FACULTAD DE TECNOLOGÍA INFORMÁTICA

INGENIERÍA EN SISTEMAS INFORMÁTICOS

MODELIZACIÓN NUMÉRICA

Sistemas de colas

TEMARIO:

- Sistemas dependientes de estado (de nacimiento-muerte):
  - Concepto
  - Diagrama de estados
  - Cálculo de probabilidades de los distintos estados
- Sistemas:
  - M/M/2
  - M/G/1
  - M/D/1
  - Sistema de colas con prioridad

GRUPO 6:

- Ricardo, Agustin
- Bovyn, Nehuen
- Dieguez, Gisela
- Tagliarini, German
- Rodriguez, Aylen

Introducción

Requisitos previos:

- ¿Qué es un sistema de colas? – lo explicó otro grupo

Sistemas dependientes de estado (de nacimiento-muerte)

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. Lo importante de este modelo es que las tasas de entrada y salida dependen de la cantidad de elementos que hay en ese momento. Es decir, el comportamiento del sistema cambia según su estado actual.

¿Qué significa dependientes de estado?

Significa que el comportamiento del sistema (como las tasas de entrada o salida) depende de su estado actual, es decir, de cuántos elementos hay dentro del sistema en ese momento. Por ejemplo, en un modelo de población:

Cuantas más personas haya, más nacimientos y muertes puede haber, porque hay más personas que pueden nacer o morir. Así, el ritmo de cambio no es fijo, sino que varía según cuántas personas haya en ese instante.

Para representar este comportamiento, se utilizan dos tasas:

- **Lambda (λ):** representa la **tasa de nacimientos**, es decir, la frecuencia con la que nuevos elementos ingresan al sistema.

- **Mu (μ):** representa la **tasa de muertes**, es decir, la frecuencia con la que los elementos salen del sistema.

Estas tasas pueden variar según cuántos elementos haya en el sistema en cada momento.

Ejemplo en la realidad

Tomando como ejemplo comparativo la evolución de la población entre **Argentina** y **China**.

Tanto Argentina como China pueden analizarse como sistemas en los que la población varía a lo largo del tiempo debido a flujos de entrada (nacimientos) y salida (muertes). Este tipo de evolución se modela adecuadamente mediante un *proceso de nacimiento y muerte dependiente del estado*, donde las tasas de ocurrencia de los eventos no son constantes, sino que dependen directamente del estado actual del sistema, es decir, de la cantidad de individuos presentes.

Pais	**Tasa de natalidad (por c/ 1000 hab.)**	**Tasa de mortalidad (por c/ 1000 hab.)**	**Población estimada**
Argentina	9,8	7,7	46 millones
China	6,4	7,9	1.400 millones

Datos demográficos comparativos

Interpretación

Argentina, con una población moderada, exhibe tasas de natalidad y mortalidad relativamente equilibradas, lo que genera un crecimiento poblacional leve. Sin embargo, dado el tamaño del sistema, las tasas absolutas de cambio son también limitadas: pocos nacimientos y muertes en términos absolutos.

En contraste, China, a pesar de presentar tasas de natalidad más bajas que Argentina, muestra un volumen absoluto mucho mayor de nacimientos y defunciones debido a su gran población. Este fenómeno evidencia el principio clave del modelo: el estado actual del sistema (población total) actúa como un multiplicador de las tasas, amplificando los efectos demográficos.

Conclusión

En un proceso de nacimiento y muerte dependiente del estado, la dinámica del sistema no está determinada únicamente por las tasas relativas, sino también por el estado del sistema en cada momento. Una población reducida genera, aún con tasas moderadas, pocos eventos demográficos. Por el contrario, un sistema de gran escala, como el caso de China, convierte incluso tasas bajas en efectos demográficos de gran magnitud. Por lo tanto, el estado del sistema no sólo influye, sino que define la magnitud del comportamiento observado.

Diagrama de estado

Es una herramienta visual que muestra cómo cambia un sistema con el tiempo. Cada **estado** representa una situación del sistema en un momento dado, y las **transiciones** muestran cómo el sistema pasa de un estado a otro debido a eventos que ocurren.

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.

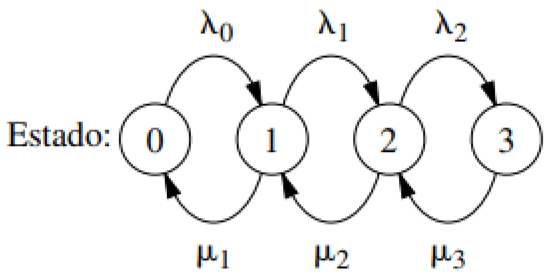
3. **Eventos**: Son los factores que causan las transiciones entre los estados. Por ejemplo, el **nacimiento** o **muerte** de una persona.

