property 3: The minimum of several independent exponential random variables has an exponential distribution

let $T_1, T_2, \ldots, T_n$ be *independent* exponential random variables with parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$.

let $U = \min\{T_1, T_2, \ldots, T_n\}$

$$\begin{aligned}
P\{U > t\} &= P\{T_1 > t, T_2 > t, \ldots, T_n > t\} \\
&= P\{T_1 > t\}P\{T_2 > t\} \cdots P\{T_n > t\} \\
&= e^{-\alpha_1 t} e^{-\alpha_2 t} \cdots e^{\alpha_n t} \\
&= \exp\left(-\sum_{i=1}^{n} \alpha_i t\right),
\end{aligned}$$

$$\alpha = \sum_{i=1}^{n} \alpha_i.$$

<u>Implications:</u>

<u>Inter-arrivel time:</u> One can choose to ignore the distinction between types of customers and still have exponential interarrival times for the queuing model

<u>Service time:</u> Is service times are same Multiple server models can be treated as a single server model with parameter *nμ*.

Properties of exponential distribution

Property 4: Relationship with Poisson distribution

Define $p_0(t)$ = Probability of no arrivals during a period of time t

Given that the interarrival time is exponential and that the arrival rate is λ customers per unit time, then

$$\begin{aligned} p_0(t) &= P\{\text{interarrival time} \geq t\} \\ &= 1 - P\{\text{interarrival time} \leq t\} \\ &= 1 - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t} \end{aligned}$$

For a sufficiently small time interval $h > 0$, we have

$$p_0(h) = e^{-\lambda h} = 1 - \lambda h + \frac{(\lambda h)^2}{2!} - \ldots = 1 - \lambda h + 0(h^2)$$

The exponential distribution is based on the assumption that during $h > 0$, at most one event (arrival) can occur. Thus, as $h \rightarrow 0$,

$$p_1(h) = 1 - p_0(h) \approx \lambda h$$

To derive the distribution of the *number* of arrivals during a period t when the interarrival time is exponential with mean $\frac{1}{\lambda}$, define

$$p_n(t) = \text{Probability of } n \text{ arrivals during } t$$

For a sufficiently small $h > 0$,

$$p_n(t+h) \approx p_n(t)(1-\lambda h) + p_{n-1}(t)\lambda h, \quad n > 0$$

Probability of *n* arrivals during *t* and zero arrival during *h* OR probability of (*n*-1) arrival during *t* and zero arrival during *h*.

$$p_0(t+h) \approx p_0(t)(1-\lambda h), \quad n = 0$$

Rearranging the terms and taking the limits as $h \rightarrow 0$, we get

$$p'_n(t) = \lim_{h\to 0} \frac{p_n(t+h) - p_n(t)}{h} = -\lambda p_n(t) + \lambda p_{n-1}(t), \; n > 0$$

$$p'_0(t) = \lim_{h\to 0} \frac{p_0(t-h) - p_0(t)}{h} = -\lambda p_0(t), \quad n = 0$$

The solution of the preceding difference-differential equations yields

$$p_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \; n = 0, 1, 2, \ldots$$

This is a **Poisson distribution** with mean $E\{n|t\} = \lambda t$ arrivals during t.

The preceding result shows that if the time between arrivals is exponential with mean $\frac{1}{\lambda}$ then the number of arrivals during a specific period t is Poisson with mean λt. The converse is true also.

	Exponential	Poisson
Random variable	*Time* between successive arrivals, t	*Number* of arrivals, n, during a specified period T
Range	$t \geq 0$	$n = 0, 1, 2, \ldots$
Density function	$f(t) = \lambda e^{-\lambda t}, t \geq 0$	$p_n(T) = \dfrac{(\lambda T)^n e^{-\lambda T}}{n!}, n = 0, 1, 2, \ldots$
Mean value	$\dfrac{1}{\lambda}$ time units	λT arrivals during T
Cumulative probability	$P\{t \leq A\} = 1 - e^{-\lambda A}$	$p_{n \leq N}(T) = p_0(T) + p_1(T) + \ldots + p_N(T)$
$P\{$no arrivals during period $A\}$	$P\{t > A\} = e^{-\lambda A}$	$p_0(A) = e^{-\lambda A}$

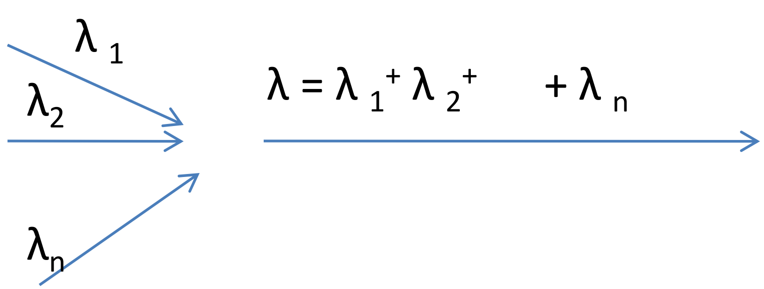
Properties of exponential distribution

Property 5: For all positive values of t, $P\{T \leq t + \Delta t \,|\, T > t\} \approx \alpha\,\Delta t$, for small Δt.

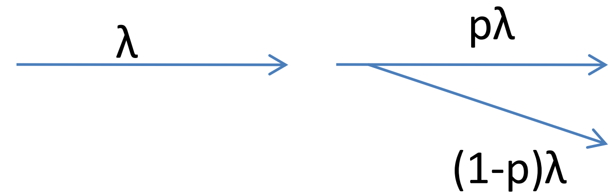
$$
\begin{aligned}
P\{T \leq t + \Delta t \,|\, T > t\} &= P\{T \leq \Delta t\} \\
&= 1 - e^{-\alpha\,\Delta t}. \\
&= 1 - 1 + \alpha\,\Delta t - \sum_{n=2}^{\infty} \frac{(-\alpha\,\Delta t)^n}{n!} \\
&\approx \alpha\,\Delta t, \qquad \text{for small } \Delta t,^{1}
\end{aligned}
$$

Property 6: Unaffected by aggregation or disaggregation.

