Cadenas de Markov de tiempo continuo

- ¿Qué propiedades debe cumplir un proceso estocástico para ser una MC de tiempo continuo?
  - Los estados deben formar un conjunto numerable (En caso contrario podemos tener un *proceso de Markov*).
  - Si $X(t)$ es el evento {*En el instante t el sistema se encuentra en estado i*} se debe cumplir:

$$P\{X(t_n) = i \mid X(t_1) = j, X(t_2) = k, \ldots\} = P(X(t_n) = i \mid X(t_1) = j\}$$

para cualquier $t_n > t_1 > t_2 > t_3 \ldots$

Cadenas de Markov de tiempo continuo

- Probabilidad de transición:

$$p_{ij}(s, t) = P\{X(t) = j \mid X(s) = i\}$$

- En forma matricial (matriz de probabilidad de transición):

$$\mathbf{P}(s,t) = \begin{pmatrix} p_{11}(s,t) & p_{12}(s,t) & \ldots \\ p_{21}(s,t) & p_{22}(s,t) & \ldots \\ \ldots & \ldots & \ldots \end{pmatrix}$$

- Si la cadena es homogénea:

$$p_{ij}(\tau) = P\{X(s+\tau) = j \mid X(s) = i\} = P\{X(\tau) = j \mid X(0) = i\}$$

$$\mathbf{P}(\tau) = \begin{pmatrix} p_{11}(\tau) & p_{12}(\tau) & \ldots \\ p_{21}(\tau) & p_{22}(\tau) & \ldots \\ \ldots & \ldots & \ldots \end{pmatrix}$$

Cadenas de Markov de tiempo continuo

- Para encontrar una ecuación análoga al caso discreto (**$\pi = \pi$ P**, **π e = 1)**, definimos la tasa de transición (*transition rate*) del estado i al j:

$$q_{ij} = \lim_{\Delta t \to 0}\{p_{ij}(\Delta t) / \Delta t\},\ i \neq j$$

- y la tasa de permanencia en el estado i:

$$q_{ii} = \lim_{\Delta t \to 0}\{(p_{ii}(\Delta t) - 1) / \Delta t\} = -\sum_{i \neq j} q_{ij}$$

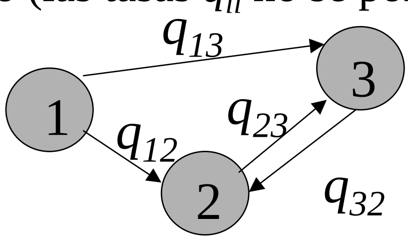
- q_{ii} es negativo porque la probabilidad de permanecer en el mismo estado decrece al aumentar t. q_{ij} es positivo porque la probabilidad de cambiar de estado aumenta con t.

Cadenas de Markov de tiempo continuo

- Se define la matriz de tasa de transición (*transition rate matrix*) o generador infinitesimal de la cadena:

$$\mathbf{Q} = \begin{pmatrix} q_{11} & q_{12} & \ldots \\ q_{21} & q_{22} & \ldots \\ \ldots & \ldots & \ldots \end{pmatrix} = \lim_{\Delta t \to \infty} \frac{\mathbf{P}(\Delta t) - \mathbf{I}}{\Delta t}$$

- Todas la filas de Q cumplen que sus elementos suman 0.
- Una cadena de Markov de tiempo continuo se especifica dando su matriz de tasa de transición, representada gráficamente como (las tasas q_{ii} no se ponen en el diagrama):



Cadenas de Markov de tiempo continuo

- Las ecuaciones de Chapman-Kolmogorov son ahora:

$$p_{ij}(\tau) = \Sigma_k\, p_{ik}(\tau\text{-}\alpha)\, p_{kj}(\alpha)$$

- De donde:

$$p_{ij}(\tau+\Delta t) - p_{ij}(\tau) = \Sigma_k [\, p_{ik}(\tau+\Delta t - \alpha) - p_{ik}(\tau\text{-}\alpha)]\, p_{kj}(\alpha)$$

tomando $\alpha \rightarrow \tau$, $\Delta t \rightarrow 0$: $p_{ik}(\tau\text{-}\alpha) \rightarrow 0$, $i \neq k$; $p_{ii}(\tau\text{-}\alpha) \rightarrow 1$, luego:

$$dp_{ij}(\tau)/d\tau = \Sigma_k\, q_{ik}\, p_{kj}(\tau)$$

y en forma matricial:

$$d\mathbf{P}(\tau)/d\tau = \mathbf{Q}\, \mathbf{P}(\tau)$$

También puede obtenerse: $d\mathbf{P}(\tau)/d\tau = \mathbf{P}(\tau)\, \mathbf{Q}$

- La ecuación matricial diferencial anterior implica:

$$\mathbf{P}(\tau) = e^{\mathbf{Q}\tau} = \mathbf{I} + \mathbf{Q}\, \tau + \mathbf{Q}^2\, \tau^2 / 2! + \mathbf{Q}^3\, \tau^3 / 3! + \ldots$$

Cadenas de Markov de tiempo continuo

- La probabilidad de estar en el estado *i* en el instante *t* es:

$$\pi_i(t) = \mathrm{P}\{X(t) = i\} = \Sigma_k\, \mathrm{P}\{X(t) = i \mid X(0) = k\}\, \mathrm{P}\{X(0) = k\} =$$

$$\Sigma_k\, p_{ki}(t)\, \pi_k(0)$$

- En forma matricial:

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\, \mathbf{P}(t)$$

- Supongamos que existe el límite:

$$\lim_{t\rightarrow 0} \boldsymbol{\pi}(t) = \boldsymbol{\pi} = \boldsymbol{\pi}(0)\, \mathbf{P}(\infty)$$

puesto que:

