**Teoría de Colas/ Sistema de espera**

Estudia el comportamiento de las líneas de espera. Se utiliza para analizar sistemas donde un cliente demanda un servicio y este debe ser atendido por uno o más servidores. Cuando el servicio no puede ser satisfecho de forma inmediata, los clientes deben esperar, formando así una cola.

* Ejemplos **reales**: Física: Bancos, Hospitales, Supermercados, Peajes, Industrias o líneas de producción, etc.
* Ejemplos en el contexto de la **informática**: Procesamiento por CPU, Routers y switches de red, Colas de impresión, etc.

Objetivos:

* Minimizar el tiempo que los clientes pasan en el sistema → Se beneficia al que recibe el servicio
* Minimizar los costos de quienes prestan el servicio → Se beneficia el Servidor

*La diferencia con una red de colas es que un sistema de colas es un sistema aislado. Tiene un solo punto donde los clientes llegan, esperan y son atendidos (cola simple). Una red de colas es una interconexión de varios sistemas de colas: los clientes pueden pasar de una cola a otra.*

**Características de los sistemas de colas**

1. Régimen de llegada de los clientes (A):

* **Estocástico (aleatorio):** no son predecibles. Se modelan con variables aleatorias que siguen una distribución de probabilidad. La distribución más común para este tipo es M (Exponencial o Poisson) con llegadas totalmente al azar, sin memoria (markoviana).
* **Programado (Determinista)**: Las llegadas se producen a intervalos fijos
* **General (G):** No se asume ninguna forma específica. Se usa cuando no se conoce la distribución exacta o es muy variable.

Consideraciones adicionales:

* **Impaciencia:** algunos clientes se van si la cola es muy larga o si esperan demasiado tiempo
* Variabilidad en el tiempo:
  + Si la tasa de llegada no cambia, el sistema es estacionario.
  + Si cambia según la hora o condiciones, es no estacionario.

1. Régimen de servicio de los servidores (B):

* Duración del servicio:
  + Determinista: todos los clientes son atendidos en el mismo tiempo.
  + Variable (aleatorio): cada cliente puede tardar más o menos.
  + Debido a esta variación, el régimen de servicio también se modela con una distribución de probabilidad, como:
    - determinista: tiempo fijo
    - exponencial: tiempo aleatorio, sin memoria (*markoviano*).
    - general: patrón irregular.
* La forma de atención puede ser individual o en grupo.

1. Número de canales de servicio (c):

* Sistemas Monocanal (M/M/1): Hay un solo servidor en el sistema. Los clientes hacen una cola única y son atendidos uno por uno.
* Sistemas Multicanal (M/M/c): Hay varios servidores en paralelo. Puede haber:
  + - Una cola única común que alimenta a varios servidores
    - Una cola para cada servidor.

1. Número de etapas de servicio (No forma parte de la notación Kendall):

* Unietapa (Teoría de colas): El cliente pasa una sola vez por una cola y un servidor. Una vez atendido, sale del sistema.
* Multietapa (Redes de cola): El cliente pasa por varias etapas secuenciales (varias colas y servidores). Puede haber reciclado: volver a una etapa anterior si hay errores.

1. Disciplina de cola (d):

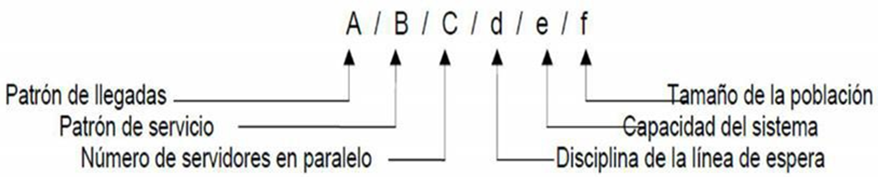
Define el orden de atención. Puede ser:

* FIFO: primero en llegar, primero en salir.
* LIFO: último en llegar, primero en salir.
* RSS: selección aleatoria de servicio.
* Por prioridad: El cliente más importante puede interrumpir al que está siendo atendido.

1. Capacidad del sistema (e|n|k): Es la cantidad máxima de clientes permitidos en la cola y servicio. Si es finita, los nuevos clientes pueden ser rechazados, lo que refleja limitaciones o impaciencia.
2. Población posible de clientes (f): Considera si hay una cantidad finita de clientes posibles, como en sistemas cerrados o limitados. Suele ser opcional.

**Notación de Kendall**

Sirve para describir de forma estandarizada un sistema de colas.