Aggregation



Disaggregation



Birth and death process

- Special type of continuous time Markov process
- Transitions are allowed to the state immediately above or below in the natural number ordering.

- Tri- diagonal transition probability matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\mu+\lambda) & \lambda & & \\ & \mu & -(\mu+\lambda) & \lambda & \\ & & \mu & -(\mu+\lambda) & \lambda \\ & & & & \ddots \end{pmatrix}$$

- Poisson process: Special case of Birth and Death process. Only forward transitions are allowed.

Generalized Poisson Queuing Models

- Arrival and departure rate follow Poisson distribution.
- Hence inter-arrival and service times are exponential
- Both arrival and departure rates are state dependent
- Important to study long-run steady state behavior

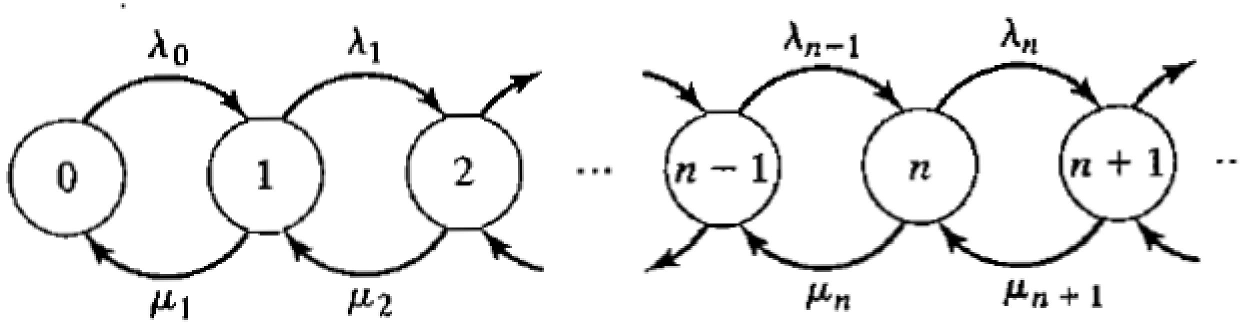
Define

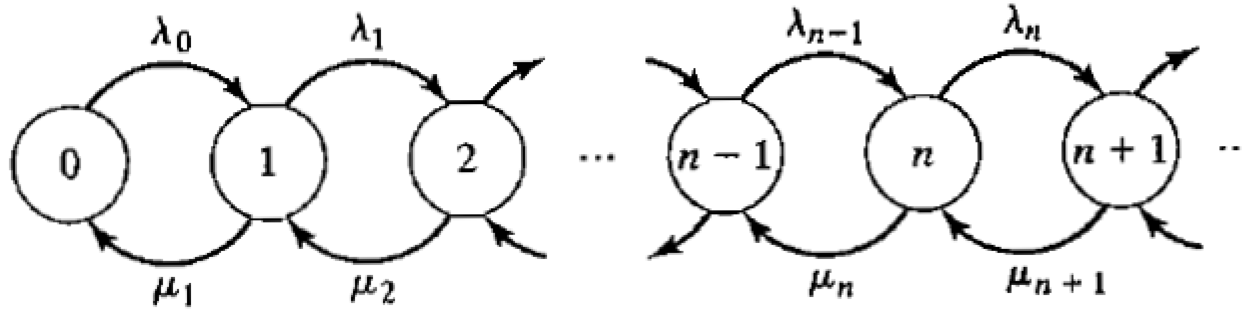
n = Number of customers in the system (in-queue plus in-service)

λ_n = Arrival rate given n customers in the system

μ_n = Departure rate given n customers in the system

p_n = Steady-state probability of n customers in the system





Under steady-state conditions, for $n > 0$, the *expected* rates of flow into and out of state n must be equal. Based on the fact that state n can be changed to states $n-1$ and $n+1$ only, we get

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

Similarly,

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

Equating the two rates, we get the following **balance equation**:

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

the balance equation associated with $n = 0$, is

$$\lambda_0 p_0 = \mu_1 p_1$$

The balance equations are solved recursively

$$\lambda_0 p_0 = \mu_1 p_1 \Rightarrow p_1 = \left(\frac{\lambda_0}{\mu_1}\right) p_0$$

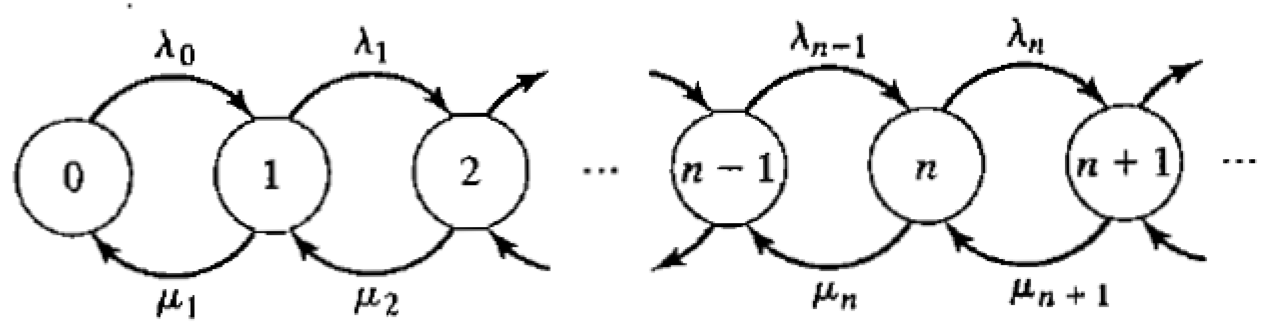
for $n = 1$, $\Rightarrow \lambda_0 p_0 + \mu_2 p_2 = (\lambda_1 + \mu_1) p_1$

Substituting $p_1 = \left(\frac{\lambda_0}{\mu_0}\right) p_0 \Rightarrow p_2 = \left(\frac{\lambda_1 \lambda_0}{\mu_2 \mu_1}\right) p_0$

In general,

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

The value of p_0 is determined from the equation $\sum_{n=0}^{\infty} p_n = 1$



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:

Problem

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.

