Clase 03

Teoría de Colas

Cola: Es una línea de espera.

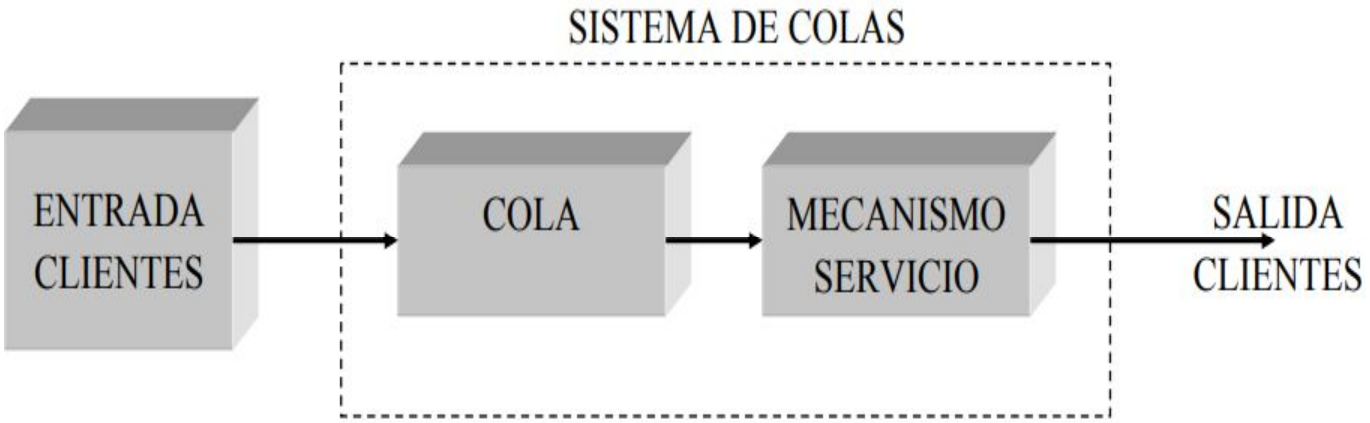
Línea de espera: Es una hilera formada por uno o varios clientes que aguardan para recibir un servicio.

Se produce por un desequilibrio temporal entre la demanda de un servicio y la capacidad del sistema para gestionarlo.

Un **cliente** puede ser un humano o piezas esperando su turno para ser procesadas o una lista de trabajo esperando para imprimir en una impresora en red o maquinas que requieren mantenimiento, etc.

Teoría de Colas

Proceso en una cola



"No importa en qué cola se sitúe: La otra siempre avanzará más rápido" (Primera Ley de Harper)

"Y si se cambia de cola, aquélla en la que estaba al principio empezará a ir más deprisa" (Segunda Ley de Harper)

Teoría de Colas

El estudio de las **colas** es importante porque proporciona tanto una base teórica del tipo de servicio que se puede esperar de un recurso, como la forma en la cual dicho recurso puede ser diseñado para proporcionar un grado de servicio a sus clientes.

Situación	Llegadas	Cola	Mecanismo de Servicio
Aeropuerto	Pasajeros	Sala de espera	Avión
Dpto. de bomberos	Alarmas de incendio	Incendios	Dpto. De Bomberos.
Compañía telefónica	Números marcados	Llamadas	Conmutador
Panadería	Clientes	Clientes con números	Vendedor
Carga de camiones	Camiones	Camiones en espera	Muelle de carga
Oficina de correos	Cartas	Buzón	Empleados de correos
Fábrica	Piezas para ensamblar	Inventario en proceso	Estación de trabajo.
Hospital	Pacientes	Personas enfermas	Médicos

Teoría de Colas

Teoría de colas: Es el estudio matemático del comportamiento de líneas de espera que se encarga de proponer modelos para el manejo eficiente de ellas.

El origen de la teoría de colas está en el esfuerzo de **Agner Kraup Erlang** (Dinamarca, 1878 - 1929) en 1909 para analizar la congestión de tráfico telefónico con el objetivo de cumplir la demanda incierta de servicios en el sistema telefónico de Copenhague.

Teoría de Colas

Sistemas de colas: Esta compuesto por clientes que son aquellos que solicitan servicios que se disponen en un área de espera ordenada y un servidor que presta un servicio.

