

Sistemas con cola

Area de Ingeniería Telemática
<http://www.tlm.unavarra.es>

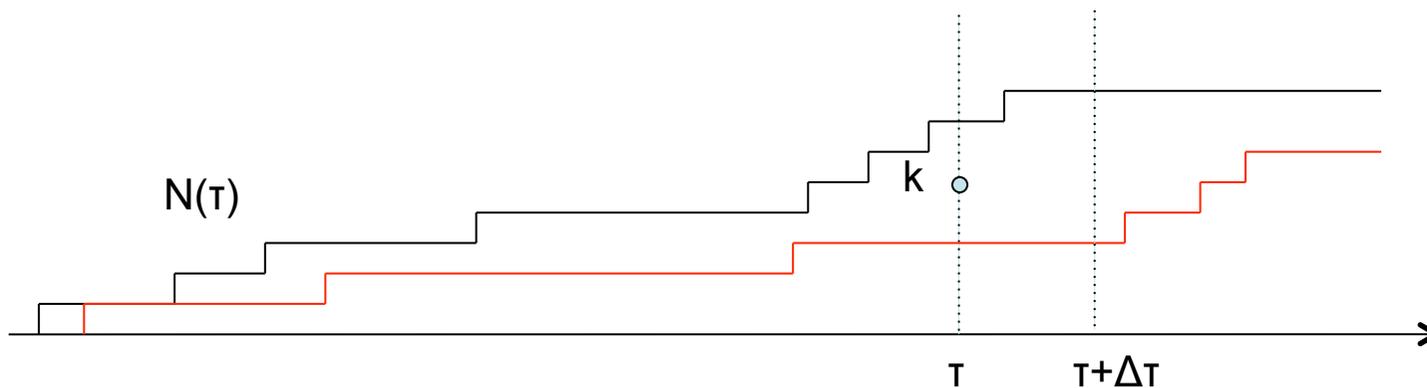
Grado en Ingeniería en Tecnologías de
Telecomunicación, 4º

Ecuaciones del sistema con cola en equilibrio

Hemos visto

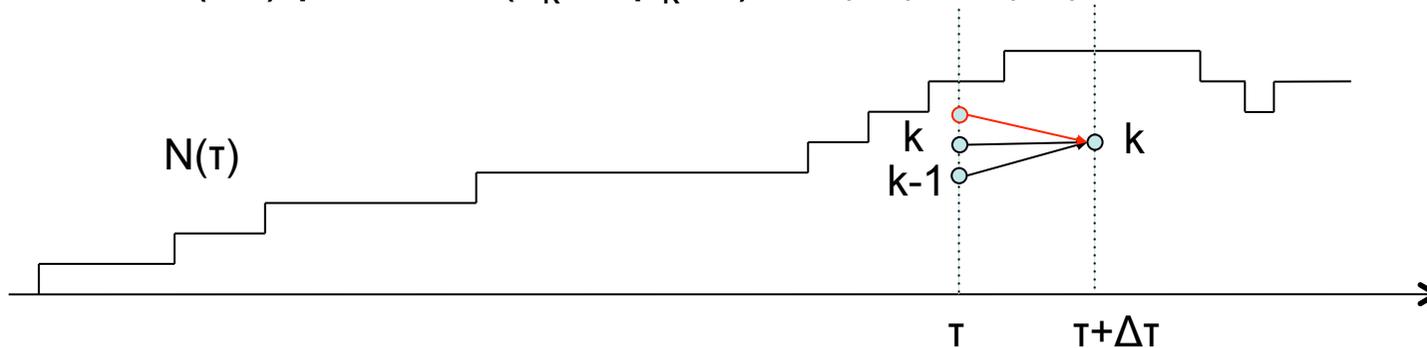
Proceso de nacimiento puro

- El estado es el número de clientes en el sistema
- Llegadas (nacimientos) independientes
- Supongamos que con Δt pequeño:
- $P[1 \text{ llegada en } (\tau, \tau + \Delta\tau) \mid k \text{ en el sistema}] = \lambda_k \Delta\tau + o(\Delta\tau)$
- $P[0 \text{ llegadas en } (\tau, \tau + \Delta\tau) \mid k \text{ en el sistema}] = 1 - \lambda_k \Delta\tau + o(\Delta\tau)$



Comportamiento de cola

- Ahora no solo hay “nacimientos” sino también “muertes”
- El estado es el número de clientes en el sistema (Markoviano)
 - $P[1 \text{ llegada en } (\tau, \tau+\Delta\tau) \mid k \text{ en el sistema }] = \lambda_k \Delta\tau + o(\Delta\tau)$
 - $P[0 \text{ llegadas en } (\tau, \tau+\Delta\tau) \mid k \text{ en el sistema }] = 1 - \lambda_k \Delta\tau + o(\Delta\tau)$
 - $P[1 \text{ salida en } (\tau, \tau+\Delta\tau) \mid k \text{ en el sistema }] = \mu_k \Delta\tau + o(\Delta\tau)$
 - $P[0 \text{ salidas en } (\tau, \tau+\Delta\tau) \mid k \text{ en el sistema }] = 1 - \mu_k \Delta\tau + o(\Delta\tau)$
- Si miramos solo las llegadas forman un proceso de Poisson
- Y tal y como hemos planteado las salida implica que **el tiempo de servicio es exponencial** con parámetro μ_k
- Ahora debemos contar la transición de una salida
- Una entrada simultánea a una salida tiene probabilidad de orden $o(\Delta\tau)$ pues es $(\lambda_k \Delta\tau \mu_k \Delta\tau) + o(\Delta\tau) = o(\Delta\tau)$