Solution

$$\lambda_n = \lambda = 10 \text{ customers per hour}, \qquad n = 0, 1, \ldots$$

$$\mu_n = \begin{cases} \frac{60}{12} = 5 \text{ customers per hour}, & n = 0, 1, 2, 3 \\ 2 \times 5 = 10 \text{ customers per hour}, & n = 4, 5, 6 \\ 3 \times 5 = 15 \text{ customers per hour}, & n = 7, 8, \ldots \end{cases}$$

$$p_1 = \left(\tfrac{10}{5}\right)p_0 = 2p_0$$

$$p_2 = \left(\tfrac{10}{5}\right)^2 p_0 = 4p_0$$

$$p_3 = \left(\tfrac{10}{5}\right)^3 p_0 = 8p_0$$

$$p_4 = \left(\tfrac{10}{5}\right)^3\left(\tfrac{10}{10}\right)p_0 = 8p_0$$

$$p_5 = \left(\tfrac{10}{5}\right)^3\left(\tfrac{10}{10}\right)^2 p_0 = 8p_0$$

$$p_6 = \left(\tfrac{10}{5}\right)^3\left(\tfrac{10}{10}\right)^3 p_0 = 8p_0$$

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

$$p_0 + p_0\{2 + 4 + 8 + 8 + 8 + 8 + 8(\tfrac{2}{3}) + 8(\tfrac{2}{3})^2 + 8(\tfrac{2}{3})^3 + \ldots\} = 1$$

$$p_0\{31 + 8(1 + (\tfrac{2}{3}) + (\tfrac{2}{3})^2 + \ldots)\} = 1$$

$$p_0\left\{31 + 8\left(\frac{1}{1 - \frac{2}{3}}\right)\right\} = 1$$

probability that there are at most three customers

$$p_1 + p_2 + p_3 = (2 + 4 + 8)(\tfrac{1}{55}) \approx .255$$

$$\begin{pmatrix}\text{Expected number}\\ \text{of idle counters}\end{pmatrix} = 3p_0 + 2(p_1 + p_2 + p_3) + 1(p_4 + p_5 + p_6)$$

$$+ 0(p_7 + p_8 + \ldots)$$

$$= 1 \text{ counter}$$

Specialized Poisson Queues

notation for summarizing the characteristics

$$(a/b/c):(d/e/f)$$

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) (symbols a and b) is

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

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

The queue discipline notation (symbol d) includes

$FCFS$ = First come, first served

$LCFS$ = Last come, first served

$SIRO$ = Service in random order

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

What does it represent??? $(M/D/10):(GD/20/\infty)$

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

system includes both the *queue* and the *service facility*.

From Definition →

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

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

Little's Formula →

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

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

$$\begin{pmatrix}\text{Expected waiting}\\ \text{time in system}\end{pmatrix} = \begin{pmatrix}\text{Expected waiting}\\ \text{time in queue}\end{pmatrix} + \begin{pmatrix}\text{Expected service}\\ \text{time}\end{pmatrix} \rightarrow W_s = W_q + \frac{1}{\mu}$$

multiplying both sides of the last formula by λ_{eff} →

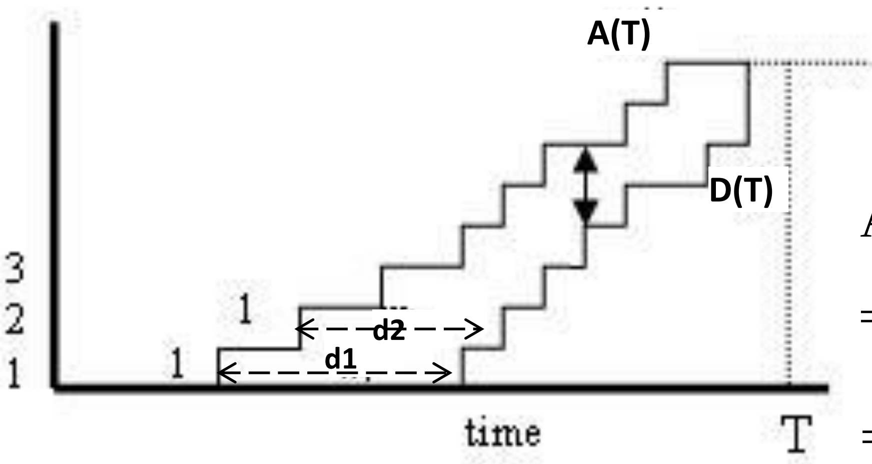
$$L_s = L_q + \frac{\lambda_{\text{eff}}}{\mu}$$

average number of *busy* servers →

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

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

Illustration of Little's Formula



No of customers in the system N(T)=A(T)-D(T)
A(T): Arrivals , D(T): Departures

Average number of customers $= L_S(T)$

$$= \frac{1}{T}\int_0^T N(t)dt = \text{area under the curves during T}$$

$$= \frac{1}{T}[1\times d_1 + 1\times d_2 ....] = \frac{\Delta D}{T}$$

Where d_i is the delay experienced by customer i

$$\Rightarrow L_S(T) = \frac{\Delta D}{T}$$

Average Arrival Rate λ(T)= $\dfrac{A(T)}{T}$

Average delay per customer (waiting time) W_s(T)= $\dfrac{\Delta D}{A(T)}$

$$L_S(T) = \frac{\Delta D}{T} = \frac{\Delta D}{A(T)}\frac{A(T)}{T} = W_s(T)\lambda(\mathrm{T})$$

At steady state when $T \to \infty,\ L_S = W_s\lambda$

Problem

Visitors' parking at Ozark 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 inside the lot 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.