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.

Estos componentes trabajan juntos para mostrar cómo un sistema cambia y evoluciona a lo largo del tiempo.



En el caso de las **poblaciones de Argentina y China**, podemos ver cómo nacimientos y muertes afectan el tamaño de la población. Estas dinámicas se pueden modelar como un proceso de **nacimiento-muerte dependiente del estado**, donde el número de nacimientos y muertes depende del tamaño actual de la población.

Ejemplo de diagrama de estado

Estados y Transiciones:

- **Estados:** Son los diferentes niveles de población. Por ejemplo:
    - **Estado 0**: Poca población.
    - **Estado 1**: Población moderada.
    - **Estado 2**: Población más alta.
- **Transiciones:**
    - **Nacimiento**: Cuando alguien nace, la población aumenta (pasa a un estado con más personas).
    - **Muerte**: Cuando alguien muere, la población disminuye (pasa a un estado con menos personas).

Probabilidades y Efectos:

Las transiciones entre estados dependen de las tasas de nacimientos y muertes:

- **Argentina** tiene una tasa de natalidad y mortalidad bajas, por lo que el crecimiento poblacional es lento o estable.

- **China** tiene una tasa de natalidad absoluta más alta y una población mucho mayor, lo que hace que el crecimiento sea más rápido y significativo.

Cálculo de probabilidades de los distintos estados

Primero, ¿Qué significa "probabilidad de un estado"?

Cuando analizamos un sistema de colas (por ejemplo, una fila en una caja), una pregunta frecuente es: ¿Cuál es la probabilidad de que haya cierta cantidad de personas en el sistema?

Por ejemplo: ¿cuál es la probabilidad de que haya 0, 2 o 10 personas?

Para responder esto, se utiliza un **modelo de nacimiento y muerte**, un tipo de **proceso de Markov en tiempo continuo** (es un modelo que representa cómo el sistema va pasando de un estado a otro, uno por uno, con el tiempo). En este modelo:

- Los estados posibles son 0, 1, 2, 3, ..., según cuántos clientes haya.
- Solo se puede pasar de un estado al siguiente ($n \rightarrow n+1$) o al anterior ($n \rightarrow n-1$).
- Cada transición ocurre a una **tasa constante**:
    - $\boldsymbol{\lambda_n}$: tasa de llegada al pasar de n a n+1 (nacimiento).
    - $\boldsymbol{\mu_n}$: tasa de salida al pasar de n a n−1 (muerte).

Estado estacionario y cálculo de probabilidades

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 a esto se le llama **estado estacionario**.

En ese estado, se cumple una **ecuación de equilibrio** entre estados consecutivos:

$$\lambda_{n-1} * P(n-1) = \mu_n * P(n)$$

Esto significa que el flujo que entra a un estado es igual al que sale. Con esta relación, se pueden calcular todas las probabilidades en función de **P(0)**, que es la probabilidad de que el sistema esté vacío:

$P(1) = \frac{\lambda_0}{\mu_0} * P(0)$, $P(2) = \frac{\lambda_1}{\mu_2} * P(1)$, etc.

¿Cómo se calcula la probabilidad de cada estado?

Usamos la idea de que, en el largo plazo, el sistema se vuelve **estable**. Es decir: la cantidad de personas que entra, en promedio, es igual a la que sale. Esto permite establecer una relación entre las probabilidades de cada estado.

La clave es una probabilidad base: **la probabilidad de que el sistema esté vacío** (la llamamos **P(0)**). A partir de ella, se pueden calcular todas las demás.

¿Cómo se obtiene esa probabilidad base (P(0))?

Sabemos que el sistema siempre se encuentra en **algún estado** (vacío, con una persona, con dos, etc.). Por lo tanto, **la suma de todas las probabilidades debe ser 100%**, o en términos matemáticos, igual a 1.

P(0)+P(1)+P(2)+P(3)+…= 1

Usando esta regla, y sabiendo que todas las probabilidades dependen de P(0), podemos armar una fórmula que nos permite calcular su valor exacto.

$$P(0) = \frac{1}{1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2\mu_3} + \dots}$$

Se representa cómo se comportan las llegadas y salidas del sistema. Pero lo importante es que, una vez que sabemos cuánto vale P(0), podemos calcular todas las demás probabilidades (P(1), P(2), etc.) fácilmente.

Cada una de esas probabilidades nos dice qué tan probable es que el sistema tenga cierta cantidad de personas en un momento dado. Por ejemplo, si P(2) es 0,25, significa que hay un 25 % de chances de que haya exactamente dos personas en el sistema.