Comportamiento de cola

- Siguiendo un razonamiento similar al que hicimos antes llegamos a:

$$dP_k(\tau)/d\tau = - (\lambda_k + \mu_k)P_k(\tau) + \lambda_{k-1}P_{k-1}(\tau) + \mu_{k+1}P_{k+1}(\tau) \quad k \geq 1$$

$$dP_0(\tau)/d\tau = - \lambda_0P_0(\tau) + \lambda_1P_1(\tau)$$

- Ecuaciones diferenciales en diferencias
- Representan la dinámica del sistema
- Vamos a ignorar el transitorio y nos preguntamos si esas $P_k(\tau)$ se estabilizan
- Si es así existirán $p_k \equiv \lim_{\tau \rightarrow \infty} P_k(\tau)$
- Y se darán cuando las derivadas frente al tiempo se anulen
- Es decir:

$$0 = - (\lambda_k + \mu_k)p_k + \lambda_{k-1}p_{k-1} + \mu_{k+1}p_{k+1} \quad k \geq 1$$

$$0 = - \lambda_0p_0 + \mu_1p_1$$

- Siendo por supuesto: $\sum_{k=0}^{\infty} p_k = 1$

Resolución del sistema $M/M/1$ en equilibrio

Sistema M/M/1

- Vamos a resolver para $\lambda_k = \lambda$, $\mu_k = \mu$

$$0 = -(\lambda + \mu)p_k + \lambda p_{k-1} + \mu p_{k+1} \quad k \geq 1$$

$$0 = -\lambda p_0 + \mu p_1$$

- Vemos que $p_1 = p_0 \lambda / \mu$
- Sustituimos en la primera para $k=1$ y queda:

$$0 = -(\lambda + \mu) p_0 \lambda / \mu + \lambda p_0 + \mu p_2$$

$$\text{• Y despejando } p_2 = p_0 (\lambda / \mu)^2$$

$$\text{• Continuando llegamos a } p_k = p_0 (\lambda / \mu)^k = p_0 \rho^k$$

- Para calcular p_0 recurrimos a que:

$$\sum_{k=0}^{\infty} p_k = 1$$

$$1 = \sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} p_0 \rho^k = p_0 \sum_{k=0}^{\infty} \rho^k$$

$$p_0 = \frac{1}{\sum_{k=0}^{\infty} \rho^k} = \frac{1}{1 + \sum_{k=1}^{\infty} \rho^k} = \frac{1}{1 + \frac{\rho}{1 - \rho}} = 1 - \rho$$

$\rho < 1$

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

Sistema M/M/1

- $p_0 = 1 - \rho$
- Y con eso $p_k = p_0(\lambda/\mu)^k = (1 - \rho) \rho^k$
- Así pues:

$$p_k = (1 - \rho) \rho^k \quad k \geq 0$$
- No es más que la distribución geométrica
- Da la probabilidad estacionaria de encontrar k clientes en el sistema
- No depende de λ y μ sino de su cociente ρ
- Podemos calcular el número medio de clientes en el sistema:

$$\begin{aligned}
 N &= \sum_{k=0}^{\infty} k p_k = (1 - \rho) \sum_{k=0}^{\infty} k \rho^k = (1 - \rho)(\rho + 2\rho^2 + 3\rho^3 + 4\rho^4 + \dots) = (1 - \rho)\rho(1 + 2\rho + 3\rho^2 + 4\rho^3 + \dots) = \\
 &= (1 - \rho)\rho \frac{d}{d\rho} (\rho + \rho^2 + \rho^3 + \dots) = (1 - \rho)\rho \frac{d}{d\rho} \sum_{k=0}^{\infty} \rho^k = (1 - \rho)\rho \frac{d}{d\rho} \frac{1}{1 - \rho} = (1 - \rho)\rho \frac{1}{(1 - \rho)^2} = \frac{\rho}{1 - \rho}
 \end{aligned}$$

$$\boxed{N = \frac{\rho}{1 - \rho}}$$

Sistema M/M/1

$$N = \frac{\rho}{1-\rho}$$

- Empleando Little, $N = \lambda S$, con lo que el tiempo medio en el sistema:

$$S = \frac{1/\mu}{1-\rho}$$

- El tiempo medio de espera en cola será $W = S - 1/\mu$, es decir:

$$W = \frac{1/\mu}{1-\rho} - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}$$

- Y el número medio de cliente en cola será (Little) $N_q = \lambda W$

$$N_q = \lambda W = \lambda \frac{\rho}{\mu(1-\rho)} = \frac{\rho^2}{1-\rho}$$

- Que evidentemente es:

$$N_q = \frac{\rho^2}{1-\rho} = \frac{\rho - \rho + \rho^2}{1-\rho} = \frac{\rho - \rho(1-\rho)}{1-\rho} = \frac{\rho}{1-\rho} - \rho = N - \rho$$

$$N_q = N - \rho$$

Estadísticos del sistema M/M/1

Sistema M/M/1

- Resumiendo:

- Número medio de clientes en el sistema (cola+servidor)

$$N = \frac{\rho}{1-\rho}$$

- Número medio de cliente en la cola

$$N_q = \frac{\rho^2}{1-\rho}$$

- Número medio de clientes en el servidor

$$N_s = \rho$$

- Tiempo medio de espera en el sistema (cola+servidor)

