

# Estadística y Procesos Estocásticos

## Tema 5: Procesos de Nacimiento y Muerte

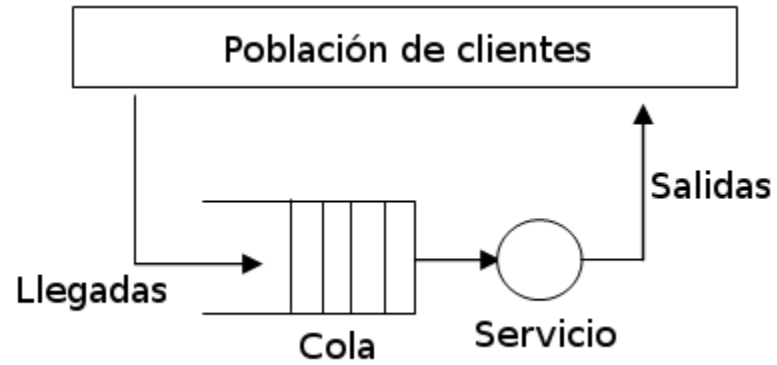
Grado en Ingeniería en Tecnologías de la Telecomunicación

A detailed illustration of a satellite in space. The satellite is positioned in the lower right quadrant, featuring a central body with various instruments and two large, rectangular solar panel arrays extending outwards. Two prominent parabolic dish antennas are visible on the satellite's structure. The background is a dark, deep space with a large, bright sun or star in the center, creating a strong lens flare effect. The sun is surrounded by a glowing orange and red aura, and several faint, circular orbits or paths are visible in the distance.

# 1. Introducción.

- Este tipo de procesos, como sugiere su propio nombre, se comenzó a utilizar inicialmente en demografía para modelar el tamaño de una población.
- Constituyen la herramienta básica para modelar el teletráfico:
  - un nacimiento representa el acceso de un *cliente* a un sistema (por ejemplo, un paquete de datos que accede a un router)
  - una muerte representa la salida de un cliente del sistema.
- Cuando las llegadas y salidas a un sistema se producen siguiendo patrones aleatorios, se forman *líneas de espera* o *colas*. El *protocolo de servicio* especifica en qué orden son atendidos los sucesivos clientes.
- Por tanto estos procesos constituyen la base para el modelado y gestión de tráfico, así como para el dimensionamiento de recursos en redes.

## Ejemplo: Sistema de colas



## 2. Procesos de nacimiento y muerte

- Sea  $N(t)$  el número de clientes en el sistema en el instante  $t$ .
- Sea  $h$  un valor pequeño ( $h \rightarrow 0$ ). Llamaremos:
  - $L_{(t,t+h]}$  = número de llegadas que ocurren en el intervalo  $(t, t + h]$
  - $S_{(t,t+h]}$  = número de salidas en el mismo intervalo
- Sea  $o(h)$  un valor infinitesimal respecto a  $h$ , esto es,  $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$

- Supondremos que en un lapso de tiempo infinitesimal **ocurre como máximo una única llegada**:
  - $\Pr (L_{(t,t+h]} = 1 | N(t) = n) = \lambda_n h + o(h)$
  - $\Pr (L_{(t,t+h]} = 0 | N(t) = n) = 1 - \lambda_n h + o(h)$        $\lambda_n = \mathbf{Tasa\ de\ llegada}$
  - $\Pr (L_{(t,t+h]} > 1 | N(t) = n) = o(h)$
  
- Supondremos también que en un lapso de tiempo corto ocurre **como máximo una única salida**:
  - $\Pr (S_{(t,t+h]} = 1 | N(t) = n) = \mu_n h + o(h)$
  - $\Pr (S_{(t,t+h]} = 0 | N(t) = n) = 1 - \mu_n h + o(h)$        $\mu_n = \mathbf{tasa\ de\ servicio}$
  - $\Pr (S_{(t,t+h]} > 1 | N(t) = n) = o(h)$
  
- Las llegadas y las salidas se producen de forma **independiente**.
- $\lambda$  y  $\mu$  dependen solo de  $n$  y no de  $t$ .