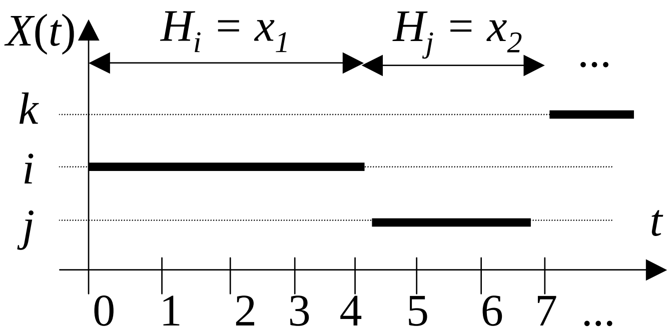
$$\lim_{t\rightarrow 0} d\mathbf{P}(\tau)/d\tau = 0,\ d\mathbf{P}(\tau)/d\tau = \mathbf{P}(\tau)\, \mathbf{Q} \Rightarrow \mathbf{P}(\infty)\, \mathbf{Q} = 0$$

- Obtenemos así la ecuación para la **distribución estacionaria**:

$$\boldsymbol{\pi}\, \mathbf{Q} = 0,\ \boldsymbol{\pi}\, \mathbf{e} = 1$$

Cadenas de Markov de tiempo continuo

- **Tiempo de permanencia** en un estado k (*Sojourn or holding time*): Es la V.A. H_k igual al tiempo de permanencia en el estado k:

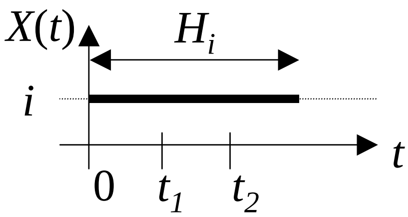


- La propiedad de Markov implica que el tiempo de permanencia en el estado i tiene una distribución exponencial de parámetro q_{ii}:

$$H_i(x) = P\{H_i \le x\} = 1 - \exp\{q_{ii}\, x\}$$

Cadenas de Markov de tiempo continuo

- **Comprobación** de que el tiempo de permanencia con distribución exponencial cumple la propiedad de Markov (ausencia de memoria):



$$P\{X(t_2) = i \mid X(t_1) = i, X(0) = i\} = P\{X(t_2) = i \mid X(t_1) = i\} \Rightarrow$$
$$P\{H_i \le t_2 \mid H_i > t_1\} = P\{H_i \le t_2 - t_1\}$$

- Comprobación:

$$P\{H_i \le t_2 \mid H_i > t_1\} = P\{H_i \le t_2, H_i > t_1\} / P\{H_i > t_1\} =$$
$$P\{t_2 \ge H_i > t_1\} / P\{H_i > t_1\} = [P\{H_i \le t_2\} - P\{H_i \le t_1\}] / P\{H_i > t_1\} =$$
$$[(1 - \exp\{q_{ii}\, t_2\}) - (1 - \exp\{q_{ii}\, t_1\})] / \exp\{q_{ii}\, t_1\} =$$
$$1 - \exp\{q_{ii}\,(t_2 - t_1)\} = P\{H_i \le t_2 - t_1\}.$$

- La distribución exponencial es la única que cumple esta propiedad.

Cadenas de Markov de tiempo continuo

Transición entre estados en una cadena de Markov de tiempo continuo:

Todas las transiciones $i \rightarrow j$, $i \rightarrow k$; $j,k \neq$ i, ocurren distribuidas exponencialmente (con tasas q_{ij}, q_{ik} respectivamente) y independientes entre si.

Cadenas de Markov de tiempo continuo

Generalización de una cadena de Markov de tiempo continuo: proceso semi-Markov:

Definimos la V.A continua H_{ij} igual al tiempo de permanencia en el estado i antes de saltar al estado j.

En un proceso semi-Markov dejamos que la distribución de H_{ij} sea arbitraria. Si H_{ij} está distribuida exponencialmente, tenemos una cadena de Markov de tiempo continuo.

NOTA: Si H_{ij} no está distribuido exponencialmente, considerar sólo el estado actual no cumple la propiedad de Markov (de ausencia de memoria) pues la evolución del proceso depende del estado actual y el tiempo de permanencia en este estado (i, t_i).

Considerar (i, t_i) si cumpliría la propiedad de Markov, pero tendríamos un *proceso de Markov* (pues t_i no es una variable discreta).

Cadenas de Markov de tiempo continuo

Cadena de Markov interna de un proceso semi-Markov (embedded Markov chain):

Si en un proceso semi-Markov no fijamos sólo en los instantes donde se produce un cambio de estados obtenemos la *cadena de Markov interna* del proceso.

Teorema: si π_i^e es la probabilidad estacionaria de la cadena interna, π_i es la probabilidad estacionaria del proceso semi-Markov, y $E[H_i]$ es el tiempo medio de permanencia en el estado i:

$$\pi_i = \frac{\pi_i^e \, E[H_i]}{\sum_j \pi_j^e \, E[H_j]}$$

Conceptes fonamentals de xarxes de computadors. Un enfocament analític.

Cadenas de Markov de tiempo continuo

Justificación de $\pi_i = \frac{\pi_i^e \, E[H_i]}{\sum_j \pi_j^e \, E[H_j]}$

π_i es la fracción de tiempo que el proceso está en el estado i y $\pi_i^e = 1/M_{ii}$ en la cadena interna. Definimos:

$f_i(n)$: fracción de tiempo que el proceso está en el estado i en n saltos de la cadena interna.

$N_i(n)$ número de veces que la cadena interna visita el estado i en estos n saltos.

$H_i(l)$: tiempo de permanencia en el estado i en la visita l que se hace a ese estado durante los n saltos.

Cadenas de Markov de tiempo continuo

Justificación de $\pi_i = \dfrac{\pi_i^{e}\, \mathrm{E}[H_i]}{\sum_j \pi_j^{e}\, \mathrm{E}[H_j]}$

$$f_i(n) = \frac{\sum_{j=1}^{N_i(n)} H_i(j)}{\sum_k \sum_{j=1}^{N_k(n)} H_k(j)} = \frac{\frac{N_i(n)}{n} \sum_{j=1}^{N_i(n)} \frac{H_i(j)}{N_i(n)}}{\sum_k \frac{N_k(n)}{n} \sum_{j=1}^{N_k(n)} \frac{H_k(j)}{N_k(n)}}$$

Puesto que:

$$\begin{cases} f_i(n) \xrightarrow{n\to\infty} \pi_i \\ \sum_{j=1}^{N_i(n)} \frac{H_i(j)}{N_i(n)} \xrightarrow{n\to\infty} \mathrm{E}[H_i] \\ \frac{N_i(n)}{n} \xrightarrow{n\to\infty} 1/M_{ii} = \pi_i^e \end{cases}$$

Substituyendo se justifica la ecuación.