**MM1N**

**✅ 2. (TEÓRICA) Explique la dinámica de un sistema M/M/1/N**

Un sistema **M/M/1/N** es un modelo de teoría de colas donde:

* Las llegadas siguen una **distribución de Poisson** (la primera M, por *Markoviana*).
* El tiempo de servicio sigue una **distribución exponencial** (segunda M).
* Hay **1 único servidor**.
* El sistema tiene una **capacidad máxima de N clientes** (incluye los que esperan y el que está siendo atendido).

**📌 Dinámica del sistema:**

* Los clientes llegan al sistema a una **tasa λ**.
* Si el sistema tiene **menos de N clientes**, el nuevo cliente **entra**.
* Si hay **N clientes en el sistema**, el nuevo cliente es **rechazado** y no entra (esto genera una **tasa de pérdida** o **tasa de rechazo**).
* El servidor atiende a una tasa **μ**.
* El comportamiento del sistema depende de **ρ = λ / μ** (factor de utilización) y del valor de **N**.

Este modelo se usa en situaciones donde hay **recursos limitados**, como cajeros automáticos, estacionamientos o líneas telefónicas.

**✅ 3. ¿Cuáles son las principales desventajas de una alta tasa de rechazo?**

Una **alta tasa de rechazo** en un sistema M/M/1/N indica que **muchos clientes no pueden ingresar** al sistema por estar lleno. Esto trae varias desventajas:

* 🔸 **Pérdida de ingresos**: Cada cliente rechazado representa una oportunidad de venta o servicio perdida.
* 🔸 **Disminución de la satisfacción del cliente**: La imposibilidad de ingresar puede generar molestias y dañar la imagen del servicio.
* 🔸 **Ineficiencia del sistema**: Significa que la capacidad del sistema no es suficiente para la demanda, lo que puede requerir rediseño o inversión.
* 🔸 **Desviación de clientes hacia la competencia**: Si los rechazos son frecuentes, los usuarios buscarán otras opciones.

**Nota:** A veces en este modelo en vez de ser M es m o k.

**Diferencia de características**

|  |  |  |
| --- | --- | --- |
|  | **M/M/1** | **M/M/1/N** |
| Capacidad del sistema | Ilimitada: no hay restricción en la cantidad de clientes en espera. | Limitada: el sistema admite un máximo de N clientes (en cola + en servicio). |
| Tasa de llegada (λ) | Constante y siempre se acepta cada cliente que llega. | Constante teóricamente, pero no se aceptan más clientes si el sistema está lleno. |
| Tasa de servicio (μ) | Constante. Se mantiene mientras haya clientes en el sistema. | Es constante, siempre que haya al menos un cliente. |
| Rendimiento del sistema | Máximo rendimiento teórico posible, ya que no hay pérdidas. | El rendimiento es menor, ya que puede haber pérdida de clientes al estar lleno el sistema. |
| Congestión del sistema | Puede aumentar sin límite si λ ≥ μ, generando largas colas. | La congestión se limita artificialmente al máximo de N clientes, pero eso puede provocar rechazo de nuevos clientes. |
| Equilibrio (estabilidad del sistema) | Se alcanza si λ < μ, ya que el sistema tiene la capacidad de vaciarse. | También requiere que λ < μ para alcanzar equilibrio, pero hay una mayor sensibilidad a la saturación debido al límite N. |
| Aplicación típica | Sistemas donde se pueden tener muchas personas en espera sin afectar el servicio (por ejemplo, atención virtual, chats, etc.). | Sistemas con capacidad física o lógica limitada o donde no se puede permitir acumulación de clientes (ej. salas de espera, buffers limitados). |

**SISTEMAS DEPENDIENTES DE ESTADOS**

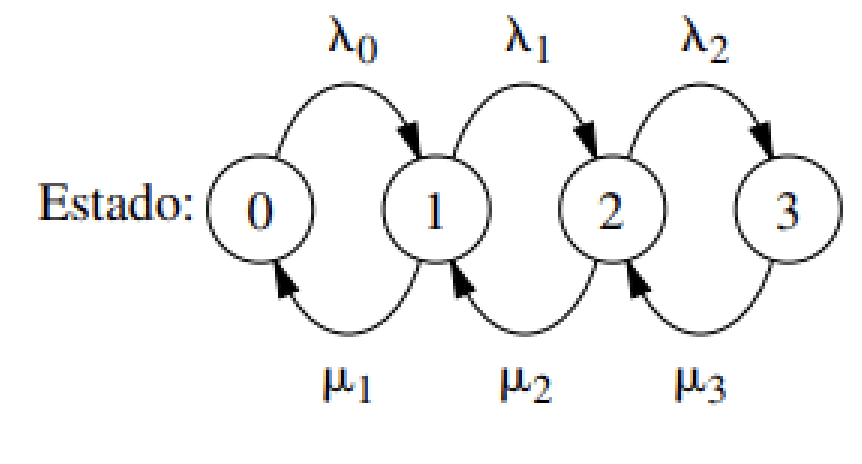
Son modelos matemáticos que demuestran cómo cambian algunos sistemas con el tiempo. Se les llama "de nacimiento-muerte" porque simulan ingresos (nacimientos) y salidas (muertes) del sistema.