Solution

No. of parallel servers c = 5

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

$$\lambda_n = 6 \text{ cars/hour}, n = 0, 1, 2, \ldots, 8$$

$$\mu_n = \begin{cases} n\left(\frac{60}{30}\right) = 2n \text{ cars/hour}, n = 1, 2, 3, 4, 5 \\ 5\left(\frac{60}{30}\right) = 10 \text{ cars/hour}, n = 6, 7, 8 \end{cases}$$

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

$$p_n = \begin{cases} \frac{3^n}{n!}p_0, & n = 1, 2, 3, 4, 5 \\ \frac{3^n}{5!\,5^{n-5}}p_0, & n = 6, 7, 8 \end{cases}$$

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

$$p_0 + p_0\left(\frac{3}{1!} + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!} + \frac{3^6}{5!\,5} + \frac{3^7}{5!\,5^2} + \frac{3^8}{5!\,5^3}\right) = 1$$

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

$p_0 = .04812$

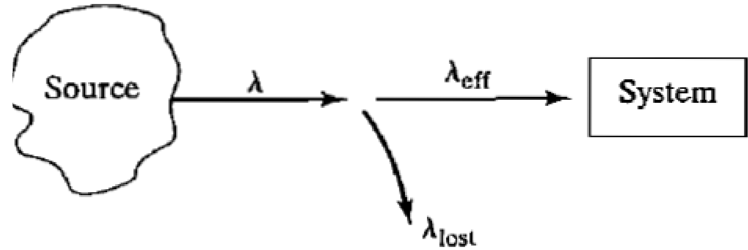
n	1	2	3	4	5	6	7	8
p_n	.14436	.21654	.21654	.16240	.09744	.05847	.03508	.02105

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

$\lambda = \lambda_{\text{eff}} + \lambda_{\text{lost}}$.

$\lambda_{\text{lost}} = \lambda p_8 = 6 \times .02105 = .1263$ cars per hour

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



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

$$L_s = 0p_0 + 1p_1 + \ldots + 8p_8 = 3.1286 \text{ cars}$$

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

$$W_q = W_s - \frac{1}{\mu}, \quad W_s = \frac{L_s}{\lambda_{\text{eff}}} = \frac{3.1286}{5.8737} = .53265 \text{ hour}, \quad W_q = .53265 - \frac{1}{2} = .03265 \text{ hour}$$

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

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

(f) The average utilization of the parking lot.

$$\text{Parking lot utilization} = \frac{\bar{c}}{c} = \frac{2.9368}{5} = .58736$$

Single-Server Models

$\boxed{(M/M/1):(GD/\infty/\infty)}$ (Exponential inter-arrival and service time, single server, general queue discipline, infinite queue size and population size)

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

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

From generalized model

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

Then

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

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

$$p_0(1 + \rho + \rho^2 + \ldots) = 1$$

$$p_0 = 1 - \rho, \text{ provided } \rho < 1$$

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

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

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

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

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

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

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

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

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

Automata car wash facility operates with only one bay. Cars arrive according to a Poisson distribution with a mean of 4 cars per hour, and may wait in the facility's parking lot if the bay is busy. The time for washing and cleaning a car is exponential, with a mean of 10 minutes. Cars that cannot park in the lot can wait in the street bordering the wash facility. This means that, for all practical purposes, there is no limit on the size of the system. The manager of the facility wants to determine the size of the parking lot.

For this situation, we have $\lambda = 4$ cars per hour, and $\mu = \frac{60}{10} = 6$ cars per hour. Because $\rho = \frac{\lambda}{\mu} < 1$, the system can operate under steady-state conditions.

```
Scenario1: (M/M/1):(GD/infinity/infinity)

Lambda =        4.00000                  Mu =        6.00000
Lambda eff = 4.00000                     Rho/c =     0.66667

Ls =       2.00000                       Lq =        1.33333
Ws =       0.50000                       Wq =        0.33333
```

n	Probability pn	Cumulative Pn	n	Probability pn	Cumulative Pn
0	0.33333	0.33333	13	0.00171	0.99657
1	0.22222	0.55556	14	0.00114	0.99772
2	0.14815	0.70370	15	0.00076	0.99848
3	0.09877	0.80247	16	0.00051	0.99899
4	0.06584	0.86831	17	0.00034	0.99932
5	0.04390	0.91221	18	0.00023	0.99955
6	0.02926	0.94147	19	0.00015	0.99970
7	0.01951	0.96098	20	0.00010	0.99980
8	0.01301	0.97399	21	0.00007	0.99987
9	0.00867	0.98266	22	0.00004	0.99991
10	0.00578	0.98844	23	0.00003	0.99994
11	0.00385	0.99229	24	0.00002	0.99996
12	0.00257	0.99486	25	0.00001	0.99997

Generally, using L_q as the sole basis for the determination of the number of parking spaces is not advisable, because the design should, in some sense, account for the maximum possible length of the queue. For example, it may be more plausible to design the parking lot such that an arriving car will find a parking space at least 90% of the time. To do this, let S represent the number of parking spaces. Having S parking spaces is equivalent to having $S + 1$ spaces in the *system* (queue plus wash bay). An arriving car will find a space 90% of the time if there are *at most* S cars in the system. This condition is equivalent to the following probability statement:

$$p_0 + p_1 + \ldots + p_s \geq .9$$

From Figure 15.5, *cumulative* p_n for $n = 5$ is .91221. This means that the condition is satisfied for $S \geq 5$ parking spaces.