$$S = \frac{1/\mu}{1-\rho}$$

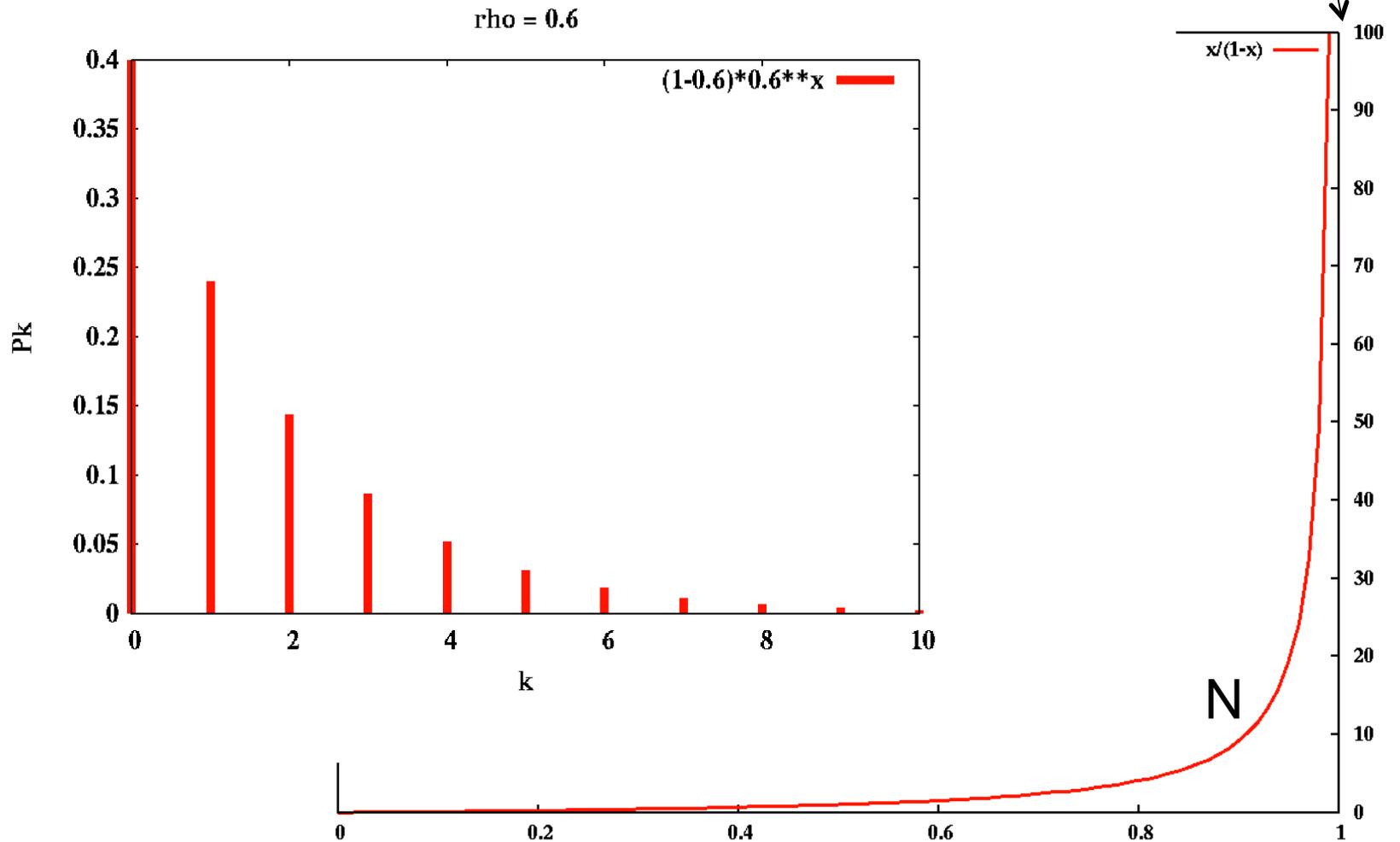
- Tiempo medio de espera en la cola

$$W = \frac{\rho}{\mu(1-\rho)}$$

- Tiempo medio de espera en el servidor es simplemente $1/\mu$

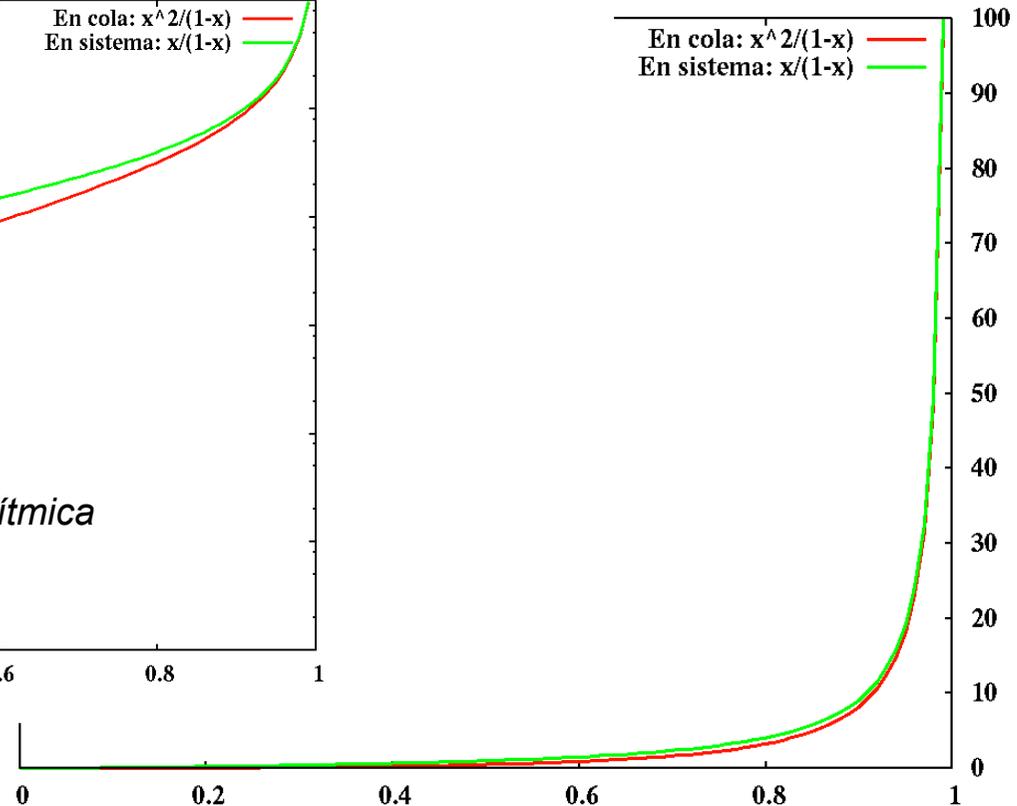
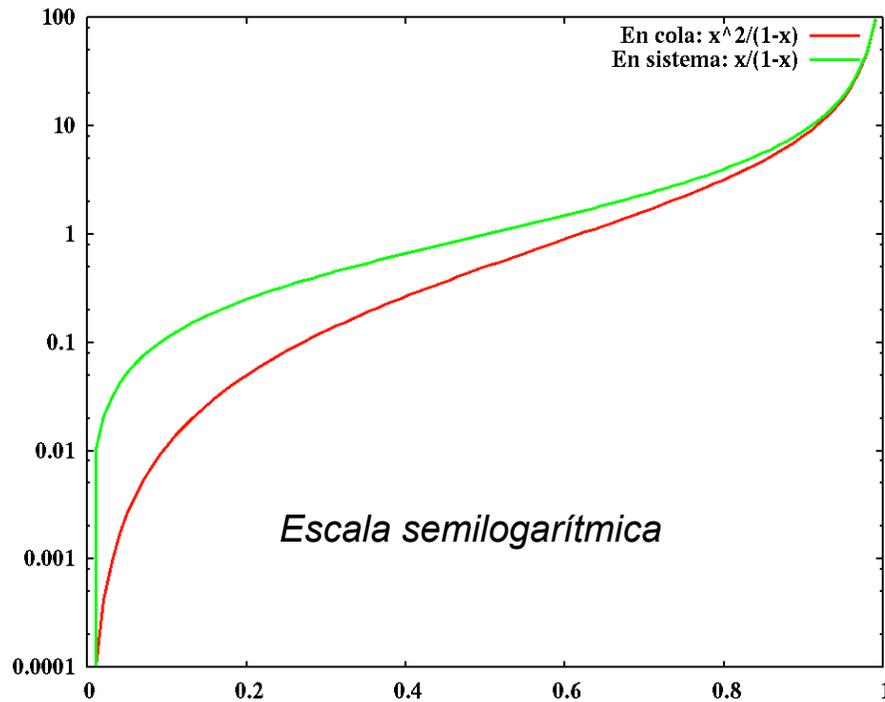
M/M/1 : Clientes en sistema