Se le dice dependientes de estado porque el comportamiento del sistema depende de cuántos elementos hay dentro del sistema en ese momento.

Tasas/Medición de ritmo de estado de cambio en el sistema:

* **Lambda** (λ): la frecuencia con la que nuevos elementos ingresan al sistema.
* **Mu** (μ): la frecuencia con la que los elementos salen del sistema

**Diagrama de estados**

Un diagrama de estado cuenta con los siguientes componentes:

1. **Estados**: Representan las diferentes situaciones del sistema en un momento dado.

2. **Transiciones**: Son las flechas que conectan los estados, mostrando cómo el sistema pasa de un estado a otro. Las transiciones ocurren debido a un evento o acción.

3. **Eventos**: Son los factores que causan las transiciones entre los estados.

4. **Condiciones de transición**: Especifican cuándo ocurre una transición, basándose en la ocurrencia de un evento o el cumplimiento de una condición.

5. **Estado inicial:** Es el estado en el que comienza el sistema.

6. **Estado final:** Es el estado en el que el sistema termina o se estabiliza.

**Cálculo de probabilidades**

Un proceso de Markov es un modelo matemático que describe un sistema que evoluciona en el tiempo siguiendo reglas probabilísticas muy específicas. El futuro del sistema depende solo del estado actual, no del pasado.

**Estado estacionario**

Si las probabilidades no cambian con el tiempo, es decir, la probabilidad de estar en el estado n no cambia ya que la tasa de entrada es igual a la de salida.

**Sistema de colas con prioridades**

Un sistema de colas con prioridades es aquel en el que no todos los clientes son tratados por igual.

Tipos de Prioridades

1. Sin Interrupción de Servicio: la prioridad solo afecta el orden de ingreso a la atención. Esto significa que el cliente con menor prioridad debe esperar a que la persona con mayor prioridad termine su atención antes de ser atendido. Sin embargo, una vez que un cliente ha comenzado a ser atendido, no se interrumpe su servicio para dar paso a otro cliente prioritario.

2. Con Interrupción de Servicio: si llega un cliente de mayor prioridad, puede interrumpir el servicio de un cliente de menor prioridad que ya está siendo atendido.

**Sistemas M/M/2**

Un sistema **M/M/2** es un modelo de colas con **dos servidores en paralelo**, donde las llegadas de clientes tienen una distribución de poisson y los **tiempos de servicio una distribución exponencial**. Cada cliente es atendido por el **primer servidor disponible**.

Ventajas:

* Mayor capacidad de atención.
* Reducción del tiempo en cola.
* Análisis matemático sencillo gracias a las propiedades de la distribución exponencial.

Limitaciones:

* Supone que ambos servidores son idénticos en velocidad, lo cual no siempre se cumple en la práctica.
* En caso de velocidades distintas, el análisis se complica y puede requiere simulación para su entendimiento.
* Si la tasa de llegada supera la capacidad conjunta de servicio (2μ ≤ λ), el sistema se vuelve inestable y la cola crecerá indefinidamente.

Pero en la práctica, los servidores pueden tener diferentes velocidades.

* Si los dos tienen la misma velocidad, el sistema es equilibrado.
* Si uno es más rápido, lo ideal sería que atienda más clientes para mejorar la eficiencia.
* Si se asignan clientes sin tener en cuenta las diferencias, puede generarse más espera en el servidor lento.

**Sistema M/D/1**

Reduce la complejidad de los cálculos y permite una descripción más precisa de algunos sistemas reales.

Ventajas:

● Más predecible: al tener tiempos de servicio constantes, es más fácil prever los tiempos de espera.

● Menor variabilidad: al eliminar la incertidumbre del servicio, el sistema es más estable.

● Simplicidad en el análisis: permite cálculos más sencillos en comparación con modelos con distribución general.

**Sistemas M/G/1**

Características del sistema M/G/1:

* Un solo servidor, al igual que el modelo M/M/1.
* La llegada de clientes sigue una distribución de Poisson.
* Los tiempos de servicio siguen una distribución general (“G”), no exponencial.
* Mayor variabilidad en los tiempos de atención (más realista).

Ventajas:

* Representa mejor escenarios reales.
* Flexible para modelar servicios con alta variabilidad.

Limitaciones:

Su análisis es más complejo que el de los modelos con distribución exponencial.

No siempre se obtienen soluciones exactas y, en muchos casos, se recurre a simulación.