The number of spaces S can be determined also by using the mathematical definition of p_n—that is,

$$(1 - \rho)(1 + \rho + \rho^2 + \ldots + \rho^S) \geq .9$$

The sum of the truncated geometric series equals $\frac{1 - \rho^{S+1}}{1 - \rho}$. Thus the condition reduces to

$$(1 - \rho^{S+1}) \geq .9$$

Simplification of the inequality yields

$$\rho^{S+1} \leq .1$$

Taking the logarithms on both sides (and noting that $\log(x) < 0$ for $0 < x < 1$, which reverses the direction of the inequality), we get

$$S \geq \frac{\ln(.1)}{\ln(\frac{4}{6})} - 1 = 4.679 \approx 5$$

Single-Server Models

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

(Exponential inter-arrival and service time, single server, general queue discipline, finite queue size and infinite population size)

$$\lambda_n = \begin{cases} \lambda, & n = 0, 1, \ldots, N-1 \\ 0, & n = N, N+1 \end{cases}$$

$$\mu_n = \mu, \quad n = 0, 1, \ldots$$

Using $\rho = \frac{\lambda}{\mu}$, the generalized model

$$p_n = \begin{cases} \rho^n p_0 & n \le N \\ 0, & n > N \end{cases}$$

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

$$p_0(1 + \rho + p^2 + \ldots + \rho^N) = 1$$

$$p_0 = \begin{cases} \dfrac{(1-\rho)}{1-\rho^{N+1}}, & \rho \neq 1 \\ \dfrac{1}{N+1}, & \rho = 1 \end{cases}$$

$$p_n = \left\{ \begin{matrix} \dfrac{(1-\rho)\rho^n}{1-\rho^{N+1}}, & \rho \neq 1 \\ \dfrac{1}{N+1}, & \rho = 1 \end{matrix} \right\}, n = 0, 1, \ldots, N$$

Since

$$\lambda_{\text{lost}} = \lambda p_N$$

$$\lambda_{\text{eff}} = \lambda - \lambda_{\text{lost}} = \lambda(1 - p_N)$$

$$\begin{aligned} L_s &= \sum_{n=1}^{N} n p_n \\ &= \frac{1-\rho}{1-\rho^{N+1}} \sum_{n=0}^{N} n\rho^n \\ &= \left(\frac{1-\rho}{1-\rho^{N+1}}\right)\rho \frac{d}{d\rho} \sum_{n=0}^{N} \rho^n \\ &= \frac{(1-\rho)\rho}{1-\rho^{N+1}} \frac{d}{d\rho}\left(\frac{1-\rho^{N+1}}{1-\rho}\right) \\ &= \frac{\rho[1-(N+1)\rho^N + N\rho^{N+1}]}{(1-\rho)(1-\rho^{N+1})}, \rho \neq 1 \end{aligned}$$

The value of ρ need *not* be less than 1 in this model, because arrivals at the system are controlled by the system limit N

$$\rho = 1, L_s = \frac{N}{2}$$

Suppose that the facility has a total of four parking spaces. If the parking lot is full, newly arriving cars balk to other facilities. The owner wishes to determine the impact of the limited parking space on losing customers to the competition.

```
Scenario 1:(M/M/1):(GD/5/infinity)

Lambda =       4.00000              Mu =        6.00000
Lambda eff = 3.80752                Rho/c =     0.66667

Ls =           1.42256              Lq =        0.78797
Ws =           0.37362              Wq =        0.20695
```

n	Probability pn	Cumulative Pn	n	Probability pn	Cumulative Pn
0	0.36541	0.36541	3	0.10827	0.87970
1	0.24361	0.60902	4	0.07218	0.95188
2	0.16241	0.77143	5	0.04812	1.00000

Because the limit on the system is $N = 5$, the proportion of lost customers is $p_5 = .04812$, which, based on a 24-hour day, is equivalent to losing $(\lambda p_5) \times 24 = 4 \times .04812 \times 24 = 4.62$ cars a day. A decision regarding increasing the size of the parking lot should be based on the value of lost business.

Multiple Server Models

$(M/M/c):(GD/\infty/\infty)$

(Exponential inter-arrival and service time, c number of identical server, general queue discipline, infinite queue size and population)

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

$$\lambda_n = \lambda, \qquad n \geq 0$$

$$\mu_n = \begin{cases} n\mu, & n < c \\ c\mu, & n \geq c \end{cases}$$

$$p_n = \begin{cases} \dfrac{\lambda^n}{\mu(2\mu)(3\mu)\dots(n\mu)} p_0 \\ = \dfrac{\lambda^n}{n!\,\mu^n} p_0 = \dfrac{\rho^n}{n!} p_0, & n < c \\ \dfrac{\lambda^n}{\left(\prod_{i=1}^{c} i\mu\right)(c\mu)^{n-c}} p_0 \\ = \dfrac{\lambda^n}{c!\,c^{n-c}\mu^n} p_0 = \dfrac{\rho^n}{c!\,c^{n-c}} p_0, & n \geq c \end{cases}$$

Letting $\rho = \frac{\lambda}{\mu}$, and assuming $\frac{\rho}{c} < 1$

$$p_0 = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \sum_{n=c}^{\infty} \left(\frac{\rho}{c}\right)^{n-c} \right\}^{-1}$$

$$= \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \left(\frac{1}{1 - \frac{\rho}{c}} \right) \right\}^{-1}, \quad \frac{\rho}{c} < 1$$