- Probabilidad: $p_k = (1 - \rho) \rho^k \quad k \geq 0$
- Media: $N = \frac{\rho}{1 - \rho}$



M/M/1: Clientes en cola

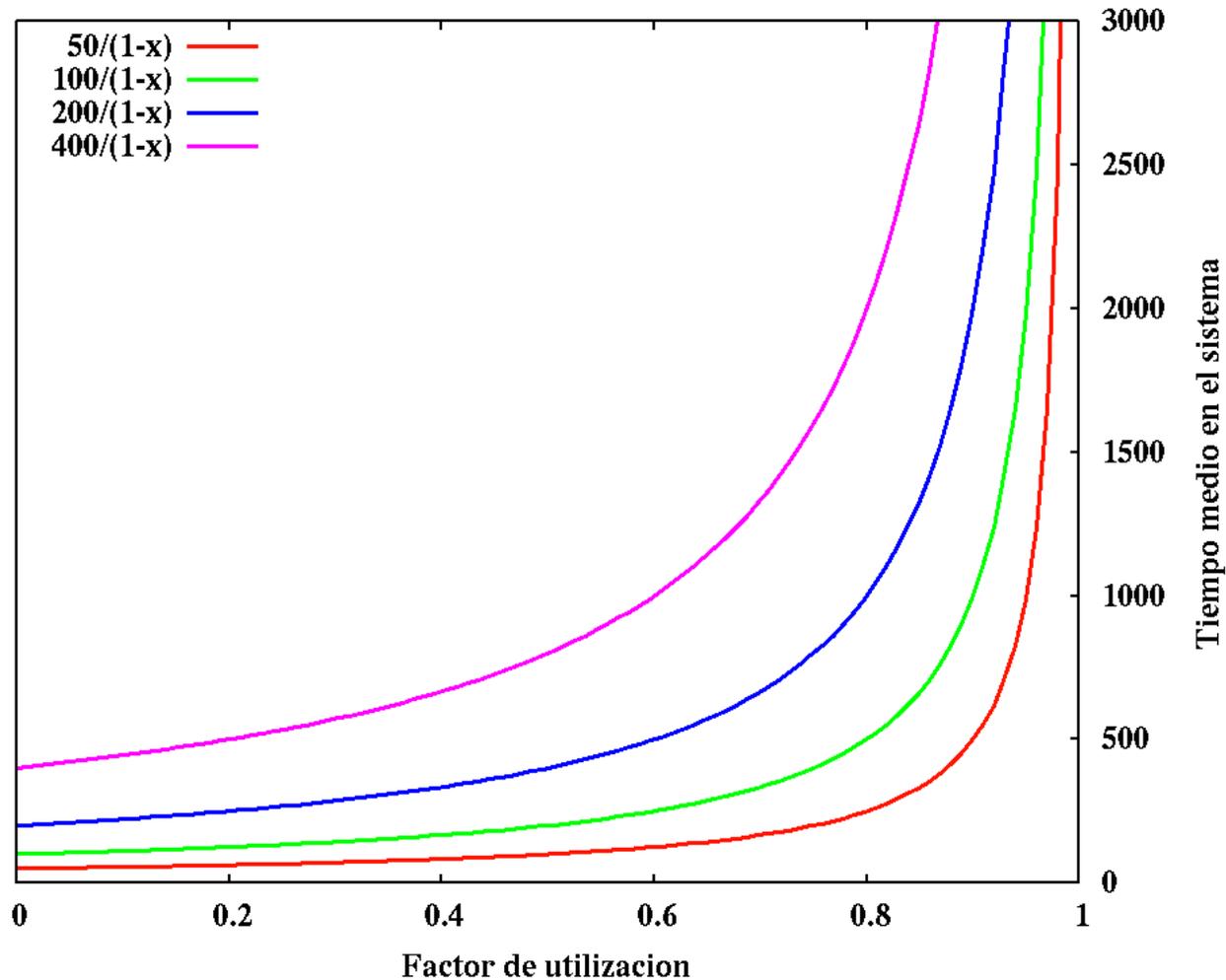
- Media en cola: $N_q = \frac{\rho^2}{1-\rho}$
- Media en el sistema era: $N = \frac{\rho}{1-\rho}$
- La diferencia es solo ρ



