Probabilidades Y Estadística II

TEMA 2. MODELOS DE COLAS

CONTENIDOS

1. **Introducción a las colas poissonianas.**
2. **Modelo de colas poissoniano con un servidor *M/M/1***
3. **Modelo con un servidor y capacidad finita *M/M/1/K***

4. Modelo *M/M/c*: *c* servidores en paralelo

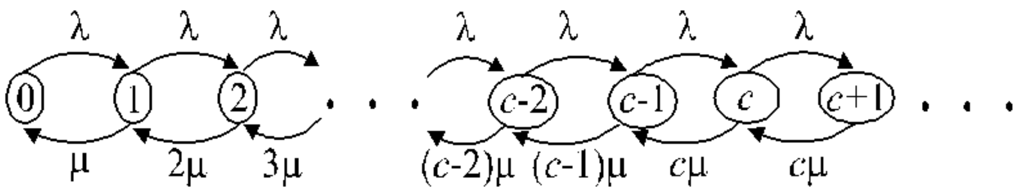
Se dispone de *c* servidores paralelos idénticos, cada uno de los cuales sirve a una tasa de μ clientes por unidad de tiempo.

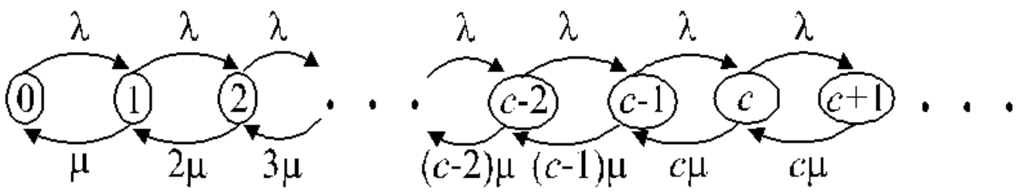
Luego, si los *c* están utilizándose, la tasa media de salida del servicio es *c*μ. Cuando hay *n* < *c* clientes en el sistema, sólo trabajan *n* servidores y, por tanto, la tasa de servicio es *n*μ. Es decir, las tasas de nacimiento y muerte son

$$\lambda_n = \lambda, \qquad n = 0, 1, 2, \ldots$$

$$\mu_n = \begin{cases} n\mu, & \text{si } n = 1, 2, \ldots, c \\ c\mu, & \text{si } n \geq c \end{cases}$$

y su diagrama de transición es





El sistema de ecuaciones en equilibrio es:

Estado	Tasa entrada	=	Tasa salida

$$
\begin{aligned}
\pi_0\lambda &= \pi_1\mu \\
\pi_1(\lambda+\mu) &= \pi_0\lambda + \pi_2 2\mu \\
\pi_2(\lambda+\mu) &= \pi_1\lambda + \pi_3 3\mu \\
&\cdots \\
\pi_{c-1}(\lambda+(c-1)\mu) &= \pi_{c-2}\lambda + \pi_c c\mu \\
\pi_c(\lambda+c\mu) &= \pi_{c-1}\lambda + \pi_{c+1}c\mu \\
\pi_{c+1}(\lambda+c\mu) &= \pi_c\lambda + \pi_{c+2}c\mu \\
&\cdots \\
\sum_{i=0}^{\infty}\pi_i &= 1
\end{aligned}
\quad\Rightarrow\quad
\begin{aligned}
\pi_0\lambda &= \pi_1\mu \\
\pi_1\lambda &= \pi_2 2\mu \\
\pi_2\lambda &= \pi_3 3\mu \\
&\cdots \\
\pi_{c-1}\lambda &= \pi_c c\mu \\
\pi_c\lambda &= \pi_{c+1}c\mu \\
\pi_{c+1}\lambda &= \pi_{c+2}c\mu \\
&\cdots \\
\sum_{i=0}^{\infty}\pi_i &= 1
\end{aligned}
\quad\Rightarrow\quad
\begin{aligned}
\pi_1 &= \frac{\lambda}{\mu}\pi_0 \\
\pi_2 &= \left(\frac{\lambda}{\mu}\right)^2\frac{1}{2!}\pi_0 \\
&\cdots \\
\pi_{c-1} &= \left(\frac{\lambda}{\mu}\right)^{c-1}\frac{1}{(c-1)!}\pi_0 \\
\pi_c &= \left(\frac{\lambda}{\mu}\right)^{c}\frac{1}{c!}\pi_0 \\
\pi_{c+1} &= \left(\frac{\lambda}{\mu}\right)^{c+1}\frac{1}{c!c}\pi_0 \\
&\cdots \\
\sum_{i=0}^{\infty}\pi_i &= 1
\end{aligned}
$$

Obteniendo finalmente (r = λ/μ es la intensidad de tráfico):

$$\pi_0 = \left[\sum_{n=0}^{c-1}\frac{r^n}{n!} + \frac{r^c}{c!}\sum_{n=c}^{\infty}(r/c)^{n-c}\right]^{-1} = \left[\sum_{n=0}^{c-1}\frac{r^n}{n!} + \frac{r^c}{c!(1-\frac{\lambda}{c\mu})}\right]^{-1}$$

y

$$\left\{ \frac{r^n}{\cdots} \right.$$

de estabilidad.