Este enfoque se basa en observar el comportamiento del sistema cuando ya pasó suficiente tiempo y se volvió estable. No importa cómo empezó, lo que nos interesa es cómo se comporta en promedio a lo largo del tiempo.

Aplicación de sistemas de nacimiento-muerte:

Los sistemas **M/M/1** y **M/M/2** son modelos clásicos dentro de la teoría de colas, que se basan en el proceso de nacimiento y muerte dependiente del estado.

La notación **M/M/c** describe un sistema de colas con las siguientes características:

- La primera **M** indica que las **llegadas** de clientes ocurren al azar, pero con una tasa promedio constante, lo que se modela mediante un **proceso de Poisson** (cantidad de eventos en un intervalo de tiempo fijo).

- La segunda **M** significa que los **tiempos de servicio** también son aleatorios y siguen una **distribución exponencial**, lo que implica que el tiempo entre atención varía.

- El número **c** representa la **cantidad de servidores** que atienden al mismo tiempo; por ejemplo, 1 en un sistema M/M/1 o 2 en M/M/2.

Sistemas M/M/2

El sistema M/M/2 es una extensión del modelo M/M/1, donde en lugar de un solo servidor, hay **dos servidores trabajando en paralelo**. Esto significa que pueden atender hasta dos clientes al mismo tiempo, lo cual reduce la espera promedio en comparación con un solo servidor.

Al igual que el M/M/1, este modelo también pertenece a los **sistemas dependientes de estado** o de **nacimiento y muerte**, ya que el número de clientes cambia según llegadas y atenciones finalizadas. Las llegadas siguen una **distribución de Poisson** (con tasa λ) y los tiempos de servicio siguen una **distribución exponencial** (con tasa μ para cada servidor).

En este sistema se suele suponer que **los dos servidores trabajan a la misma velocidad**, es decir, atienden clientes al mismo ritmo. Esto hace que el sistema sea más fácil de analizar y los clientes se asignan al primer servidor libre de forma aleatoria.

Pero en la práctica, **los servidores pueden tener diferentes velocidades**. Por ejemplo, uno puede ser más rápido que el otro.

- 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.

Cuando las velocidades son distintas, el análisis del sistema se complica un poco y muchas veces se usa simulación para entender cómo se comporta.

Para que el sistema funcione de forma estable, es necesario que la **capacidad conjunta de los servidores sea mayor que la tasa de llegada de clientes**. Es decir, que $2\mu > \lambda$. Si no se cumple, la cola crecerá indefinidamente.

Este modelo también permite analizar el rendimiento del sistema a través de ciertas probabilidades y medidas de rendimiento.

Comparado con el M/M/1, el modelo M/M/2 reduce el tiempo de espera y mejora la atención cuando el flujo de clientes es alto, pero también aumenta la complejidad del análisis y los costos, ya que se requiere más personal o recursos para sostener dos servidores.

Sistema MG1

Hasta ahora hemos analizado el **modelo MM2**, donde tanto las llegadas como los tiempos de servicio siguen distribuciones exponenciales, lo que implica que las llegadas de clientes son completamente impredecibles y los tiempos de servicio son aleatorios, pero con una tasa constante.

En la práctica, los tiempos de servicio no siempre siguen una distribución exponencial, ya que pueden depender de factores como la complejidad de la tarea o la disponibilidad de recursos. Por eso se utiliza el modelo M/G/1, que extiende al M/M/1 permitiendo que los tiempos de servicio sigan cualquier distribución general.

En el modelo **MG1**, la "G" indica que la distribución de los tiempos de servicio puede ser cualquier distribución general. Al pasar del MM2 a MG1, estamos considerando un escenario más realista, donde la variabilidad de los tiempos de servicio puede reflejar situaciones como variaciones en la carga de trabajo o en la capacidad de los servidores, lo que hace que el **MG1** sea más adecuado para modelar sistemas con una mayor variabilidad en el servicio.

Diferencias clave entre M/M/2 y M/G/1:

- **Cantidad de servidores**:
  - *M/M/2* cuenta con **dos servidores** que atienden en paralelo, lo cual mejora la capacidad del sistema y reduce los tiempos de espera en la cola.
  - *M/G/1* tiene **un único servidor**, lo que simplifica el sistema, pero puede generar mayores tiempos de espera si la demanda es alta.
- **Distribución de los tiempos de servicio**:
  - En *M/M/2*, los tiempos de atención siguen una **distribución exponencial**, es decir, tienen una media constante y baja variabilidad. Esto facilita el análisis, pero no siempre refleja la realidad.
  - En *M/G/1*, los tiempos de servicio pueden seguir **cualquier distribución general**. Esto permite modelar situaciones más realistas, donde algunos clientes pueden tardar más que otros, generando mayor variabilidad en los tiempos de atención.