M/M/1 : Espera en sistema

- Veamos el cambio para diferentes tiempos medios de servicio

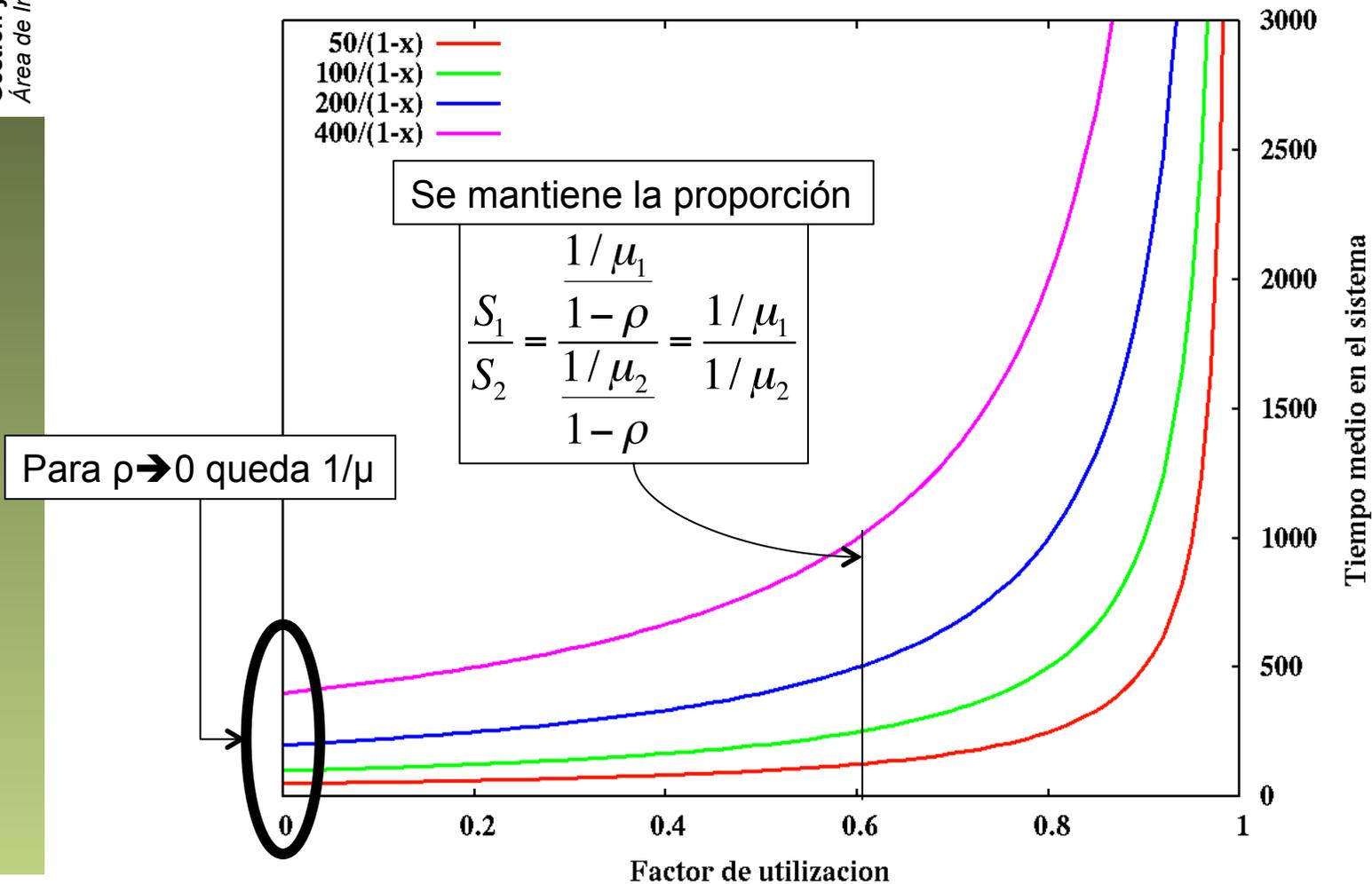
- Media:
$$S = \frac{1/\mu}{1-\rho}$$



M/M/1 : Espera en sistema

- Veamos el cambio para diferentes tiempos medios de servicio

- Media: $S = \frac{1/\mu}{1-\rho}$



M/M/1 otro cálculo de S

- Otra forma de calcular el tiempo medio en el sistema
- Tiempo medio de servicio $1/\mu$
- Supongamos una llegada
- El tiempo medio en el sistema será la suma de (*):
 - El tiempo medio de servir a todos los que hay en el sistema
 - Supongamos que hay N clientes en el sistema
 - Tardan N/μ en salir del sistema (**)
 - El tiempo medio de servir a esta llegada
- Suman $S = N/\mu + 1/\mu$
- Y por Little Sabemos que $N = \lambda S$
- Luego $S = \lambda S/\mu + 1/\mu = \rho S + 1/\mu$
- Y despejando

$$S = \frac{1/\mu}{1-\rho}$$
- (*) gracias a que las llegadas sean de Poisson y (**) a que los tiempos de servicio no tengan memoria

Sistema ¿"M/M/1"?

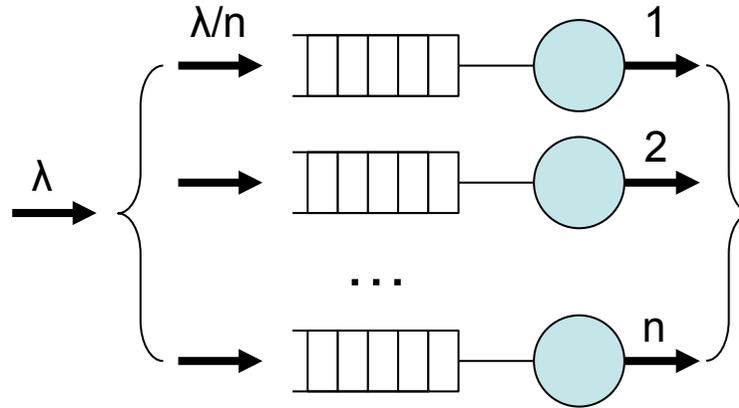
- Notación de Kendall
- A/S/c/K/N/D
- A : Describe los tiempos entre llegadas
 - M : i.i.d. exponenciales (Poisson)
 - D : deterministas (siempre el mismo valor)
 - G : distribución general (suelen ser i.i.d. aunque algunos ponen GI para indicarlo expresamente)
- S : Describe los tiempos de servicio
 - Mismos casos que para A (en este caso M son tiempos de servicio exp.)
- c : Número de servidores
- K : Capacidad máxima del sistema (clientes), igual a la capacidad máxima de la cola + c
- N : Tamaño de la población (número máximo de clientes existente, pueden salir y volver a entrar)
- D : La disciplina de cola (FIFO/LIFO/PS, etc)