Una probabilidad de interés en este modelo es la probabilidad de tener que esperar en la cola (todos los servidores están ocupados), es decir, $P(N \geq c)$, que se denota como $C(c, r)$, llamada fórmula C de Erlang:

$$C(c,r) = P(N \geq c) = 1 - P(N < c) = 1 - \sum_{n=0}^{c-1} \pi_n = 1 - \sum_{n=0}^{c-1} \frac{r^n}{n!}\pi_0$$

$$= 1 - \pi_0 \left(\frac{1}{\pi_0} - \frac{r^c}{c!(1-\frac{\lambda}{c\mu})} \right) = \pi_0 \frac{r^c}{c!(1-\rho)} = \frac{\pi_c}{1-\rho}$$

$$= \frac{r^c/c!}{(1-\rho)\left[\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{r^c}{c!(1-\rho)}\right]},$$

Normalmente, se deja al software (por ejemplo, WinQSB) que calcule los valores $C(c,r)$, si bien tradicionalmente se obtenían de forma aproximada a partir de su representación gráfica (Allen, 1978). Hoy en día es muy sencillo programar estas fórmulas.

Medidas de rendimiento

Comenzamos calculando L_q, por ser más sencillo que L

$$L_q = E(N_q) = \sum_{n=c}^{\infty}(n-c)\pi_n = \sum_{n=0}^{\infty} n\pi_{n+c} = \sum_{n=0}^{\infty} n\frac{r^{n+c}}{c!c^n}\pi_0 = \sum_{n=0}^{\infty} n\frac{r^c r^n}{c!c^n}\pi_0 =$$

$$= \sum_{n=0}^{\infty} n\frac{r^c\rho^n}{c!}\pi_0 = \pi_0\frac{r^c}{c!}\sum_{n=0}^{\infty} n\rho^n = \pi_0\frac{r^c}{c!}\frac{\rho}{(1-\rho)^2} = C(c,r)\frac{\rho}{1-\rho},$$

$$E(q \mid q > 0) = \frac{W_q}{P(q > 0)} = \frac{W_q}{C(c,r)} = \frac{1}{c\mu(1-\rho)}.$$

A partir de aquí, podemos conseguir expresiones para *W* y *L* :

$$W = W_q + W_s = \frac{1}{\mu}\left(1 + \frac{C(c,r)}{c(1-\rho)}\right)$$
$$L = \lambda W = c\rho + C(c,r)\frac{\rho}{1-\rho}$$

Obtengamos las distribuciones de las v.a. *q* y *w*.

Para *q*, debemos tener en cuenta que el cliente que no espera en cola ($q = 0$) es el que al llegar encuentra en el sistema $N = n < c$ clientes. En caso contrario, con $n \geq c$, la longitud de la cola es $n - c$ y el cliente tendrá que esperar a que se sirvan n-c+1 clientes (el que está siendo servido también cuenta).

De este modo, su tiempo en cola es $q = s_1 + \ldots + s_{n-c+1}$, con s_i v.a.i.i.d. según Exp($c\mu$), que conduce a que *q* siga una distribución gamma de parámetros $p = n-c+1$ y $a = c\mu$.

Por lo tanto, para $t \geq 0$

$$F_q(t) = P(q=0) + P(0 < q \leq t) = 1 - C(c,r) + \sum_{n=c}^{\infty} P(q \leq t \mid N = n)\pi_n$$
$$= 1 - C(c,r) + \sum_{n=c}^{\infty}\left[\int_0^t c\mu e^{-c\mu x}\frac{(c\mu x)^{n-c}}{(n-c)!}dx\right]\frac{r^n}{c!c^{n-c}}\pi_0$$
$$= 1 - C(c,r) + \frac{\pi_0 r^c}{\cdots}\int_0^t \mu e^{-c\mu x}\left(\sum^{\infty}\frac{(r\mu x)^{n-c}}{\cdots}\right)dx$$
$$\cdots = 1 - C(c,r) + C(c,r)(1 - e^{\cdots}) = 1 - C(c,r)e^{\cdots}.$$

En los dos últimos pasos utilizamos que $c - r = c\,(1 - \rho)$.

q tiene un punto (t=0) con probabilidad positiva: $P(q=0) = P(N < c) = 1 - C(c, r)$.