## Ecuaciones del proceso de Nacimiento-Muerte

Llamando:  $P_n(t) = \Pr(N(t) = n)$

Se tiene que:

$$P_n(t+h) = P_n(t)(1 - \lambda_n h)(1 - \mu_n h) + P_{n-1}(t)\lambda_{n-1}h(1 - \mu_{n-1}h) + \\ + P_{n+1}(t)(1 - \lambda_{n+1}h)\mu_{n+1}h + o(h)$$

y para  $n = 0$ :

$$P_0(t+h) = P_0(t)(1 - \lambda_0 h) + P_1(t)(1 - \lambda_1 h)\mu_1 h + o(h)$$

De donde:

$$P'_n(t) = -(\lambda_n + \mu_n)P_n(t) + \lambda_{n-1}P_{n-1}(t) + \mu_{n+1}P_{n+1}(t)$$

$$P'_0(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t)$$

## Ecuaciones del proceso de Nacimiento-Muerte

Si el sistema anterior alcanza el equilibrio (es ergódico), entonces:

$$\lim_{t \rightarrow \infty} P_n(t) = p_n \quad \text{y por tanto} \quad \lim_{t \rightarrow \infty} P'_n(t) = 0$$

y las ecuaciones anteriores se transforman en:

$$0 = -(\lambda_n + \mu_n) p_n + \lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1} ; \quad n \geq 1$$

$$0 = -\lambda_0 p_0 + \mu_1 p_1$$

## Ecuaciones del proceso de Nacimiento-Muerte

Estas ecuaciones se pueden resolver recursivamente: de la segunda ecuación se obtiene:

$$0 = -\lambda_0 p_0 + \mu_1 p_1 \Rightarrow p_1 = \frac{\lambda_0}{\mu_1} p_0$$

Asimismo, de la primera ecuación se sigue que:

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

Por tanto:

$$p_2 = \frac{1}{\mu_2} ((\lambda_1 + \mu_1) p_1 - \lambda_0 p_0) = \frac{1}{\mu_2} \left( (\lambda_1 + \mu_1) \frac{\lambda_0}{\mu_1} p_0 - \lambda_0 p_0 \right) = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} p_0$$

Podemos proceder reiteradamente del mismo modo para obtener  $p_3, p_4, \dots$



## Ecuaciones del proceso de Nacimiento-Muerte

La solución general es de la forma:

$$p_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} \cdot p_0$$

Dado que la sucesión  $\{p_n; n = 0, 1, \dots\}$  es una distribución de probabilidad, verificará que  $\sum_{n=0}^{\infty} p_n = 1$ , de donde se deduce que:

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}}$$

# Sistemas de Colas

- Los procesos de nacimiento-muerte permiten modelar aquellos sistemas de colas en los que las tasas de llegada ( $\lambda_n$ ) y servicio ( $\mu_n$ ) de los clientes dependen únicamente de la ocupación del sistema en cada momento.
- Las ecuaciones anteriores nos proporcionan las probabilidades de que haya  $n$  clientes en el sistema ( $n \geq 0$ ) cuando éste se encuentra en equilibrio.
- **Sistema de colas en equilibrio:** se considera como tal un sistema de colas una vez que ha transcurrido el tiempo suficiente para que la distribución de probabilidad de la ocupación del sistema alcance el estado estacionario. La cola fluctúa constantemente de forma dinámica, llenándose y vaciándose, y el valor de  $p_n$  representa la probabilidad de encontrar  $n$  clientes en el sistema en cualquier futuro instante arbitrario.

# Elementos de un sistema de colas.

- $C_n$ : representa el n-ésimo cliente que llega al sistema.
- $\tau_n$ : instante de llegada de  $C_n$ .
- $T_n$ : tiempo entre las llegadas de  $C_{n-1}$  y  $C_n$ , ( $T_n = \tau_n - \tau_{n-1}$ ).
- $X_n$ : **Duración del servicio** para el cliente  $C_n$ .
- $W_n$ : **Tiempo de espera en cola** del cliente  $C_n$ .
- $S_n$ : **Tiempo en el sistema** ( $W_n + X_n$ ) del cliente  $C_n$ .
- $N(t)$ : **Número de clientes en el sistema** en el instante  $t$ .
- $Q(t)$ : **Tamaño de la cola** en el instante  $t$ .

# Intensidad de tráfico

$$\rho_k = \frac{\lambda_k}{\mu_k}$$

- Mide la relación entre el número de llegadas y el número de salidas por unidad de tiempo.
- **Condición de equilibrio:** La condición necesaria y suficiente para que un proceso de nacimiento-muerte sea ergódico (alcance el equilibrio) es que exista un valor  $k_0$  tal que para todo  $k \geq k_0$ :

$$\rho_k = \frac{\lambda_k}{\mu_k} < 1$$

- Esta condición indica que el sistema sólo alcanza el equilibrio si a partir de cierto número de clientes en el sistema ( $k_0$ ), la velocidad a que llegan los clientes es inferior a la velocidad con que son atendidos.