Sistema MD1

Una **variante particular del modelo M/G/1** es el sistema **M/D/1**.

Ya que se **especializa** aún más el comportamiento del servicio: **todos los tiempos de atención son constantes**, es decir, **determinísticos**. Este tipo de sistema puede considerarse como una simplificación útil del M/G/1 cuando se sabe que la variabilidad del servicio es muy baja o casi nula.

Por lo tanto, el modelo M/D/1 surge **como un caso especial de M/G/1**, donde el **desvío estándar resulta nulo, es decir, igual a cero**, lo que reduce la complejidad de los cálculos y permite una descripción más precisa de algunos sistemas reales.

La diferencia clave con M/G/1 está en los **tiempos de servicio**:

- En M/G/1, los tiempos de servicio pueden seguir **cualquier distribución estadística**.
- En cambio, en M/D/1, los tiempos de servicio son **determinísticos**, es decir, **iguales para todos los clientes**. Es decir, que al asumir tiempos de servicio constantes, el **desvío estándar es cero** ($\sigma = 0$).

La diferencia clave entre **M/M/1** y **M/D/1** radica en la **distribución de los tiempos de servicio**:

- **M/M/1**: Los **tiempos de servicio** siguen una distribución **exponencial**, lo que significa que el tiempo de atención es aleatorio y tiene una variabilidad constante.
- **M/D/1**: Los **tiempos de servicio**, sin embargo, son **determinísticos** (**D** significa determinístico), es decir, todos los clientes son atendidos en un tiempo fijo y conocido, sin variabilidad.

Algunas ventajas del modelo M/D/1:

- **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.

El ejemplo más común de este modelo son las líneas de producción automatizadas, donde cada pieza tarda exactamente lo mismo en ser procesada.

Sistema de colas con prioridades

Un sistema de colas con prioridades es aquel en el que **no todos los clientes son tratados por igual**: existen **categorías o clases de prioridad**, y los clientes con mayor prioridad **reciben atención antes** que los de menor prioridad, siguiendo ciertas reglas de atención.

Un ejemplo cotidiano es el de una caja de supermercado con cartel de **“Prioridad para embarazadas y personas con discapacidad”**. Allí, estas personas tienen un **trato preferencial**, aunque no siempre se interrumpe al cliente que ya está siendo atendido.

Tipos de Prioridades

1. **Sin Interrupción de Servicio** (Ejemplo: Supermercado):

   En este tipo de sistema, 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.

   Ejemplo: Una persona está siendo atendida en una caja de supermercado, llega una persona embarazada (de mayor prioridad). La persona embarazada debe esperar a que la persona común termine su atención. La atención de la persona común no se interrumpe.

2. **Con Interrupción de Servicio** (Ejemplo: Hospital):

   En este sistema, si llega un **cliente de mayor prioridad**, puede **interrumpir el servicio de un cliente de menor prioridad** que ya está siendo atendido, tomando su lugar en la fila. Este tipo de sistema es común en contextos en los cuales la **urgencia o la gravedad de la situación** (como en hospitales) justifica interrumpir la atención de un cliente para atender a otro cuya situación lo requiere.

   Ejemplo: Un paciente con gravedad leve está siendo atendido en la guardia de un hospital cuando, de repente, ingresa otro paciente con un grado de gravedad superior. Ante esta situación, los médicos interrumpen la atención del paciente para asistir de inmediato al paciente con mayor prioridad.

Ejemplos:

Escenario	**Cliente ya atendido**	**Llega prioritario**	**¿Se interrumpe?**
Supermercado	Persona común	Embarazada	No
Guardia hospital	Tos leve	Accidente grave	Si

Clases de Prioridad

En un sistema de colas con prioridades, los clientes se agrupan en **clases de prioridad** que determinan el orden de atención. Las clases se enumeran según el nivel de urgencia:

- **Clase 1**: Mayor prioridad (se atienden primero).
- **Clase 2**: Menor prioridad (se atienden después de los de clase 1).

El sistema puede extenderse para incluir más clases (por ejemplo, **Clase 3**, **Clase 4**, etc.), y cada **clase tiene su propia cola**. Los clientes de **clases más altas** se atienden antes que aquellos de **clases más bajas**.

En resumen, los sistemas de colas con prioridades permiten gestionar situaciones donde la **urgencia o el tipo de cliente** deben influir en el orden de atención, y este orden puede implicar la **interrupción del servicio** o no, dependiendo del contexto y las reglas del sistema.

¿Qué se analiza en estos sistemas?

- Tiempo medio que cada tipo de cliente pasa en el sistema.
- Tiempo medio de espera en la cola.
- Carga del sistema, longitud de cola promedio, uso del servidor, etc.