Tipos de sistemas de colas

- ✓ Una cola y un servidor (garita de seguridad).
- ✓ Una cola y múltiples servidores (cajero automático).
- ✓ Varias colas y un servidor (farmacia).
- ✓ Varias colas y múltiples servidores (supermercado).
- ✓ Una cola y servidores secuenciales (examen preocupacional)

Teoría de Colas

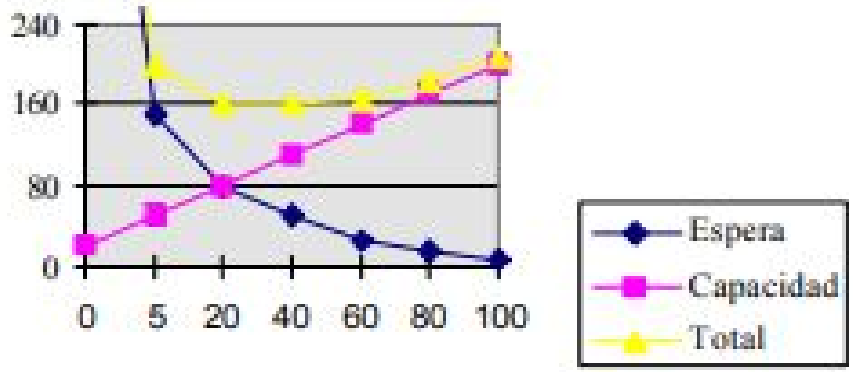
Elementos comunes en toda línea de espera

- ✓ Una población de clientes que genera clientes potenciales.
- ✓ Una línea o fila de espera formada por los clientes.
- ✓ La instalación del servicio, forma por una persona o personas, una maquina o maquinas que se requiera para proveer el servicio que el cliente solicita.
- ✓ Una regla de prioridad para seleccionar al siguiente cliente que será atendido por la instalación del servicio.

Teoría de Colas

Costes asociados a un sistema de colas

- ✓ Los costes asociados a la espera de los clientes.
- ✓ Los costes asociados a la expansión de la capacidad de servicio.
- ✓ Los costes totales del sistema de servicio (suma de los costes anteriores).



Teoría de Colas

Características de los sistemas de colas

- ✓ Patrón de llegada de los clientes (como es la llegada de los clientes).
- ✓ Patrón de servicio de los servidores (como se atiende).
- ✓ Disciplina de cola (como se seleccionan los clientes para ser atendidos).
- ✓ Capacidad del sistema (numero máximo de clientes que pueden estar dentro del sistema).
- ✓ Número de canales de servicio (un canal o canales de servicio en paralelo).
- ✓ Número de etapas de servicio (una o varias etapas de servicio).

Teoría de Colas

En los sistemas de colas interesa distinguir la distribución de probabilidad de los arribos, la distribución de probabilidad de los tiempos que el/los servidor/res tarda/n con cada uno de los clientes (tiempos que por lo general son todos diferentes ya que depende del servicio que solicite el cliente y también del propio cliente), la cantidad de servidores y en ocasiones se necesita alguna otra aclaración. Para ello se incorporo la notación de Kendall.

Notación de Kendall: Es una forma sintética de describir el sistema de colas y mostrar sus características pudiendo clasificar los diferentes tipos de colas por medio de iniciales y un numero.

Teoría de Colas

Por ejemplo

M / M / 1

- **Número de servidores**
- **Distribución de probabilidad de los tiempos de servicios**
- **Distribución de probabilidad de los arribos**

A / B / X / Y / Z

A: indica la distribución de tiempo entre llegadas consecutivas

B: alude al patrón de servicio de servidores

X: es el número de canales de servicio

Y: es la restricción en la capacidad del sistema

Z: es la disciplina de cola

Teoría de Colas

Simbología de la notación

Característica	Símbolo	Explicación
Distribución de tiempos de llegada (A) Distribución de tiempos de servicio (B)	M	Exponencial
	D	Determinista
	Ek	Erlang tipo-k (k=1,2,...)
	Hk	Mezcla de k exponenciales
	PH	Tipo fase
	G	General
Número de servidores	1,2,...,∞	
Disciplina de cola	FIFO	Servir al primero que llega
	LIFO	El último que llega se sirve primero
	RSS	Selección aleatoria de servicio
	PR	Prioridad
	GD	Disciplina general

Teoría de Colas

Sistemas M/M/1: Son los sistemas mas sencillos.