# Fórmula de Little

Esta fórmula indica que en un sistema de colas en equilibrio:

$$E[N] = \frac{E[S]}{E[T]}$$

donde:

- $E[T]$ : tiempo esperado entre llegadas.
- $E[S]$ : tiempo esperado que permanece en el sistema un cliente arbitrario.
- $E[N]$ : número esperado de clientes en el sistema en un instante arbitrario.

Si  $\lambda$  es el número esperado de llegadas por unidad de tiempo, se tiene que  $\lambda = 1/E[T]$  y la fórmula de Little puede expresarse también como:

$$E[N] = \lambda E[S]$$

# Justificación de la fórmula de Little

En un sistema de colas en equilibrio, **el número esperado de clientes a la llegada de un cliente arbitrario debe coincidir con el número esperado a su salida**. Tenemos:

- Cuando un cliente arbitrario accede al sistema en equilibrio, el número esperado de clientes que encuentra es  $E[N]$ .
- Cuando se marcha del sistema, habrá permanecido en él durante un tiempo esperado  $E[S]$ .
- Como el tiempo medio entre llegadas es  $E[T]$ , durante el tiempo  $E[S]$  habrá llegado un número esperado de clientes  $E[S] / E[T]$ .
- Si la disciplina de la cola es FIFO (First In First Out, los clientes son atendidos en su orden estricto de llegada), cuando el cliente se vaya, los  $E[S] / E[T]$  que han llegado durante su permanencia en el sistema son los que quedan en el sistema.

# Notación de Kendall

Establece una clasificación de los distintos modelos de colas en los que cada cliente debe recibir un solo servicio. Consiste en un código de la forma **L/S/m/K/N**, donde:

- **L**: distribución de probabilidad de los tiempos entre llegadas de clientes a la cola.
- **S**: distribución de los tiempos de servicio.
- **m**: número de servidores.
- **K**: capacidad (buffer) disponible en el sistema para acumular los clientes en cola (si no se especifica, se entiende que la capacidad del sistema es infinita)
- **N**: tamaño de la población de la que proceden los clientes. No se especifica si el tamaño es infinito.

# Notación de Kendall

Los códigos más habituales para las distribuciones de los tiempos entre llegadas y de servicio son: M (exponencial), G (general), y D (determinista). Ejemplos de sistemas descritos por la notación de Kendall son:

- **M/M/1**: Llegadas con distribución de Poisson con tasa constante  $\lambda$  (y por tanto tiempos entre llegadas con distribución exponencial), tiempos de servicio exponenciales y un solo servidor.
- **M/M/1/K**: idéntico al anterior, pero con un buffer de capacidad finita K.
- **M/G/1**: Llegadas con distribución de Poisson con tasa constante  $\lambda$ , tiempos de servicio con distribución general, un único servidor.



# Colas M/M/1

Consiste en un único canal de servicio o servidor en el que las llegadas y salidas se producen de acuerdo con un proceso de nacimiento y muerte con tasas de llegada y salida respectivas:

$$\begin{aligned}\lambda_n &= \lambda & n &= 0, 1, 2, \dots \\ \mu_n &= \mu & n &= 1, 2, 3, \dots\end{aligned}$$

En este sistema:

- Los tiempos entre dos llegadas consecutivas,  $\{T_n; n \in N\}$ , son variables aleatorias independientes y con distribución de probabilidad exponencial de parámetro  $\lambda$ .
- Los sucesivos tiempos de servicio  $\{X_n; n \in N\}$  son también variables aleatorias independientes y con distribución de probabilidad exponencial de parámetro  $\mu$ .
- La distribución de equilibrio es:

$$p_n = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{\mu_1 \mu_2 \dots \mu_n} \cdot p_0 = \left(1 - \frac{\lambda}{\mu}\right) \cdot \left(\frac{\lambda}{\mu}\right)^n ; \quad n = 0, 1, 2, \dots$$

# Colas M/M/1:

Por tanto, en el equilibrio, llamando  $\rho = \frac{\lambda}{\mu}$  la probabilidad de que haya  $N$  clientes en el sistema sigue una **distribución geométrica**:

$$P(N = n) = (1 - \rho) \cdot \rho^n$$

Para calcular la esperanza y la varianza de  $N$  hallamos la función característica:

$$\varphi(t) = E[e^{itN}] = \sum_{n=0}^{\infty} e^{itn} P(N = n) = \sum_{n=0}^{\infty} e^{itn} (1 - \rho) \rho^n = (1 - \rho) \sum_{n=0}^{\infty} (\rho e^{it})^n = \frac{1 - \rho}{1 - \rho e^{it}}$$