Cadenas de Markov de tiempo continuo

- **Cadena de Markov interna** de una cadena de Markov de tiempo continuo.
- Las probabilidades de transición p_{ij}^{e} de la cadena interna valen la probabilidad de que haya una transición del estado i al j en la cadena de tiempo continuo. Estas probabilidades valen:

$$p_{ij}^{e} = q_{ij} / \Sigma_{i \neq j}\ q_{ij}\,,\ i \neq j$$
$$p_{ii}^{e} = 0$$

La distribución estacionaria de la cadena a tiempo continuo π_ι puede calcularse a partir de la distribución estacionaria de la cadena interna con la fórmula:

$$\pi_i = \frac{\pi_i^{e} / q_i}{\sum_j \pi_j^{e} / q_j}\,,\ q_i = \sum_{j \neq i} q_{ij}$$

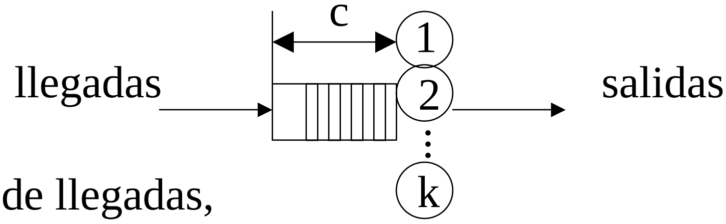
Cadenas de Markov de tiempo continuo

- Justificación: la probabilidad de que haya una transición del estado *i* al *j* en la cadena de tiempo continuo es la probabilidad que que la distribución exponencial de tasa q_{ij} sea la de valor menor de las distribuciones exponenciales $\{q_{ik}\}_{i \neq j}$. Esta probabilidad vale $q_{ij}/\Sigma_{i \neq j}\, q_{ij}$ C.Q.D.
- La distribución del tiempo de permanencia en el estado *i* de la cadena a tiempo continuo es la distribución del mínimo de las distribuciones exponencialmente distribuidas de tasas $\{q_{ij}\}_{i \neq j}$. Esta distribución es una distribución exponencial de tasa $\Sigma_{i\neq j} q_{ij}$. Luego $\mathrm{E}[H_i] = 1/\Sigma_{i \neq j}\, q_{ij}$. Substituyendo:

$$\pi_i = \frac{\pi_i^{\mathrm{e}}\ \mathrm{E}[H_i]}{\sum_j \pi_j^{\mathrm{e}}\ \mathrm{E}[H_j]} = \frac{\pi_i^{\mathrm{e}} / q_i}{\sum_j \pi_j^{\mathrm{e}} / q_j},\ q_i = \sum_{j \neq i} q_{ij}$$

Teoría de colas (*queueing theory*)

- **Notación** de Kendal: A/S/k/c/p



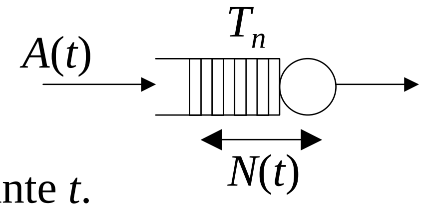
- A: proceso de llegadas,
- S: proceso de servicios,
- k: número de servidores,
- c: tamaño de la cola,
- p: población (si “c” o “p” no se indican, significa infinitos).

Procesos habituales:

G: general (no se supone un proceso concreto), M: Markoviano (distribución exponencial), D: determinista, Er: Erlang...

Teoría de colas (*queueing theory*)

- **Teorema de Little**:



Definimos los procesos estocásticos:

- $A(t)$: número de llegadas hasta el instante t.
- T_n: tiempo en el sistema del usuario n.
- $N(t)$: número de usuarios en el sistema en el instante t.

Y las cantidades:

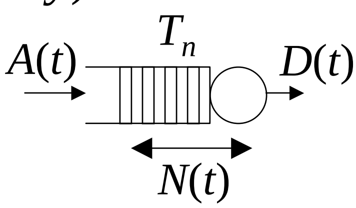
- Número medio de usuarios en el sistema: $N = \lim_{t\to\infty} \frac{1}{t}\int_0^t N(s)ds$
- Tasa media de llegada: $\lambda = \lim_{t\to\infty} A(t) / t$
- Tiempo medio en el sistema: $T = \lim_{t\to\infty} \Sigma_n T_n / A(t)$

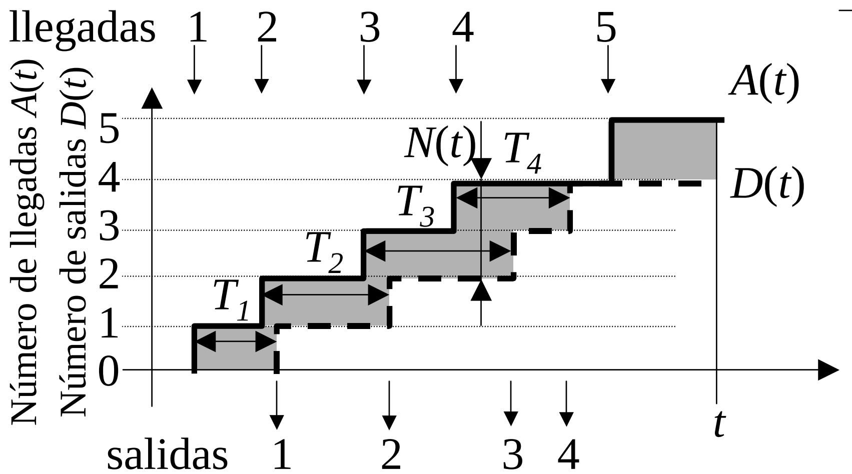
Se cumple que:

$$N = \lambda\, T$$

Teoría de colas (*queueing theory*)