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

$$= \sum_{k=0}^{\infty} k p_{k+c}$$

$$= \sum_{k=0}^{\infty} k \frac{\rho^{k+c}}{c^k c!} p_0$$

$$= \frac{\rho^{c+1}}{c!\,c} p_0 \sum_{k=0}^{\infty} k \left(\frac{\rho}{c}\right)^{k-1}$$

$$= \frac{\rho^{c+1}}{c!\,c} p_0 \frac{d}{d\left(\frac{\rho}{c}\right)} \sum_{k=0}^{\infty} \left(\frac{\rho}{c}\right)^k$$

$$= \frac{\rho^{c+1}}{(c-1)!\,(c-\rho)^2} p_0$$

Multiple Server Models

$(M/M/c):(GD/N/\infty), c \le N.$

(Exponential inter-arrival and service time, c number of identical server, general queue discipline, finite queue size and infinite population)

$$\lambda_n = \begin{cases} \lambda, & 0 \le n \le N \\ 0, & n > N \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & 0 \le n \le c \\ c\mu, & c \le n \le N \end{cases}$$

Substituting λ_n and μ_n in the general expression and noting that $\rho = \frac{\lambda}{\mu}$,

$$p_n = \begin{cases} \dfrac{\rho^n}{n!} p_0, & 0 \le n < c \\ \dfrac{\rho^n}{c!\, c^{n-c}} p_0, & c \le n \le N \end{cases}$$

$$p_0 = \begin{cases} \left(\displaystyle\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c \left(1 - \left(\frac{\rho}{c}\right)^{N-c+1}\right)}{c! \left(1 - \frac{\rho}{c}\right)} \right)^{-1}, & \dfrac{\rho}{c} \ne 1 \\ \left(\displaystyle\sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} (N - c + 1) \right)^{-1}, & \dfrac{\rho}{c} = 1 \end{cases}$$

$\frac{\rho}{c} \ne 1 \Rightarrow$

$$\begin{aligned} L_q &= \sum_{n=c}^{N} (n - c) p_n \\ &= \sum_{j=0}^{N-c} j p_{j+c} \\ &= \frac{\rho^c \rho}{c! c} p_0 \sum_{j=0}^{N-c} j \left(\frac{\rho}{c}\right)^{j-1} \\ &= \frac{\rho^{c+1}}{cc!} p_0 \frac{d}{d\left(\frac{\rho}{c}\right)} \sum_{j=0}^{N-c} \left(\frac{\rho}{c}\right)^j \\ &= \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} \left\{ 1 - \left(\frac{\rho}{c}\right)^{N-c+1} - (N - c + 1)\left(1 - \frac{\rho}{c}\right)\left(\frac{\rho}{c}\right)^{N-c} \right\} p_0 \end{aligned}$$

$\frac{\rho}{c} = 1 \Rightarrow$

$$L_q = \frac{\rho^c (N - c)(N - c + 1)}{2c!} p_0, \quad \frac{\rho}{c} = 1$$

To determine W_q W_s and L_s:

$$\lambda_{\text{lost}} = \lambda p_N$$

$$\lambda_{\text{eff}} = \lambda - \lambda_{\text{lost}} = (1 - p_N)\lambda$$

A community is served by two cab companies. Each company owns two cabs and both share the market equally, as evidenced by the fact that calls arrive at each company's dispatching office at the rate of eight per hour. The average time per ride is 12 minutes. Calls arrive according to a Poisson distribution, and the ride time is exponential. The two companies recently were bought by an investor who is interested in consolidating them into a single dispatching office to provide better service to customers. Analyze the new owner's proposal.

c	Lambda	Mu	L'da eff	p0	Ls	Ws	Lq	Wq
2	8.000	5.000	8.00	0.110	4.444	0.556	2.844	0.356
4	16.000	5.000	16.00	0.027	5.586	0.349	2.386	0.149

Figure 15.7 provides the output for the two scenarios. The results show that the waiting time for a ride is .356 hour (≈21 minutes) for the two-cab situation and .149 (≈9 minutes) for the consolidated situation, a remarkable reduction of more than 50% and a clear evidence that the consolidation of the two companies is warranted.

In the consolidated cab company problem of Example 15.6-5, suppose that new funds cannot be secured to purchase additional cabs. The owner was advised by a consultant that one way to reduce the waiting time is for the dispatching office to inform new customers of potential excessive delay once the waiting list reaches 6 customers. This move is certain to get new customers to seek service elsewhere, but will reduce the waiting time for those on the waiting list. Assess the friend's advice.

Limiting the waiting list to 6 customers is equivalent to setting $N = 6 + 4 = 10$ customers. We are thus investigating the model $(M/M/4):(GD/10/\infty)$, where $\lambda = 16$ customers per hour and $\mu = 5$ rides per hour. The following input data provide the results in Figure 15.8.

Lambda	Mu	c	System limit	Source limit
16	5	4	10	infinity

The average waiting time, W_q, before setting a limit on the capacity of the system is .149 hour ($\approx$9 minutes) (see Figure 15.7), which is about twice the new average of .075 hour ($\approx$4.5 minutes). This remarkable reduction is achieved at the expense of losing about 3.6% of potential customers ($p_{10} = .03574$). However, this result does not reflect the effect of possible loss of customer goodwill on the operation of the company.

An investor invests $1000 a month on average in one type of stock market security. Because the investor must wait for a good "buy" opportunity, the actual time of purchase is totally random. The investor usually keeps the securities for about 3 years on the average but will sell them at random times when a "sell" opportunity presents itself. Although the investor is generally recognized as a shrewd stock market player, past experience indicates that about 25% of the securities decline at about 20% a year. The remaining 75% appreciate at the rate of about 12% a year. Estimate the investor's (long-run) average equity in the stock market.

Who are the customers? The securities

What are servers? The investor

Any time securities can be bought or sold. They do not have to wait to get server's time. So the server is equivalent to infinitely many servers running parallel.