Análogamente, podemos obtener la distribución de *w:*

$$F_w(t) = \begin{cases} 1 + \dfrac{r - c + 1 - C(c,r)}{c - 1 - r} e^{-\mu t} + \dfrac{C(c,r)}{c - 1 - r} e^{-(1-\rho)c\mu t}, & \text{si } r \neq c - 1 \\ 1 - [1 + C(c,r)\mu t]e^{-\mu t}, & \text{si } r = c - 1 \end{cases}$$

Obviamente, tomando c = 1, recuperamos las fórmulas del modelo *M/M/1.*

5. Modelo *M/M/∞*: infinitos servidores

El sistema de espera tiene un número ilimitado de servidores, lo que significa que cada cliente que llega es servido inmediatamente.

A pesar de no haber competencia ni compartición de recursos, los resultados de este modelo pueden servirnos para estimar cantidades de interés en sistemas con un número c suficientemente grande de servidores.

Las tasas de nacimiento y muerte son

$$\begin{aligned}\pi_0\lambda &= \pi_1\mu \\ \pi_1(\lambda+\mu) &= \pi_0\lambda + 2\mu\pi_2 \\ \pi_2(\lambda+2\mu) &= \pi_1\lambda + 3\mu\pi_3 \\ &\dots \\ \pi_n(\lambda+n\mu) &= \pi_{n-1}\lambda + (n+1)\mu\pi_{n+1} \\ &\dots \\ \sum \pi_i &= 1\end{aligned} \Rightarrow \begin{aligned}\pi_0\lambda &= \pi_1\mu \\ \pi_1\lambda &= 2\mu\pi_2 \\ \pi_2\lambda &= 3\mu\pi_3 \\ &\dots \\ \pi_n\lambda &= (n+1)\mu\pi_{n+1} \\ &\dots \\ \sum \pi_i &= 1\end{aligned} \Rightarrow$$

$$\Rightarrow \begin{aligned}\pi_1 &= \left(\frac{\lambda}{\mu}\right)\pi_0 \\ \pi_2 &= \frac{1}{2!}\left(\frac{\lambda}{\mu}\right)^2\pi_0 \\ \pi_3 &= \frac{1}{3!}\left(\frac{\lambda}{\mu}\right)^3\pi_0 \\ &\dots \\ \pi_n &= \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\pi_0 \\ &\dots \\ \sum \pi_i &= 1\end{aligned} \Rightarrow \pi_0 = \left[\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right]^{-1} = e^{-\frac{\lambda}{\mu}}$$

$$\Downarrow$$

$$\pi_n = \frac{(\lambda/\mu)^n}{n!}e^{-\frac{\lambda}{\mu}}, \quad n = 0, 1, \dots$$

Se obtiene que la v.a. N sigue una distribución de Poisson de parámetro $r = \lambda/\mu$, con la condición de estabilidad $r < 1$.

Por tanto, $L = \lambda/\mu$, que indica el número medio de servidores ocupados. Además, $\sigma_N^2 = \lambda/\mu$.

Como no hay cola, $L_q = W_q = 0$.

Ejercicios. Modelo de colas I

Al supercomputador de un centro de cálculo llegan usuarios según un proceso de Poisson con tasa 5 usuarios cada hora. Sabiendo que éstos consumen un tiempo de cómputo aleatorio cuya distribución puede suponerse exponencial de media 10 minutos y que la disciplina de cola es una FIFO, se pide:

a) El número medio de usuarios en el supercomputador y esperando para poder utilizarlo.
b) Suponiendo que hay usuarios esperando, obtenga el tamaño medio de la cola.

c) Si en la sala de espera hay 4 sillas ¿cuál es la probabilidad de que un usuario que llega a la sala tenga que esperar de pie?
d) ¿Cuántas sillas se necesitarían para que un usuario al llegar al sistema tenga una probabilidad menor del 10% de esperar de pie?

e) ¿Qué porcentaje de usuarios que llegan al servidor lo encuentran desocupado?
f) Obtenga el tiempo medio de los usuarios en el sistema y en la cola del mismo.

g) Obtenga la probabilidad de que un usuario espere más de una hora.

Ejercicios. Modelo de colas II

El tráfico en un centro de computación de mensajes, para una de las líneas de salida, llega según un patrón aleatorio de Poisson con un promedio de 240 mensajes por minuto. La línea tiene una velocidad de transmisión de 800 octetos por segundo. La longitud del mensaje es aleatoria con distribución aproximadamente exponencial con longitud media de 176 octetos. Se pide:

a) Calcular las medidas estadísticas de las prestaciones del sistema desde el punto de vista del usuario suponiendo un número elevado de buffers para mensa-