- **Demostración** gráfica del teorema de Little:





- Del gráfico:

$$\frac{1}{t}\int_0^t N(s)ds = \frac{1}{t}\sum_{i=1}^{A(t)} T_i = \frac{A(t)}{t}\,\frac{\sum_{i=1}^{A(t)} T_i}{A(t)} \Rightarrow$$

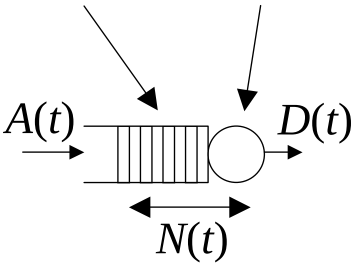
- Tomando el límite:

$$N = \lambda\, T$$

Teoría de colas (*queueing theory*)

- Podemos aplicar el teorema de Little a la cola de espera y al servidor:

Espera en la cola: W_n Tiempo de servicio: S_n



Tiempo en el sistema:

$$T_n = W_n + S_n$$

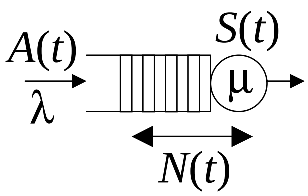
- Número medio de usuarios en la cola: $N_Q = \lambda\, W$
- Número medio de usuarios en el servidor: $\rho = \lambda\, S$
- ρ es la utilización o carga del servidor.

NOTA: $N = N_Q + \rho = \lambda\,(W + S) = \lambda\, T$

Conceptes fonamentals de xarxes de computadors. Un enfocament analític.

Teoría de colas (*queueing theory*)

- La **cola M/M/1**:



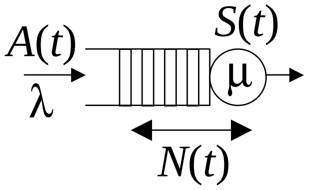
- Llegadas Markovianas de tasa $\lambda \Rightarrow$ tiempo entre llegadas exponencialmente distribuido de media $1/\lambda$: $P\{A_n \leq x\} = 1 - e^{-\lambda x}$ $\Rightarrow A(t)$ es un proceso de Poisson de parámetro $\lambda\, x$: $P\{A(t) \leq n\} = (\lambda\, x)^n\, e^{-\lambda\, x} / n!$.
- Servicios Markovianos de tasa $\mu \Rightarrow$ tiempo de servio exponencialmente distribuido de media $1/\mu$: $P\{S_n \leq x\} = 1 - e^{-\mu x}$.

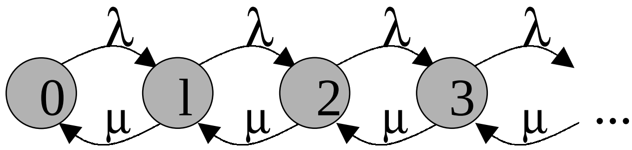
Teoría de colas (*queueing theory*)

- La **cola M/M/1**:
  - El proceso $N(t)$ = número de usuarios en el sistema en el instante t es una cadena de Markov.

OBSERVACIÓN: para un servicio no Markoviano, el proceso $N(t)$ no seria una cadena de Markov!

- Diagrama de tasas de transición:





- Matriz de tasa de transición:

$$\mathbf{Q}=\begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ 0 & \mu & -(\lambda+\mu) & \lambda & \cdots \\ 0 & 0 & \mu & -(\lambda+\mu) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Teoría de colas (*queueing theory*)

- La cola M/M/1, cálculo de la **distribución estacionaria**:

$$\mathbf{Q}=\begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ 0 & \mu & -(\lambda+\mu) & \lambda & \cdots \\ 0 & 0 & \mu & -(\lambda+\mu) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

"Global balance equations":

$$\pi\,\mathbf{Q} = 0,\ \pi\,\mathbf{e} = 1$$

$$\left\{\begin{array}{l} -\pi_0\,\lambda + \pi_1\,\mu = 0 \\ \pi_0\,\lambda - \pi_1\,(\lambda+\mu) + \pi_2\,\mu = 0 \\ \ldots \\ \pi_{i-1}\,\lambda - \pi_i\,(\lambda+\mu) + \pi_{i+1}\,\mu = 0 \end{array}\right. \quad \rho = \frac{\lambda}{\mu} \quad \Rightarrow \left\{\begin{array}{l} \pi_1 = \rho\,\pi_0 \\ \pi_2 = (1+\rho)\,\pi_1 - \rho\,\pi_0 = \rho^2\,\pi_0 \\ \ldots \\ \pi_{i+1} = (1+\rho)\,\pi_{i-1}\,\lambda - \rho\,\pi_i = \rho^{i+1}\,\pi_0 \end{array}\right.$$

- Normalización:

$$\Sigma_k\,\pi_k = 1 \Rightarrow \pi_0\,\Sigma_k\,\rho^i = 1 \Rightarrow \pi_0 = 1 - \rho \Rightarrow \boxed{\pi_i = (1-\rho)\,\rho^i}$$

Teoría de colas (*queueing theory*)

- La cola M/M/1, **propiedades**:
- Número medio de usuarios en el sistema:

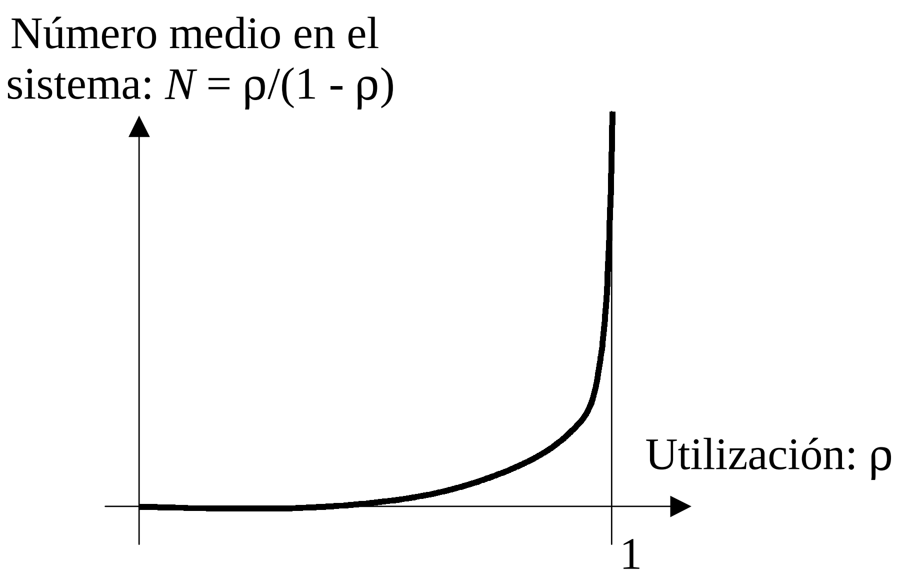
$$N = \lim_{t\to\infty} \frac{1}{t}\int_0^t N(s)ds = \sum_{i=0}^{\infty} i\,\pi_i = \sum_{i=0}^{\infty} i\,(1-\rho)\rho^i = \frac{\rho}{(1-\rho)}$$

- Tiempo medio en el sistema:
  Aplicando Little: $N = \lambda\, T \Rightarrow T = N / \lambda = \rho/[\lambda\,(1-\rho)] = 1/(\mu - \lambda)$
- Tiempo medio en la cola: $W = T - 1/\mu = \rho/(\mu - \lambda)$
- Número medio en la cola: $N_Q = \lambda\, W = \rho^2/(1 - \rho)$
- Número medio en el servidor: $N_s = N - N_Q = \rho$
  NOTA: $\rho = 1 - \pi_0$

Conceptes fonamentals de xarxes de computadors. Un enfocament analític.

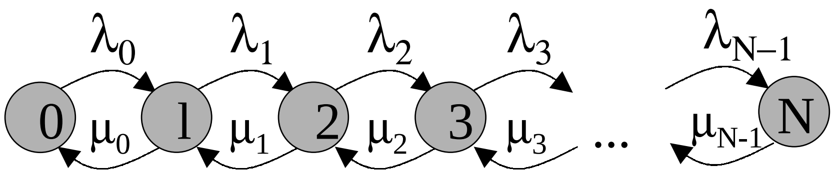
Teoría de colas (*queueing theory*)

- La cola M/M/1, **inestabilidad cuando $\rho\to 1$**:
- N, T son $\propto 1/(1 - \rho) \Rightarrow$ cuando $\rho\to 1$; N, $T\to\infty$



Teoría de colas (*queueing theory*)

- Generalización de la cola M/M/1, **procesos de nacimiento y muerte** ("birth death processes"): son procesos donde sólo puede haber transiciones entre estados adyacentes. El número de estados (N) puede ser finito o infinito.



Teoría de colas (*queueing theory*)

- "**Método del balance de flujos**" para la solución de cadenas de Markov.
- Definimos el "flujo" entre estados de una cadena de Markov:
  - Flujo del estado u al v: $F(u,v) = \pi_u \, q_{uv}$
  - Flujo entre dos conjuntos de estados U, V:

$$F(U,V) = \sum_{u \in U, v \in V} F(u,v)$$

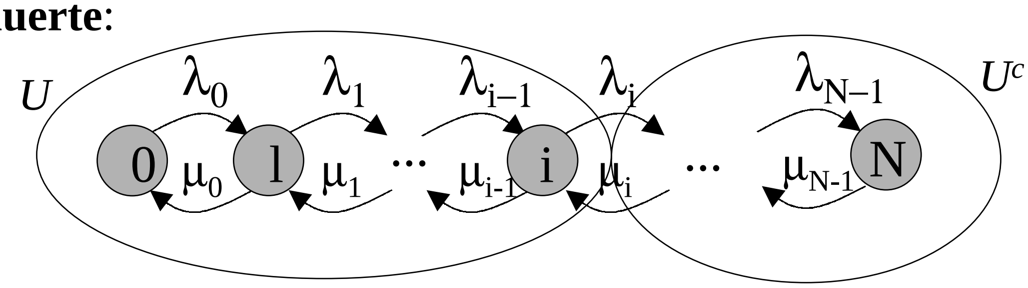
- De las "Global balance equations" ($\pi \, \mathbf{Q} = \mathbf{0}$) se puede deducir:

$$F(U,U^c) = F(U^c,U)$$

- Para cadenas de tiempo discreto se obtiene el mismo resultado cambiando las tasas (q_{uv}) por las probabilidades de transición en un salto (p_{uv}).

Teoría de colas (*queueing theory*)

- Distribución estacionaria de un **procesos de nacimiento y muerte**:



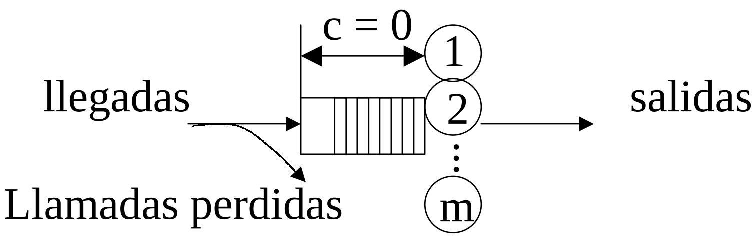
- El balance de flujos entre *U* y U^c implica: $\lambda_i \pi_i = \mu_i \pi_{i+1}$
- Iterando y normalizando se obtiene:

$$\pi_i = \frac{\Psi_i}{\sum_{k=0}^{N} \Psi_k} \qquad \text{donde:} \quad \Psi_i = \begin{cases} 1, & i = 0. \\ \prod_{k=0}^{i-1} \rho_k, & i \neq 0 \end{cases}$$