Su derivada es:

$$\varphi'(t) = \frac{(1 - \rho) i \rho e^{it}}{(1 - \rho e^{it})^2}$$

y por tanto:

$$E[N] = \frac{1}{i} \varphi'(0) = \frac{\rho}{1 - \rho} = \frac{\lambda/\mu}{1 - \lambda/\mu} = \frac{\lambda}{\mu - \lambda}$$

# Colas M/M/1:

- Para que exista solución de equilibrio debe ocurrir que  $\lambda < \mu$ .
- $E[N] = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$
- $Var(N) = E[N^2] - (E[N])^2 = \frac{1}{i^2} \varphi''(0) - \left(\frac{\rho}{1-\rho}\right)^2 = \dots = \frac{\rho}{(1-\rho)^2}$

- Aplicando la fórmula de Little:

$$E[N] = \lambda E[S] \Rightarrow \frac{\rho}{1-\rho} = \lambda E[S] \Rightarrow E[S] = \frac{\rho}{\lambda(1-\rho)} = \frac{1/\mu}{1-\rho}$$

- El tiempo medio en cola es  $E[W] = E[S] - E[X] = E[S] - \frac{1}{\mu}$ :

$$E[W] = \frac{\rho/\mu}{1-\rho}$$

- De las expresiones anteriores se sigue que cuando  $\rho \rightarrow 0$ :  $E[S] \rightarrow \infty$  y  $E[W] \rightarrow \infty$

# Cola con capacidad finita M/M/1/C:

- En las condiciones del sistema M/M/1, si la capacidad del buffer es finita  $C$  podemos modelar su comportamiento mediante un proceso de nacimiento-muerte con:

$$\lambda_k = \begin{cases} \lambda & k < C \\ 0 & k \geq C \end{cases} \quad \mu_k = \mu, \quad k = 1, 2, 3, \dots, C$$

- Este sistema siempre alcanza el equilibrio, pues para  $k \geq C$  se tiene  $\lambda_k/\mu_k = 0$ .
- Las probabilidades de equilibrio son:

$$p_k = \begin{cases} \frac{1-(\lambda/\mu)}{1-(\lambda/\mu)^{C+1}} \left(\frac{\lambda}{\mu}\right)^k & 0 \leq k \leq C \\ 0 & k > C \end{cases}$$

# Sistemas con capacidad finita M/M/1/C:

- En este sistema la probabilidad de que un cliente sea rechazado (se pierda) coincide con la probabilidad de encontrar el sistema lleno a su llegada:

$$P(\text{Pérdida}) = p_C = \frac{1 - (\lambda/\mu)}{1 - (\lambda/\mu)^{C+1}} \left(\frac{\lambda}{\mu}\right)^C$$

- En la práctica resulta de interés calcular el valor de  $C$  que garantice que la probabilidad de pérdida está por debajo de un umbral preespecificado  $p_u$

# Sistemas con más de un servidor: M/M/m

- Generalizan el modelo M/M/1 al caso de  $m$  servidores. El cliente que ocupa la primera posición de la cola es atendido por el primer servidor que queda libre.
- Podemos modelar este sistema mediante las tasas de llegada y servicio siguientes:

$$\lambda_k = \lambda \quad \forall k$$

$$\mu_k = \begin{cases} k\mu & 0 \leq k \leq m \\ m\mu & k \geq m \end{cases}$$

- La intensidad de tráfico en este sistema es:

$$\rho = \lambda/m\mu$$

# Sistemas con más de un servidor: M/M/m

- Distribución de equilibrio:

$$p_k = \begin{cases} p_0 \frac{(m\rho)^k}{k!} & k \leq m \\ p_0 \frac{\rho^k m^m}{m!} & k \geq m \end{cases} \quad p_0 = \left[ \sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!(1-\rho)} \right]^{-1}$$

- Probabilidad de que un cliente tenga que hacer cola (fórmula C de Erlang):

$$Q_m = \sum_{k=m}^{\infty} p_k = \frac{\frac{(m\rho)^m}{m!} \left( \frac{1}{1-\rho} \right)}{\left[ \sum_{k=0}^{m-1} \frac{(m\rho)^k}{k!} + \frac{(m\rho)^m}{m!(1-\rho)} \right]}$$

# Sistemas con más de un servidor: M/M/m

- Número medio de clientes en el sistema:

$$E[N] = \frac{\rho}{1 - \rho} Q_m + m\rho$$

- Tiempo medio en el sistema:

$$E[S] = \frac{E[N]}{\lambda} = \frac{1}{m\mu - \lambda} Q_m + \frac{1}{\mu}$$

- Tiempo medio en cola:

$$E[W] = \frac{1}{m\mu - \lambda} Q_m$$