Servidores en paralelo vs servidor más rápido

Servidores en paralelo

- Llegadas se reparten entre n servidores (colas infinitas)
- Reparto tal que cada uno atiende a un proceso de Poisson λ/n



$$S_1 = \frac{1/\mu}{1-\rho_1}$$

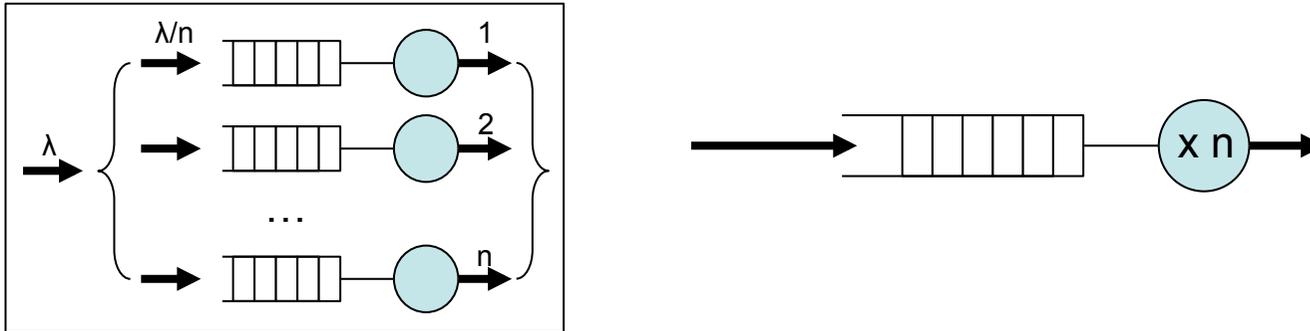
1 servidor

- El tiempo en el sistema ahora es: $S = \frac{1/\mu}{1-\frac{1}{n}\rho_1}$
- Con $n \gg 1$ el tiempo en el sistema tiende al tiempo de servicio

- Definimos ahora: $\rho = \frac{\lambda}{n\mu}$
- Debe ser $\rho < 1$

Un servidor más rápido

- Comparemos lo anterior con tener un solo servidor pero n veces más rápido



- Los n servidores con tiempo medio de servicio $1/\mu$
- Este servidor con tiempo medio de servicio $1/\mu_n = 1/(n\mu)$
- Tiempo en sistema:

$$S_n = \frac{1/\mu}{1 - \frac{1}{n}\rho_1} \quad \longleftrightarrow \quad S = \frac{1/n\mu}{1 - \frac{1}{n}\rho_1} = \frac{S_n}{n}$$

- Tiempo en el sistema n veces menor

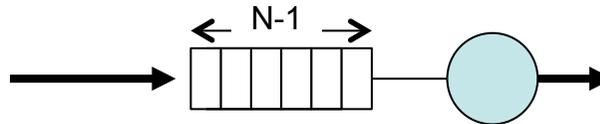


M/M/1/N



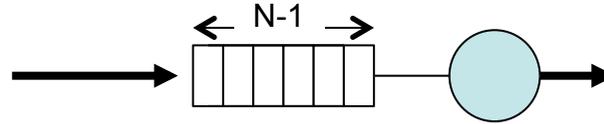
M/M/1/N

- Supongamos una cola en que el número máximo de clientes en ella es $N-1$
- Es decir, el número máximo de cliente en el sistema es N



- Las llegadas son un proceso de Poisson
- El tiempo de servicio es exponencial (i.i.d.)
- ¿Cuál es la probabilidad de pérdida?

M/M/1/N



$$0 = -(\lambda + \mu)p_k + \lambda p_{k-1} + \mu p_{k+1} \quad N \geq k \geq 1$$

$$0 = -\lambda p_0 + \mu p_1$$

- Los p_k con $k > N$ no tienen sentido (valen 0)
- Es decir, la primera ecuación es para $N \geq k \geq 1$
- Se resuelve y la probabilidad estacionaria de estado sigue siendo $p_k = p_0 \rho_1^k$

- Pero ahora el p_0 vale:
$$p_0 = \frac{1 - \rho_1}{1 - \rho_1^{N+1}}$$

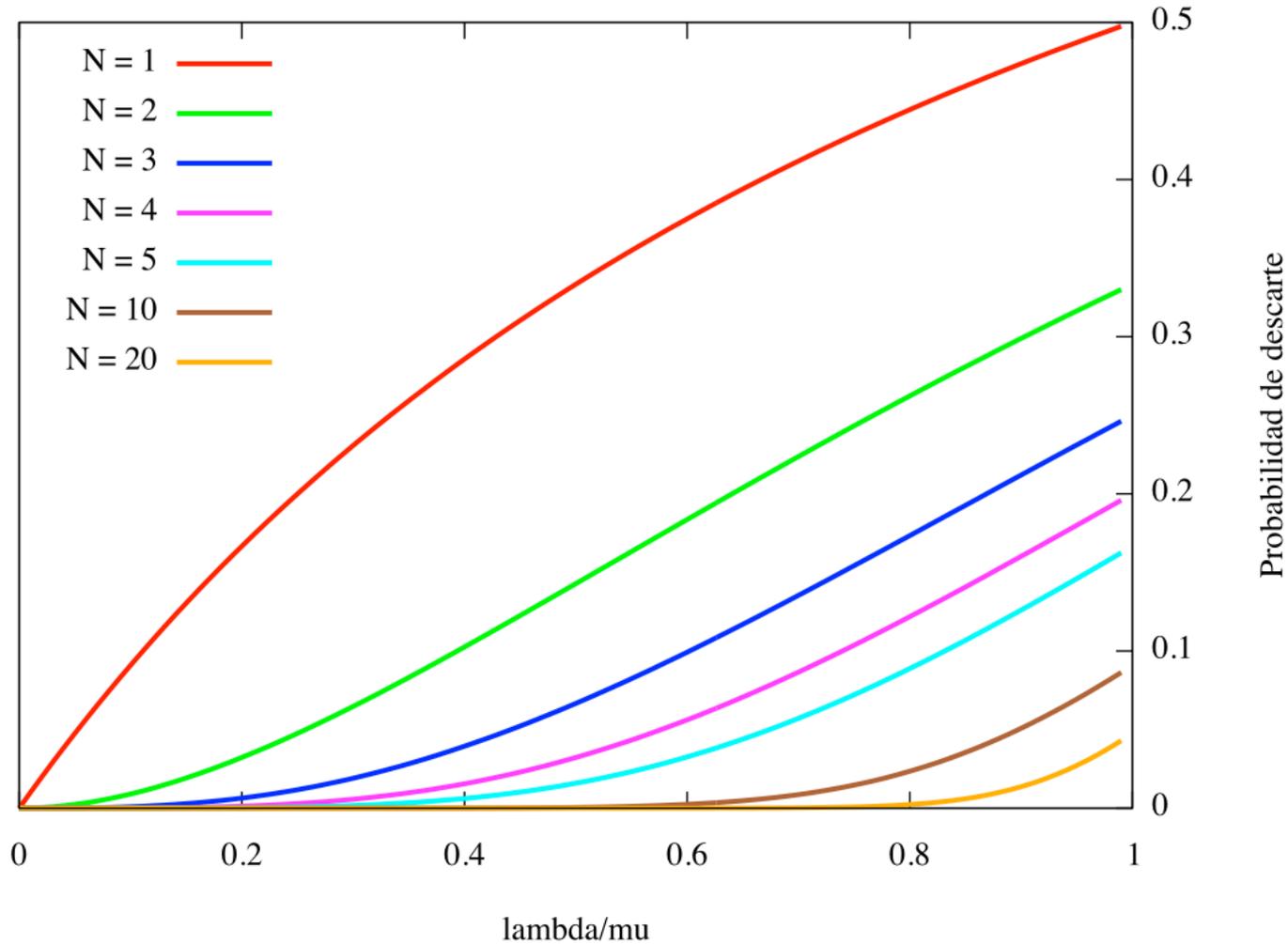
- Así que ($0 \leq k \leq N$):
$$p_k = \frac{(1 - \rho_1) \rho_1^k}{1 - \rho_1^{N+1}}$$