La primera **M** indica que los arribos son Markovianos, es decir que se distribuyen Poisson.

La segunda **M** indica que los tiempo de servicio son Markovianos, es decir que tienen distribución exponencial.

El numero **1** indica que existe un solo servidor.

¿Cómo es que los arribos y los tiempos de servicio tienen distinta distribución y ambos son Markovianos?

Teoría de Colas

Los arribos son una variable discreta cuyos sucesos, puntos aislados en el tiempo, se distribuyen Poisson.

Siempre que una variable ocurre como puntos aislados en el tiempo y los tiempos entre ocurrencia y ocurrencia son irregulares, distintos entre si, esos tiempos tienen distribución exponencial.

Con los tiempos de servicio ocurre lo mismo. Los tiempos de servicios corresponden a una variable continua, aleatoria, cuyos valores están comprendidos entre dos puntos.

Esa variable continua se distribuye exponencial y por lo tanto se dice que es Markoviana.

Teoría de Colas

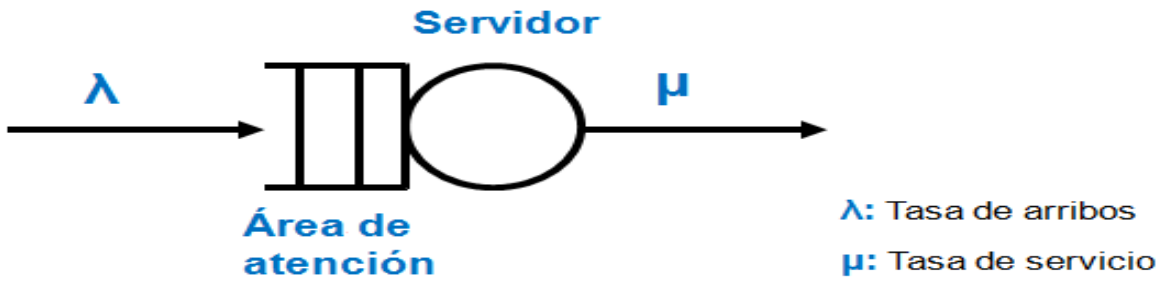
Características de una M/M/1

- ✓ Los arribos se distribuyen Poisson.
- ✓ Los tiempo de servicios tienen distribución exponencial.
- ✓ Hay un solo servidor con cola única
- ✓ Las salidas son independiente de las entradas.

  Significa que independientemente de la llegada de los clientes, el servidor trabaja a su velocidad, no toma en cuenta a que velocidad que llegan los clientes ni cuantos hay en la cola.
- ✓ La disciplina de atención es FIFO.

Teoría de Colas

Esquema básico



Tasa de arribos (λ): Es la velocidad promedio de llegada de los clientes. Los clientes llegan en intervalos de irregulares. Su unidad es: nros de clientes que arriban / unidad de tiempo. Por ejemplo **λ = 3 cli / seg**.

Tasa de servicio (μ): Es el número de clientes que en promedio atiende el servidor por unidad de tiempo. Los clientes son atendidos en intervalos irregulares y el servidor no considera tiempos ociosos. Su unidad es: nros de clientes atendidos / unidad de tiempo. Por ejemplo **μ = 6 cli / seg**.

Teoría de Colas

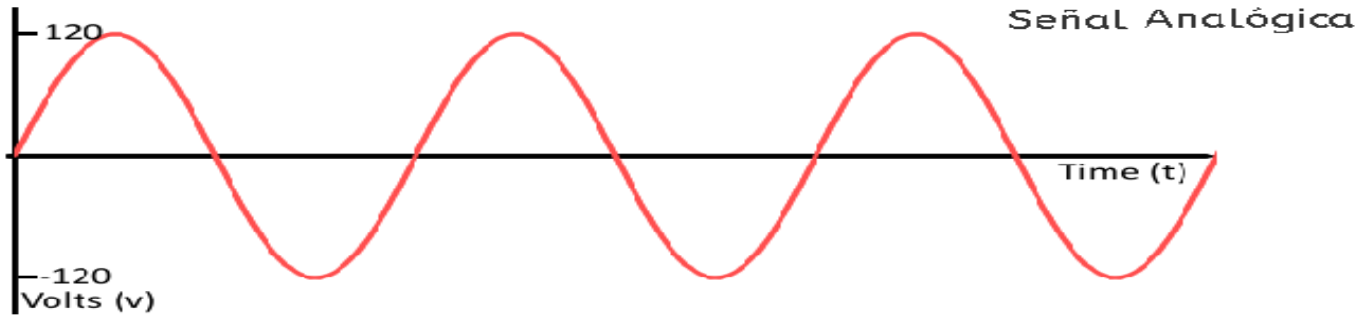
Las **tasas** **no** tienen **distribución** de **probabilidad**