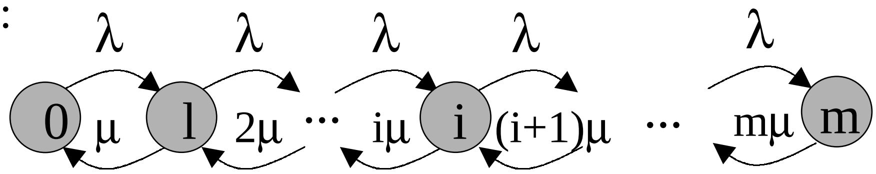
$$\rho_i = \frac{\lambda_i}{\mu_i}, \; i = 0, ..., N-1$$

Teoría de colas (*queueing theory*)

- Ejemplo probabilidad de pérdida en una **centralita telefónica**:
- Hipótesis: centralita con *m* circuitos disponibles con "llamada perdida", población infinita, llegadas Markovianas de tasa λ y duración de las llamadas exponencial de media $1/\mu \Rightarrow$ **cola M/M/m/0**.



- Puesto que la superposición de *n* VA exponencialmente distribuidas de tasas λ_1,.., λ_n es una VA exponencial de tasa $\lambda_1 + \ldots + \lambda_n$:



Teoría de colas (*queueing theory*)

- Distribución estacionaria de la **cola M/M/m/0**:
- Aplicando la solución obtenida para un proceso de nacimiento y muerte:

$$\rho_i = \frac{\lambda}{(i+1)\mu}, \ i = 0, ..., m-1 \qquad \Psi_i = \begin{cases} 1, \ i = 0. \\ \prod_{k=0}^{i-1} \rho_k = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}, \ i \neq 0 \end{cases}$$

$$\pi_i = \frac{\Psi_i}{\sum_{k=0}^{m} \Psi_k} = \frac{\left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!}}{\sum_{k=0}^{m} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}}$$

Teoría de colas (*queueing theory*)

- **Teorema PASTA**: “Poisson Arrivals See Time Average”.
- El tiempo medio que la cadena está en el estado *i* es $\pi_i \Rightarrow$ por PASTA, la probabilidad de que una llegada Marcoviana encuentre el sistema en el estado *i* es $\pi_{i.}$
- Ejemplo: en el modelo de la centralita telefónica con *m* circuitos, la probabilidad de tener una llamada perdida es la probabilidad de que la cola M/M/m/0 este en el estado *m*:

$$\pi_m = \frac{\left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!}}{\sum_{k=0}^{m} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}}$$

“Fórmula de Erlang B”

Teoría de colas (*queueing theory*)

La cola **M/G/1**:

- El proceso $N(t)$ = número de usuarios en el sistema en el instante t en general NO es una cadena de Markov (sólo si G es Markoviano).
- Supongamos que los servicios se completan en los instantes t_i (t_i son también los instantes de salida). Definimos el proceso estocástico de tiempo discreto: $X(t_i)$ = número de usuarios en el sistema en el instante t_i (inmediatamente después del servicio).
- El proceso $X(t_i)$ es una cadena de Markov de tiempo discreto. Justificación: el valor de $X(t_i)$ depende sólo del número de llegadas en intervalos que no se solapan, que es un proceso sin memoria. $X(t_i)$ es la cadena interna del proceso semi-Markov (i, t_i), donde t_i es el tiempo entre los servicios i y i-1.

Teoría de colas (*queueing theory*)

- La cola M/G/1: deducción de la matriz de **probabilidad de transición** en un salto:
- Sea $f_S(x)$ la función de densidad de la V.A. S igual al tiempo de servicio.
- Definimos la V.A. V = {número de llegadas en un tiempo de servicio}, y las probabilidades: $v_i = \mathrm{P}\{V = i\}$.
- Condicionando a la duración del servicio:

$$v_i = \int_{x=0}^{\infty} \mathrm{P}\left\{\begin{matrix} i \text{ llegadas en} \\ \text{un tiempo } x \end{matrix} \,\middle|\, S = x\right\} f_S(x)dx = \int_{x=0}^{\infty} \frac{(\lambda x)^i}{i!} \mathrm{e}^{-\lambda x} f_S(x)dx$$

- Las probabilidades de transición en un paso p_{ij} valen:

$$p_{ij} \begin{cases} 0, & j < i - 1, \quad \text{(la cola sólo puede decrementarse en 1)} \\ v_j, & i = 0,\ j \geq 0, \quad \text{(si la cola estaba vacía, quedan los que llegan)} \\ v_{j-i+1}, & i > 0,\ j \geq i - 1. \quad \text{(si la cola no estaba vacía, se decrementa en 1)} \end{cases}$$

Teoría de colas (*queueing theory*)

- La cola M/G/1: hemos obtenido que la matriz de probabilidad de transición:

$$p_{ij} \begin{cases} 0, & j < i - 1, \\ v_j, & i = 0,\ j \geq 0, \\ v_{j-i+1}, & i > 0,\ j \geq i - 1. \end{cases} \Rightarrow \quad \mathbf{P} = \begin{pmatrix} v_0 & v_1 & v_2 & v_3 & \cdots \\ v_0 & v_1 & v_2 & v_3 & \cdots \\ 0 & v_0 & v_1 & v_2 & \cdots \\ 0 & 0 & v_0 & v_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- **Propiedades** de la solución estacionaria ($\boldsymbol{\pi = \pi}\ \mathbf{P}$, $\boldsymbol{\pi}\ \mathbf{e = 1}$):
- Aplicando el teorema “Level Crossing Law”: en una cola con llegadas y salidas unitarias se cumple que la P{un usuario al llegar encuentra *i* en el sistema} = P{un usuario al salir deja *i* en el sistema} $\Rightarrow$

$$\boxed{\pi_i = \text{P}\{\text{un usuario al llegar encuentra } i \text{ en el sistema}\}.}$$

- Aplicando PASTA:

$$\boxed{\pi_i = \text{P}\{\text{en el sistema hay } i \text{ usuarios en un instante arbitrario}\}.}$$

Teoría de colas (*queueing theory*)

- Justificación del teorema “**Level Crossing Law**”:
- Sea:

$A_i(t)$ = {número de llegadas que encuentran el sistema con *i* usuarios}

$D_i(t)$ = {número de salidas que dejan el sistema con *i* usuarios}

$$\Downarrow$$

P{un usuario al llegar encuentra *i* en el sistema} = $\lim_{t\to\infty} A_i(t) / A(t)$

P{un usuario al salir deja *i* en el sistema} = $\lim_{t\to\infty} D_i(t) / D(t)$

- Un usuario que llega y ve *i* en el sistema provoca una transición $i \to i+1$, un usuario que deja *i* provoca una transición $i+1 \to i$.
- Puesto que las llegadas y salidas son unitarias, el número de transiciones $i \to i+1$ y $i+1 \to i$ puede diferir como máximo en 1:

$$|A_i(t) - D_i(t)| \leq 1$$

$$\Downarrow$$

$$\lim_{t\to\infty}\left\{\frac{A_i(t)}{A(t)} - \frac{D_i(t)}{D(t)}\right\} = \lim_{t\to\infty}\left\{\frac{A_i(t) - D_i(t)}{A(t)} - \frac{D_i(t)}{D(t)}\frac{D(t) - A(t)}{A(t)}\right\} = 0$$

Teoría de colas (*queueing theory*)

- Cálculo del tiempo medio en la cola M/G/1:W (**fórmula de Pollaczek-Khinchin**, P-K).
- Método de los momentos:

$$W = \mathrm{E}[S]\, N_Q + \rho\, R$$

donde:

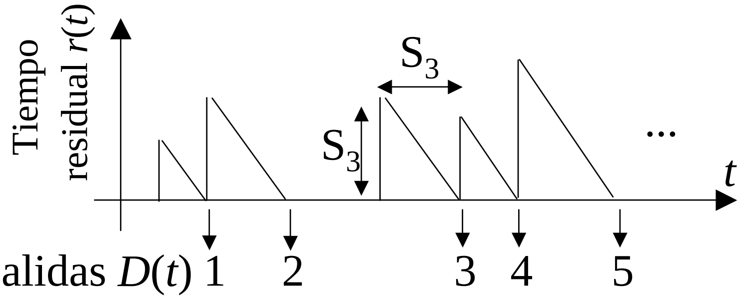
- – $\mathrm{E}[S]$ es el tiempo medio de servicio,
- – N_Q es el número medio en la cola,
- – R es el tiempo medio residual = E[tiempo que queda al usuario que se está sirviendo al haber una llegada].
- – $\rho = \lambda\, \mathrm{E}[S]$ es la probabilidad de que el servidor esté ocupado.

- Aplicando Little a la cola: $N_Q = \lambda\, W \Rightarrow W = \mathrm{E}[S]\, \lambda\, W + \rho\, R \Rightarrow$

$$W = \rho\, R\, /\, (1 - \rho).$$

Teoría de colas (*queueing theory*)

- La cola M/G/1, fórmula de Pollaczek-Khinchin (P-K): cálculo del **tiempo medio residual**:



- De la figura:

$$R = \lim_{t\to\infty} \frac{\sum_{i=0}^{D(t)} \int_0^{S_i} r(\tau)\,d\tau}{\sum_{i=0}^{D(t)} S_i} = \lim_{t\to\infty} \frac{\frac{1}{D(t)} \sum_{i=0}^{D(t)} \frac{1}{2} S_i^{\,2}}{\frac{1}{D(t)} \sum_{i=0}^{D(t)} S_i} = \frac{\mathrm{E}[S^2]}{2\mathrm{E}[S]}$$

- Substituyendo:

$$W = \rho\, R\, /\, (1 - \rho) \Rightarrow \boxed{W = \frac{\lambda\, \mathrm{E}[\, S^{\,2}\,]}{2(1 - \rho)},}\quad \rho = \lambda\, \mathrm{E}[S].$$

Teoría de colas (*queueing theory*)

- La cola M/G/1, fórmula de Pollaczek-Khinchin, P-K: **consecuencias**:
- Tiempo medio en el sistema: $T = \mathrm{E}[S] + W = \mathrm{E}[S] + \dfrac{\lambda\,\mathrm{E}[S^2]}{2(1-\rho)}$
- Para la cola M/M/1: $\mathrm{E}[S^2] = 2/\mu^2 \Rightarrow W = \dfrac{\rho}{\mu(1-\rho)}$
- Para la cola M/D/1: $\mathrm{E}[S^2] = 1/\mu^2 \Rightarrow W = \dfrac{\rho}{2\mu\,(1-\rho)}$
- Observaciones:
- La cola M/M/1 tiene W, N_Q igual a la mitad de la M/D/1 y T, N iguales para valores $\rho \to 0$ y la mitad para $\rho \to 1$.
- La cola M/D/1 tiene el valor mínimo posible de $\mathrm{E}[S^2] \Rightarrow$ es una cota mínima de W, T, N_Q y N para una M/G/1.

Conceptes fonamentals de xarxes de computadors. Un enfocament analític.

Teoría de colas (*queueing theory*)

- **Distribución de la cola en la cola M/G/1:**

Definimos

V_i = {número de llegadas en un tiempo de servicio del usuario i}.

N_i = {número de usuarios en el sistema al finalizar el servicio del usuario i} $\Rightarrow N_i = (N_{i-1} - 1)^+ + V_i$.

Si hacemos $i \to \infty$, N_i tiende a la V.A. igual al número de usuarios en el sistema al final de un servicio. Por el teorema *Level Crossing Law* esto es igual al número de usuarios al haber una llegada y por PASTA, también es igual al número de usuarios en el sistema.