What is expected inter-arrival time? Invests $1000 a month on average. Invests once in a month OR 12 times in a year.

What is expected service time? He sells the securities on an average after every 3 years (of their purchase) OR sells 1/3 securities per year.

$\lambda = 12$ securities per year

$\mu = \frac{1}{3}$ security per year

During each year, an average of 3 ice cream shops open up in Smalltown. The average time that an ice cream shop stays in business is 10 years. On January 1, 2525, what is the average number of ice cream shops that you would find in Smalltown? If the time between the opening of ice cream shops is exponential, what is the probability that on January 1, 2525, there will be 25 ice cream shops in Smalltown?

Who are the customers? The shops

What are servers? The shops

Any time shops can be opened.

What is expected inter-arrival time? 3 shops per year . Expected inter-arrival time 1/3 years.

What is expected service time? 10 years per shop

$\lambda = 3$ shops per year and $\frac{1}{\mu} = 10$ years per shop

$(M/M/\infty):(GD/\infty/\infty)$—Self-Service Model

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

$$\mu_n = n\mu, n = 0, 1, 2, \ldots$$

$$p_n = \frac{\lambda^n}{n!\,\mu^n} p_0 = \frac{\rho^n}{n!} p_0, n = 0, 1, 2, \ \ldots$$

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

$$p_0 = \frac{1}{1 + \rho + \frac{\rho^2}{2!} + \ \ldots} = \frac{1}{e^\rho} = e^{-\rho}$$

$$p_n = \frac{e^{-\rho}\rho^n}{n!}, n = 0, 1, 2, \ldots$$

(Poisson with mean $L_s = \rho$.)

What is expected, L_q and W_q ?

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

An investor invests \$1000 a month on average in one type of stock market security. Because the investor must wait for a good "buy" opportunity, the actual time of purchase is totally random. The investor usually keeps the securities for about 3 years on the average but will sell them at random times when a "sell" opportunity presents itself. Although the investor is generally recognized as a shrewd stock market player, past experience indicates that about 25% of the securities decline at about 20% a year. The remaining 75% appreciate at the rate of about 12% a year. Estimate the investor's (long-run) average equity in the stock market.

$\lambda = 12$ securities per year
$\mu = \frac{1}{3}$ security per year

$$L_s = \rho = \frac{\lambda}{\mu} = 36 \text{ securities}$$

The estimate of the (long-run) average *annual* net worth of the investor is

$$(.25L_s \times \$1000)(1 - .20) + (.75L_s \times \$1000)(1 + .12) = \$63{,}990$$

During each year, an average of 3 ice cream shops open up in Smalltown. The average time that an ice cream shop stays in business is 10 years. On January 1, 2525, what is the average number of ice cream shops that you would find in Smalltown? If the time between the opening of ice cream shops is exponential, what is the probability that on January 1, 2525, there will be 25 ice cream shops in Smalltown?

$\lambda = 3$ shops per year and $\frac{1}{\mu} = 10$ years per shop

$L_s = 3(10) = 30$ shops

$$P_{25} = \frac{(30)^{25} e^{-30}}{25!} = .05$$

Toolco operates a machine shop with a total of 22 machines. Each machine is known to break down once every 2 hours, on the average. It takes an average of 12 minutes to complete a repair. Both the time between breakdowns and the repair time follow the exponential distribution. Toolco is interested in determining the number of repairpersons needed to keep the shop running "smoothly."

Machine Servicing Model—$(M/M/R):(GD/K/K)$, $R < K$

K machines.

R available repairpersons

λ breakdowns per unit time *per machine*

μ machines per unit time **gets repaired by a repairperson**

$$\lambda_n = \begin{cases} (K-n)\lambda, & 0 \le n \le K \\ 0. & n \ge K \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & 0 \le n \le R \\ R\mu, & R \le n \le K \end{cases}$$

From the generalized model

$$p_n = \begin{cases} C_n^K \rho^n p_0, & 0 \le n \le R \\ C_n^K \dfrac{n!\,\rho^n}{R!\,R^{n-R}} p_0, & R \le n \le K \end{cases}$$

(Average number of breakdowns are proportional to the number of working machine.)

$$p_0 = \left(\sum_{n=0}^{R} C_n^K \rho^n + \sum_{n=R+1}^{K} C_n^K \frac{n!\,\rho^n}{R!\,R^{n-R}} \right)^{-1}$$

no closed form expression for $L_s = \sum_{n=0}^{K} n p_n$

$$\lambda_{\text{eff}} = E\{\lambda(K-n)\} = \lambda(K - L_s)$$

$$\lambda_{eff} = \sum_{n=0}^{K} \lambda_n p_n = \sum_{n=0}^{K} \lambda(K-n) p_n$$

$$= \sum_{n=0}^{K} \lambda K p_n - \sum_{n=0}^{K} \lambda n p_n$$

$$= \lambda K \sum_{n=0}^{K} p_n - \lambda \sum_{n=0}^{K} n p_n$$

$$= \lambda K - \lambda L_S$$

Toolco operates a machine shop with a total of 22 machines. Each machine is known to break down once every 2 hours, on the average. It takes an average of 12 minutes to complete a repair. Both the time between breakdowns and the repair time follow the exponential distribution. Toolco is interested in determining the number of repairpersons needed to keep the shop running "smoothly."

$$\lambda = \frac{1}{2} \text{ per hour}, \mu = \frac{60}{12} = 5 \text{ per hour}$$

Number of repair persons hired must substantially improve productivity

$$\begin{pmatrix}\text{Machines}\\ \text{productivity}\end{pmatrix} = \frac{\text{Available machines} - \text{Broken machines}}{\text{Available machines}} \times 100 = \frac{22 - L_s}{22} \times 100$$

c	Ls
1	12.0040
2	4.3677
3	2.4660
4	2.1001

Repairperson, R	1	$\boxed{2}$	3	4
Machines productivity (100%)	45.44	80.15	88.79	90.45
Marginal increase (100%)	—	34.71	8.64	1.66

Automata car wash facility operates with only one bay. Cars arrive according to a Poisson distribution with a mean of 4 cars per hour, and may wait in the facility's parking lot if the bay is busy. The time for washing and cleaning a car is exponential, with a mean of 10 minutes.

suppose that a new system is installed so that the service time for all cars is constant and equal to 10 minutes. How does the new system affect the operation of the facility?