Tiempo medio de servicio: Es el tiempo que, en promedio, tarda el servidor con cada cliente.

ts = Tiempo de servicio para un cliente determinado

Ts = Tiempo medio de servicio

Por analogía



Teoría de Colas

T (periodo de tiempo) = Es el tiempo que tarda en producirse una onda completa. Por ejemplo T = 0,5 seg / ciclo.

f (frecuencia) = Es la cantidad de ciclos que se producen por unidad de tiempo. Por ejemplo f = 20 ciclos / seg.

La frecuencia es la inversa del periodo (T = 1 / f).

Existe una analogía entre el tiempo medio de servicio y el periodo al igual que la tasa de servicio y la frecuencia.

Si T = 1 / f entonces **Ts = 1 / μ.**

Por ejemplo Ts = 1 / (9 cli / seg).

Teoría de Colas

Utilización del sistema (ρ): Es la relación que existe entre la tasa de arribos y la tasa de servicio. Permite evaluar el buen funcionamiento del sistema. No tiene unidades. **$\rho = \lambda / \mu$**. Por ejemplo λ = 10 cli / seg, μ = 20 cli / seg entonces $\rho = \lambda / \mu$ = 10 cli/seg / 20 cli/seg = 0,5.

Parámetros: Son las constantes que caracterizan al sistemas. Por ejemplo λ y μ son unidades constantes y son los parámetros del sistema de colas.

Congestionamiento: Se dice que un sistema se congestiona si y solo si la cola tiende a infinito, es decir crece indefinidamente.

Teoría de Colas

Una **M/M/1** se **congestiona** si solo si **ρ ≥ 1**

De esta afirmación surgen dos casos:

Caso 1 Si **ρ > 1**, es **λ > μ,** es decir, que los clientes llegan más rápidos de lo que el servidor puede atender, eso hace que la cola crezca infinitivamente por lo tanto el sistema se congestiona.

Caso 2 Si **ρ = 1**, es **λ = μ,** los clientes no llegan a intervalos regulares sino que llegan por ráfagas (la tasa de arribos, es un valor promedio); eso provoca tiempos ociosos en el servidor que no pueden recuperarse.

Por otra parte, la tasa de arribos y la tasa de servicio están calculadas considerando tiempos diferentes.

Teoría de Colas

Para el calculo de la tasa de arribos se toma en cuenta el tiempo total de funcionamiento del sistema (T total). Para el calculo de la tasa de servicio solo se toman en cuenta los tiempos de efectiva ocupación del servidor, (T ocupación) que es la suma de los tiempos en que el servidor esta ocupado atendiendo clientes, que siempre es menor que el tiempo total porque al principio se producen tiempos ociosos.

Téngase en cuenta que la tasa de servicio no es la verdadera tasa de salida del sistema. La tasa de servicio esta tomando en cuenta el trabajo del servidor; la tasa de salida se calcula observando la salida de los clientes desde afuera, como si los clientes fuesen atendidos en la ventanilla de un banco y un observador, que esta afuera del banco

Teoría de Colas

calculara la velocidad a que los clientes salen, considerando el tiempo total.

En base a las condiciones precedentes, resulta:

$$\lambda = \text{Nº clientes ingresados}/T_{total}$$

$$\mu = \text{Nº clientes atendidos}/T_{ocupacion}$$

Se partió de $\lambda = \mu$

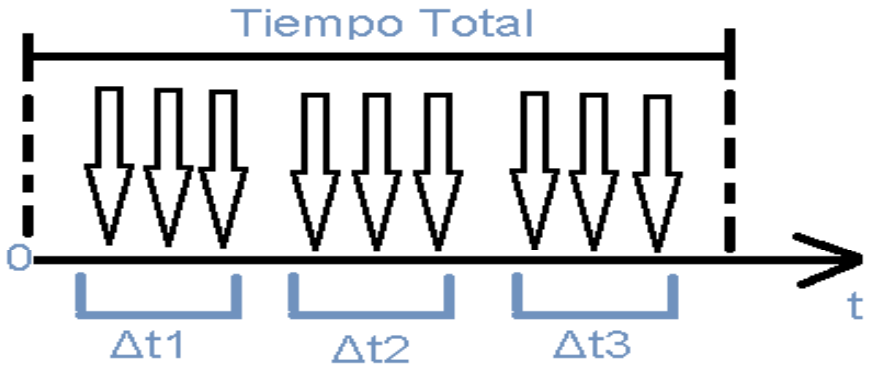
Teniendo en cuenta que el tiempo total (T total) es mayor que el tiempo de ocupación (T ocupación), Nº clientes ingresados debe ser mayor que Nº clientes atendidos, para que se cumpla la igualdad.

En consecuencia el sistema se congestiona.

Teoría de Colas

Tasa de Salida ≠ **Tasa de Servicio**

(Toma tiempos muertos) (No toma tiempos muertos)



$$\lambda = \frac{\text{nro cli arribos}}{\text{T Total}} \qquad \mu = \frac{\text{nro cli atendidos}}{\Delta t1 + \Delta t2 + \ldots + \Delta tn}$$

$$\wedge \text{ T Total} \quad \sum_{i=1}^{n} \Delta ti \wedge \lambda = \mu \Rightarrow \text{nro cli arribos}$$

$$\Rightarrow \text{se congestiona porque } \lambda = \mu > \text{tasa de salida}$$

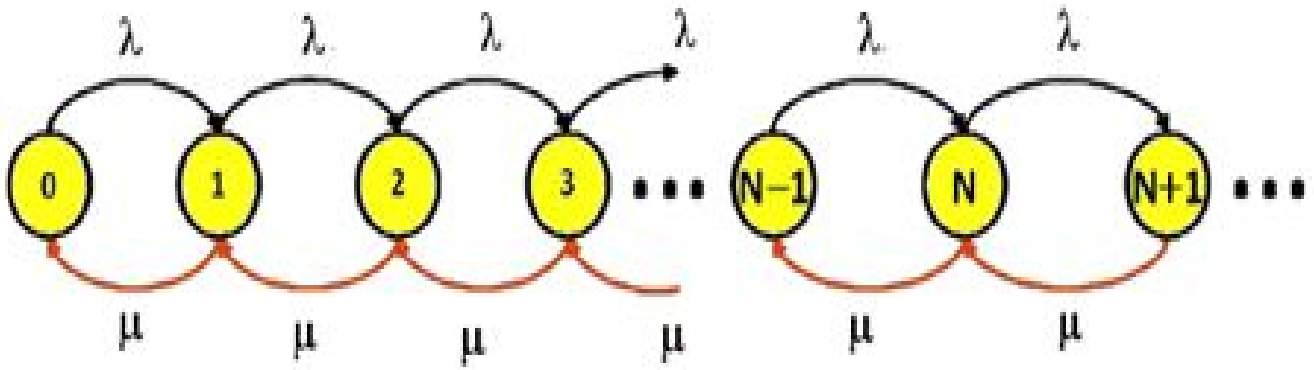
$$\Rightarrow \lambda > \text{tasa de salida}$$

Sucede porque los **tiempos muertos** **no** se **recuperan** nunca

Teoría de Colas

Estado de un sistema: Es el valor que toma las variables de un sistema en un momento determinada. En un sistema de colas el estado es el numero de clientes que hay en un momento determinado en el sistema.

Diagrama de estado de un M/M/1



Teoría de Colas

Los círculos con sus números representan los diferentes estados en que puede estar el sistema. El menor estado posible es 0, no puede haber un numero negativo de clientes.

Las flechas de arriba, que van de izquierda a derecha representan el arribo de un cliente y en promedio los clientes arriban con tasa λ.

Las flechas de abajo que van de derecha a izquierda representan la salida de un cliente y en promedio los clientes son procesados con tasa μ

Si el sistema esta en estado 0, para pasar al estado 1 llega un cliente y lo hace con tasa λ.

Si el sistema esta en estado 1 para pasar al estado 0 tiene que salir un cliente y eso lo hace con tasa μ. Y así siguiendo…….