La transformada z de N vale:

$$\mathrm{N}(z) = \Sigma_i \mathrm{P}[N=i]\, z^i = \mathrm{E}[z^N] = \mathrm{E}[z\wedge\{(N-1)^+ + V\}] =$$

$$\mathrm{E}[z\wedge(N-1)^+]\ \mathrm{E}[z^V]$$

Teoría de colas (*queueing theory*)

- Cálculo de $N(z) = E[z\wedge(N-1)^+]\, E[z^V]$

$$E[z\wedge(N-1)^+] = \Sigma_i P[(N-1)^+ = i]\, z^i$$

$$P[(N-1)^+ = i] = \begin{cases} P[N=0] + P[N=1], & i = 0 \\ P[N = i+1]+, & i > 0 \end{cases}$$

$$E[z\wedge(N-1)^+] = P[N=0] + 1/z\, \Sigma_{i>0} P[N = i]\, z^i =$$
$$P[N=0] + 1/z\, \{N(z) - P[N=0]\}$$

$$E[z^V] = P[V=i]\, z^i = \Sigma_i v_i\, z^i = \Sigma_i \int_{x=0}^{\infty} \frac{(\lambda x)^i}{i!} e^{-\lambda t} f_S(x)dx =$$

$$\int_{x=0}^{\infty} \Sigma_i \frac{(\lambda x z)^i}{i!} e^{-\lambda x} f_S(x)dx = \int_{x=0}^{\infty} e^{-\lambda(1-z)x} f_S(x)dx = L_S(\lambda(1-z))$$

Teoría de colas (*queueing theory*)

- Cálculo de $N(z) = E[z\wedge(N-1)^+]\, E[z^V]$

Substituyendo y teniendo en cuenta que $P[N = 0] = 1-\rho$

(Ecuación transformada de Pollaczek-Kinchin)

$$N(z) = L_S(\lambda(1-z)) \frac{(1-\rho)(1-z)}{L_S(\lambda(1-z)) - z}$$

Usando la igualdad $(1-x)^{-1} = \Sigma_{n\geq 0}\, x^n$, podemos escribir la ecuación anterior como:

$$N(z) = (1-\rho)(1-z)\sum_{n=0}^{\infty} \left(\frac{z}{L_S(\lambda(1-z))} \right)^n$$

Teoría de colas (*queueing theory*)

- **Distribución del tiempo en el sistema (*response time*) en la cola M/G/1.**

Hemos deducido que si X es una V.A. continua con transformada de Laplace $L_X(s)$, el número de llegadas de un proceso de poisson con tasa λ durante sucesivos valores de X vale:

$$E[z^X] = L_X(\lambda(1\text{-}z))$$

Para calcular el tiempo de respuesta T maquemos un usuario J. El número de usuarios en el sistema a la salida de J es el número de usuarios llegados durante T_J. Por *level crossig law*, es igual al número de usuarios a la llegada de J, que por PASTA es igual al número de usuarios en la cola en un instante arbitrario: N. Así pues:

$$E[z^X] = E[z^N] = N(z) = L_T(\lambda(1\text{-}z))$$

Teoría de colas (*queueing theory*)

- **Distribución del tiempo en el sistema (*response time*) en la cola M/G/1.**

Substituyendo:

$$E[z^X] = E[z^N] = N(z) = L_T(\lambda(1\text{-}z))$$

$$N(z) = L_S(\lambda(1-z))\frac{(1-\rho)(1-z)}{L_S(\lambda(1-z))-z} = L_T(\lambda(1-z))$$

Con el cambio: $s = \lambda(1\text{-}z)$; $1\text{-}z = s/\lambda$; $z = 1\text{-}\ s/\lambda$

$$L_T(s) = L_S(s)\frac{(1-\rho)s/\lambda}{L_S(s)+s/\lambda\text{-}1}$$

Teoría de colas (*queueing theory*)

Ejemplo distribución de la cola en la colaM/M/1: $P[S \le x]=1-e^{-\mu x}$

$L_s(s)= \mu/(\mu+s)$; $L_s(\lambda(1-z))=\mu/(\mu+\lambda(1-z))=1/(1+\rho(1-z))$ Puesto

$$N(z)=(1-\rho)(1-z)\sum_{n=0}^{\infty}\left(\frac{z}{L_S(\lambda(1-z))}\right)^n=(1-\rho)(1-z)\sum_{n=0}^{\infty}z^n[1+\rho(1-z)]^n=$$

$$(1-\rho)(1-z)\sum_{n=0}^{\infty}z^n\sum_{j=0}^{n}\binom{n}{j}\rho^j(1-z)^j=(1-\rho)\sum_{j=0}^{\infty}\rho^j(1-z)^{j+1}\sum_{n=j}^{\infty}\binom{n}{j}z^n=$$

$$\sum_{j=0}^{\infty}(1-\rho)\rho^j z^j$$ Puesto que $N(z) = \Sigma_i P[N=j]\, z^j \Rightarrow P[N=j] = (1-\rho)\, \rho^j$

Nota: se han usado las relaciones:

$$\begin{cases}\binom{n}{i}=\binom{n}{n-i},\ \sum_{k=0}^{\infty}\binom{n+k}{n}x^n=\frac{1}{(1-x)^{j+1}}\\ \sum_{n=j}^{\infty}\binom{n}{j}z^n=\sum_{n=j}^{\infty}\binom{n}{n-j}z^n=\sum_{i=0}^{\infty}\binom{i+j}{i}z^{i+j}=\frac{z^j}{(1-z)^{j+1}}\end{cases}$$

Teoría de colas (*queueing theory*)

Ejemplo: distribución del tiempo de respuesta en la cola M/M/1:

$P[S \le x]=1-e^{-\mu x}$; $L_s(s)= \mu/(\mu+s)$

$$L_T(s)=L_S(s)\frac{(1-\rho)s/\lambda}{L_S(s)+s/\lambda-1}=\frac{\mu(1-\rho)}{s+\mu-\lambda}$$

Usando $L_s^{-1}[1/(s+\alpha)]= e^{-\alpha x}$:

$$f_T(x)=L^{-1}\left[\frac{\mu(1-\rho)}{s+\mu-\lambda}\right]=\mu(1-\rho)e^{-\mu(1-\rho)x},\ P[T\le x]=1-e^{-\mu(1-\rho)x}$$

Cuyo valor medio vale: $T = 1/[\mu(1-\rho)] = 1/(\mu-\lambda)$