$(M/G/1):(GD/\infty/\infty)$—POLLACZEK-KHINTCHINE (P-K) FORMULA

Let λ be the arrival rate at the single-server facility
the service time, t, is represented by any probability distribution with mean $E\{t\}$ and variance $\text{var}\{t\}$

it can be shown using sophisticated probability/Markov chain analysis that

$$L_s = \lambda E\{t\} + \frac{\lambda^2(E^2\{t\} + \text{var}\{t\})}{2(1 - \lambda E\{t\})}, \lambda E\{t\} < 1$$

$$p_0 = 1 - \lambda E\{t\} = 1 - \rho$$

Automata car wash facility operates with only one bay. Cars arrive according to a Poisson distribution with a mean of 4 cars per hour, and may wait in the facility's parking lot if the bay is busy. The time for washing and cleaning a car is exponential, with a mean of 10 minutes.

suppose that a new system is installed so that the service time for all cars is constant and equal to 10 minutes. How does the new system affect the operation of the facility?

$\lambda_{\text{eff}} = \lambda = 4$ cars per hour

The service time is constant

$E\{t\} = \frac{10}{60} = \frac{1}{6}$ hour and $\text{var}\{t\} = 0$

$$L_s = 4\left(\tfrac{1}{6}\right) + \frac{4^2\left(\left(\frac{1}{6}\right)^2 + 0\right)}{2\left(1 - \frac{4}{6}\right)} = 1.33 \text{ cars}$$

$$L_q = 1.333 - \left(\tfrac{4}{6}\right) = .667 \text{ cars}$$

$$W_s = \frac{1.333}{4} = .333 \text{ hour}$$

$$W_q = \frac{.667}{4} = .167 \text{ hour}$$

	$(M/M/1):(GD/\infty/\infty)$	$(M/D/1):(GD/\infty/\infty)$
W_s (hr)	.500	.333
W_q (hr)	.333	.167

Service time exponential with mean 10 minutes

Service time constant with mean 10 minutes

Queuing Cost Models

Cost models attempt to balance two conflicting costs:

1. Cost of offering the service.
2. Cost of delay in offering the service (customer waiting time)

Letting x (= μ or c) represent the *service level*, the cost model can be expressed as

$$ETC(x) = EOC(x) + EWC(x)$$

where

$$ETC = \text{Expected total cost } \textit{per unit time}$$

$$EOC = \text{Expected cost of operating the facility } \textit{per unit time}$$

$$EWC = \text{Expected cost of waiting } \textit{per unit time}$$

The simplest forms $EOC(x) = C_1 x$

$$EWC(x) = C_2 L_s$$

where

C_1 = *Marginal* cost per unit of x per unit time

C_2 = Cost of waiting per unit time per (waiting) customer

x can be anything that is responsible for increasing or decreasing cost of operating the service facility (Increasing the number of servers, adding a better server)

KeenCo Publishing is in the process of purchasing a high-speed commercial copier. Four models whose specifications are summarized below have been proposed by vendors.

i Copier model	C_{1i} Operating cost ($/hr)	Speed (sheets/min)
1	15	30
2	20	36
3	24	50
4	27	66

$\lambda = 4$ jobs/day

Jobs arrive at KeenCo according to a Poisson distribution with a mean of four jobs per 24-hour da
Job size is random but averages about 10,000 sheets per job. Contracts with the customers specif a penalty cost for late delivery of $80 per jobs per day. Which copier should KeenCo purchase?

$$\text{Average time per job} = \frac{10{,}000}{30} \times \frac{1}{60} = 5.56 \text{ hours} \rightarrow \boxed{\mu_1 = \frac{24}{5.56} = 4.32 \text{ jobs/day}}$$

$$ETC_i = EOC_i + EWC_i = C_{1i} \times 24 + C_{2i}L_{si} = 24C_{1i} + 80L_{si},\ i = 1, 2, 3, 4$$

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

Model i	λ_i (Jobs/day)	μ_i (Jobs/day)	L_{si} (Jobs)	EOC_i ($)	EWC_i ($)	ETC_i ($)
1	4	4.32	12.50	360.00	1000.00	1360.00
2	4	5.18	3.39	480.00	271.20	751.20
3	4	7.20	1.25	**576.00**	**100.00**	**676.00**
4	4	9.50	0.73	648.00	58.40	706.40

In a multiclerk tool crib facility, requests for tool exchange occur according to a Poisson distribution at the rate of 17.5 requests per hour. Each clerk can handle an average of 10 requests per hour. The cost of hiring a new clerk in the facility is \$12 an hour. The cost of lost production per waiting machine per hour is approximately \$50. Determine the optimal number of clerks for the facility.

$\lambda = 17.5$ requests per hour, $\mu = 10$ requests per hour, C_1 \$12 an hour, C_2 \$50 an hour

$(M/M/c):(GD/\infty/\infty)$

$$ETC(c) = C_1 c + C_2 L_s(c) = 12c + 50L_s(c)$$

c	$L_s(c)$ (requests)	ETC(c) (\$)
2	7.467	397.35
3	2.217	146.85
4	**1.842**	**140.10**
5	1.769	148.45
6	1.754	159.70