Teoría de Colas

Equilibrio: Un sistema se encuentra en equilibrio si solo si el valor de sus variables permanece constante atreves del tiempo.

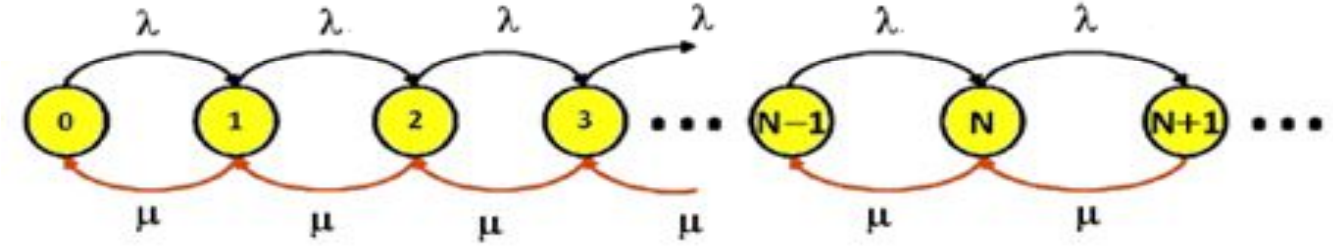
Equilibrio estático: No se producen cambios cuantitativos ni cualitativos a través del tiempo. La velocidad de entrada es igual a la velocidad de salida y ambas son iguales a 0. Nada sale del sistema y nada entra al sistema.

Equilibrio dinámico: No se producen cambios cuantitativos pero si cualitativos a través del tiempo. La velocidad de entrada es igual a la velocidad de salida, ambas son distintas de 0. Cada vez que entra un elemento, al mismo tiempo se produce la salida de otro, lo cual hace que la cantidad de elementos no cambie, a pesar que los elementos, al cabo del tiempo, no van a ser los mismos.

Teoría de Colas

Ecuación de estado estable o Steady State

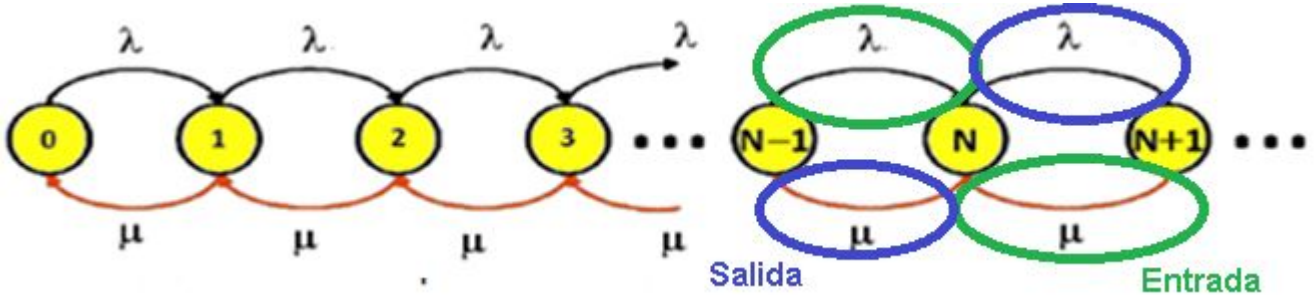
En el siguiente diagrama n es un estado de equilibrio dinámico. Los estados n - 1 y n +1 son el anterior y el posterior respectivamente.



Las flechas encerradas en un circulo verde constituyen las entradas al estado n; es decir todo lo que conduce al estado n.

Las flechas encerradas en un circulo azul constituyen las salidas de n, es decir todo lo que me aparta del estado n.

Teoría de Colas



Por tratarse de un estado de equilibrio dinámico, las entradas son iguales a las salidas. Por lo tanto:

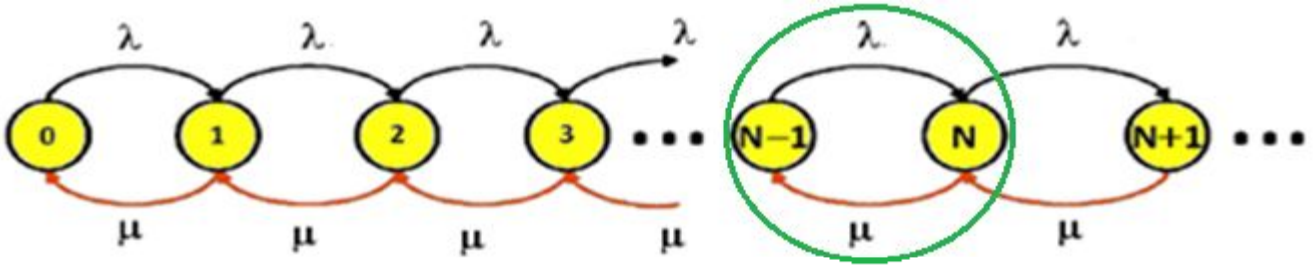
πn = Probabilidad del estado n

Entradas	=	Salidas
$\lambda . \pi n-1 + \mu . \pi n+1$	=	$\lambda . \pi n + \mu . \pi n$
$\lambda . \pi n-1 + \mu . \pi n+1$	=	$\pi n\ (\lambda + \mu)$
$\lambda . \pi n-1 + \mu . \pi n+1$	=	πn
$\dfrac{\lambda . \pi n-1 + \mu . \pi n+1}{(\lambda + \mu)}$	=	πn

⇒ **Ecuación de Estado Estable**

Teoría de Colas

Ecuación general de estado



No considera tres estados sino dos. Si asumimos que entre esos dos estados hay un equilibrio dinámico, es decir que solo consideramos las interacciones entre el estado n - 1 y el estado n, y como simplificación, consideramos que no interactúan con ningún otro estado, entonces todo lo que pasa desde n - 1 hacia n por unidad de tiempo, es igual a todo lo que pasa de n hacia n - 1 en el mismo tiempo.

Teoría de Colas

Entonces:

$$\lambda \cdot \pi_{n-1} = \mu \cdot \pi_n$$

$$\frac{\lambda}{\mu} \cdot \pi_{n-1} = \pi_n$$

$$\rho \cdot \pi_{n-1} = \pi_n$$

π_n = Probabilidad del estado n

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

Ecuación General de Estado

n - 1 es anterior a **n** entonces concluimos que siempre se puede calcular la probabilidad de un estado anterior como **ρ** por la probabilidad del estado anterior. Por ejemplo si se quiere calcular **π 5** multiplicamos **ρ** por **π 4.**

Teoría de Colas

Deducción de la formula $\pi_n = f(n)$

Las probabilidades solo admiten valores mayores o iguales que 0 y menores o iguales que 1.

Para el caso en que $\rho \leq 1$ donde ρ es la probabilidad que el sistema este siendo utilizado. Lo contrario es que el sistema no este siendo utilizado, es decir que el sistema este ocioso (que este ocioso el servidor) y eso tiene, en consecuencia, probabilidad $1 - \rho$, es decir $\pi_0 = 1 - \rho$.

Basándose en las expresiones $\pi_0 = 1 - \rho$ y $\pi_n = \rho \cdot \pi_{n-1}$ se deduce la formula de $\pi_n = f(n)$.

Teoría de Colas

$$\pi_0 = 1 - \rho$$

Estado anterior

$$\pi_1 = \rho \cdot \pi_0 = \rho\,(1 - \rho)$$

$$\pi_2 = \rho \cdot \pi_1 = \rho \cdot \rho\,(1 - \rho) = \rho^2\,(1 - \rho)$$

$$\pi_3 = \rho \cdot \pi_2 = \rho \cdot \rho^2\,(1 - \rho) = \rho^3\,(1 - \rho)$$

$$\vdots$$

$$\pi_n = \rho^n\,(1 - \rho)$$

$$\pi_n = \rho \cdot \pi_{n-1}$$

$$\pi_0 = 1 - \rho$$

Teorema de Little

Sirve para hallar el numero medio de clientes en el sistema (en una M/M/1) conociendo el tiempo medio de permanencia de los clientes en el sistema y la tasa de arribos.

Teoría de Colas

En toda M/M/1 el numero medio de clientes en el sistema es directamente proporcional al tiempo medio de permanencia de los clientes en el sistema siendo el factor de proporcionalidad la tasa de arribos.

$$N = \lambda . W$$

W = Tiempo medio de permanencia de los clientes en el sistema.

N = Numero medio de clientes en el sistema.

λ = Tasa de arribos.

Preguntas