- Y la probabilidad de sistema lleno es:
$$p_N = \frac{(1 - \rho_1) \rho_1^N}{1 - \rho_1^{N+1}}$$
- Que sería la probabilidad de pérdida

$\rho_1 = \lambda/\mu$ porque ahora ya no es el factor de utilización

M/M/1/N: pérdidas

$$p_N = \frac{(1 - \rho_1)\rho_1^N}{1 - \rho_1^{N+1}}$$

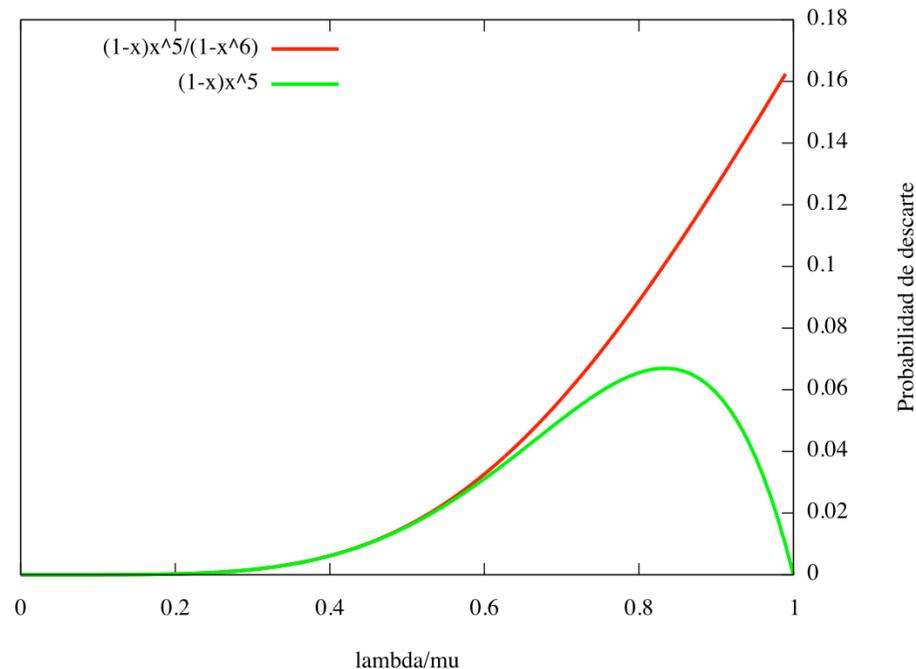


M/M/1/N: ρ pequeño

- Para un factor de utilización pequeño podemos aproximar:

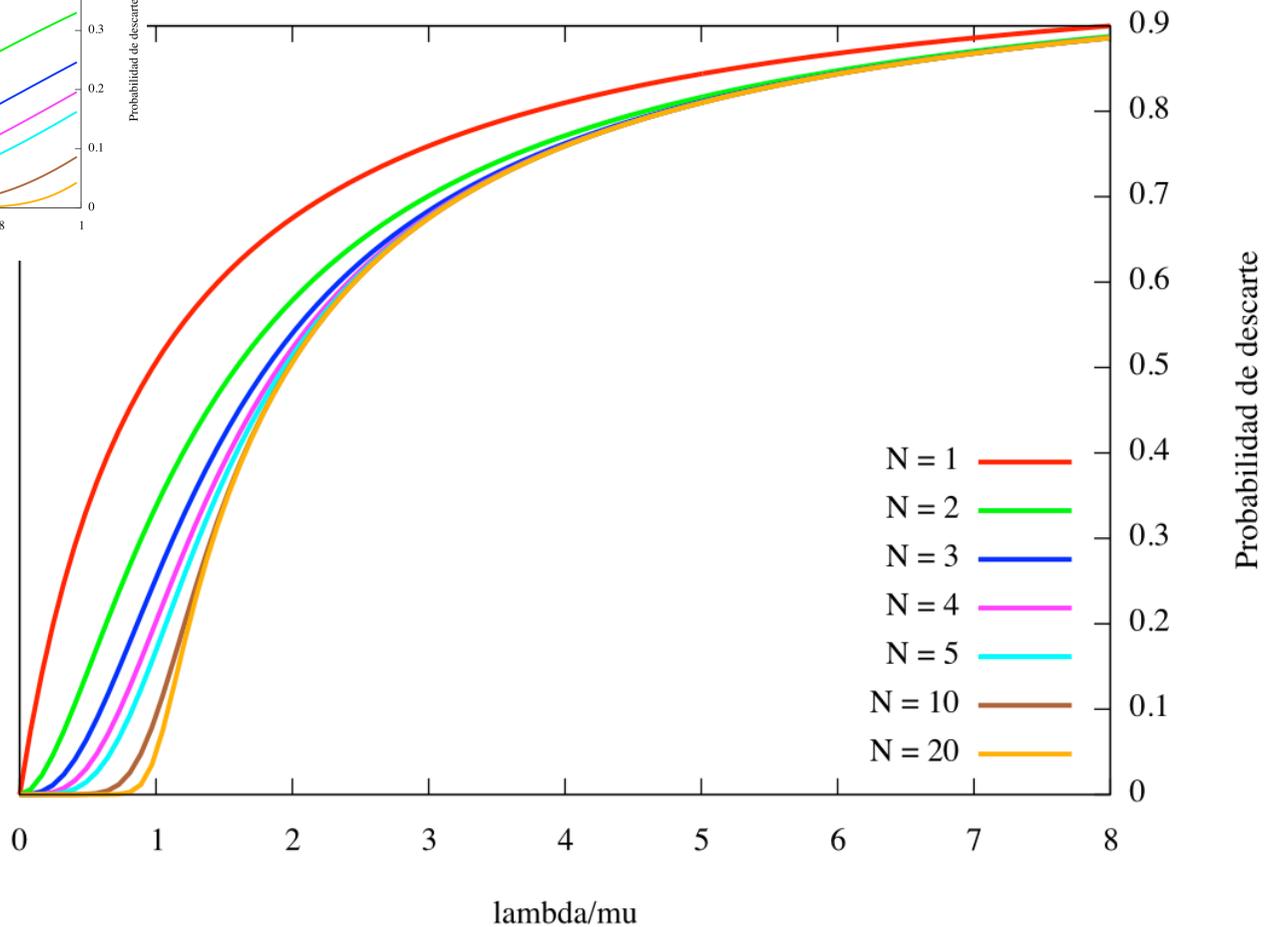
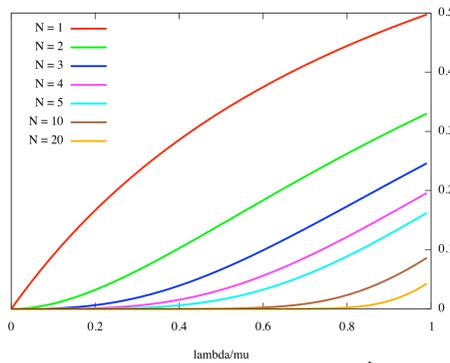
$$p_N = \frac{(1 - \rho_1)\rho_1^N}{1 - \rho_1^{N+1}} \sim (1 - \rho_1)\rho_1^N$$

- Esto es precisamente la probabilidad de que en un M/M/1 (cola infinita) nos encontremos en el estado N
- Es decir, para $\rho_1 \ll 1$ podemos aproximar la probabilidad de estar en el estado N con el valor para el caso de cola infinita



M/M/1/N: ρ grande

- Este sistema no requiere que $\rho < 1$
- Al ser finito el número de estados siempre existe el equilibrio



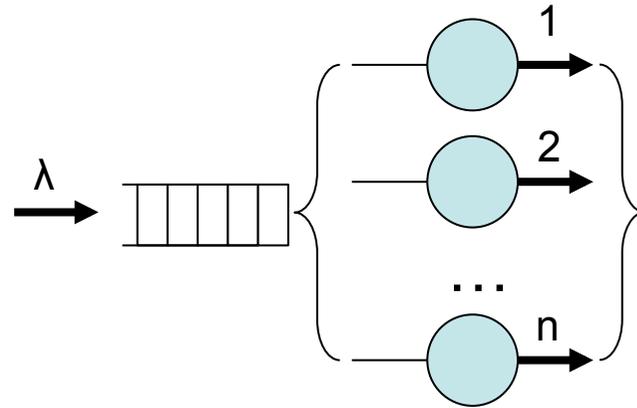
$$p_N = \frac{(1-\rho)\rho^N}{1-\rho^{N+1}}$$

Probabilidad de descarte

Algunos sistemas multi-servidor

M/M/n

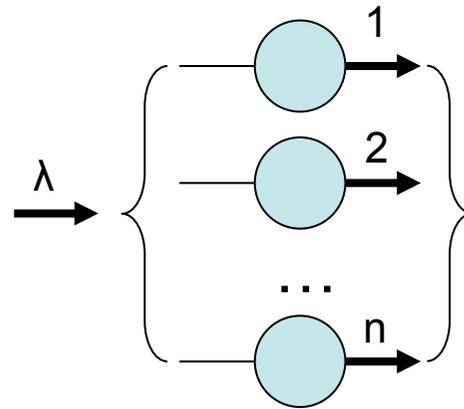
- Una cola compartida para n servidores
- Cuando queda un servidor libre atiende al primero de la cola



- Si hay k servidores ocupados con tiempos de servicio exponenciales entonces hay salidas a una tasa $k\mu$

M/M/n/n

- n servidores sin cola
- Mientras todos los servidores están ocupados se rechazan llegadas



- Probabilidad estacionaria:
- (...)

$$p_k = \frac{\rho^k}{k!} \bigg/ \sum_{i=0}^n \frac{\rho^i}{i!} \quad 0 \leq k \leq n$$

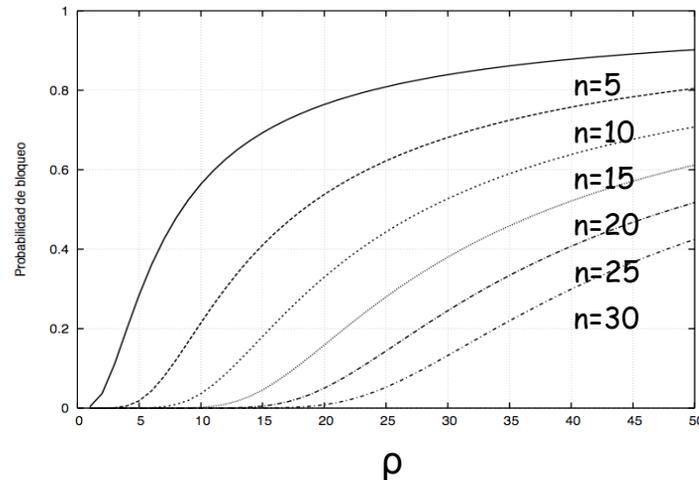
M/M/n/n

- Probabilidad estacionaria:

$$p_k = \frac{\rho^k}{\sum_{i=0}^n \frac{\rho^i}{i!}} \quad 0 \leq k \leq n$$
- Probabilidad de sistema lleno (y descartes ante llegadas) es p_n

$$p_n = \frac{\rho^n}{\sum_{i=0}^n \frac{\rho^i}{i!}}$$

- Y por si a alguien aún no le suena, eso es la Erlang-B



Otras preguntas

- ¿Y la evolución con el tiempo de las probabilidades de estado hasta estabilizarse?
- ¿El proceso de salida cómo es? ¿Es de Poisson?
- ¿Y la distribución del tiempo de espera?
- ¿Y si a continuación viene otra cola?
- ¿Y si el número de clientes es finito?
- ¿Y si los tiempos de servicio no son exponenciales?
- ¿Y si las llegadas no son de Poisson?
- ¿Y si puede haber realimentación?

