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***

1. Introducción a los modelos de colas poissonianos

Las colas poissonianas (o exponenciales o markovianas) son modelos del tipo *M/M*, con llegadas de Poisson y servicio exponencial, que son las más estudiadas analíticamente.

Las llegadas de clientes y su servicio demandado son completamente aleatorios en el sentido de que la evolución del sistema depende sólo de su estado actual, y no de su pasado.

Los procesos de nacimiento y muerte introducidos sirven para describir muchos modelos de colas. Asociaremos el término nacimiento con la llegada de un cliente al sistema y el término muerte con la salida de un cliente del sistema después de completado su servicio. El número de clientes en el sistema en el instante t, $N(t)$, indica el estado del mismo.

Estudiaremos el comportamiento de las probabilidades $p_n(t)$ en el límite $\pi_n = \lim_{t\to\infty} p_n(t)$, que indica la proporción de tiempo que el sistema permanece con n clientes.

Tema 2. Modelos de Colas Poissonianos Probabilidades y Estadística II

La solución de equilibrio (tema 11) se obtenía de las ecuaciones que igualaban las tasas de entrada y salida de cada estado, dando lugar a

$$\pi_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1}\cdots\lambda_0}{\mu_n\cdots\mu_1}} \qquad \pi_n = \frac{\lambda_{n-1}\cdots\lambda_0}{\mu_n\cdots\mu_1}\pi_0, \quad n = 1, 2, 3\ldots \qquad \textbf{(10.1)}$$

Para que exista dicha solución de equilibrio se debe satisfacer

las que existe solución de equilibrio π_n, mientras que con las ecuaciones de π_0 y π_n obtendremos dicha solución.

2. Modelo de colas poissoniano con un servidor *M/M/1*

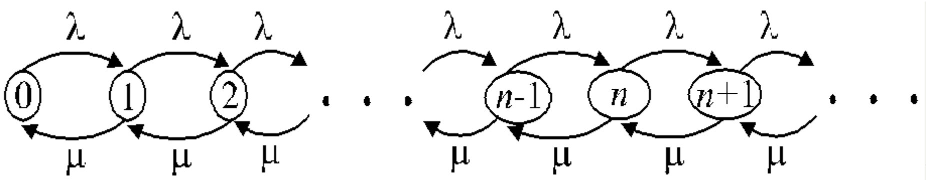
En este modelo se dispone sólo de un canal para dar servicio, las llegadas siguen un proceso de Poisson y la distribución del tiempo de servicio es exponencial.

Así, las tasas de nacimiento y muerte no dependen del número de clientes en el sistema y

$$\lambda_n = \lambda, \ \ n = 0,1,2,\ldots \qquad \mu_n = \mu, \ \ n = 1,2,3,\ldots$$

La capacidad del sistema es ilimitada y la disciplina de la cola es FIFO.

La siguiente figura representa el diagrama de transición



que conduce al sistema de ecuaciones en equilibrio

Estado	Tasa entrada	=	Tasa salida
0	$\mu\pi_1$	=	$\lambda\pi_0$
$n \geq 1$	$\lambda\pi_{n-1} + \mu\pi_{n+1}$	=	$(\lambda+\mu)\pi_n$

$$\pi_0\lambda = \pi_1\mu \qquad \pi_0\lambda = \pi_1\mu \qquad \pi_1 = \rho\pi_0$$

$$\pi_1(\lambda+\mu) = \pi_0\lambda + \pi_2\mu \qquad \pi_1\lambda = \pi_2\mu \qquad \pi_2 = \rho\pi_1 = \rho^2\pi_0$$

Sustituyendo las expresiones de los π_i en la última ecuación y despejando π_0 obtenemos (teniendo en cuenta que el factor de utilización es $\rho = r = \lambda/\mu$):

$$\pi_0 = \frac{1}{\sum_{n=0}^{\infty} \rho^n} = \frac{1}{1/(1-\rho)} = 1-\rho$$

$$\pi_n = (1-\rho)\rho^n, \qquad n = 0, 1, 2, 3...$$

que corresponde a una distribución geométrica de parámetro 1- ρ.

$$S_1 = \sum_{n=0}^{\infty} \rho^n < \infty \quad S_2 = \frac{1}{\lambda} \sum_{n=0}^{\infty} \frac{1}{\rho^n} = \infty$$

La serie de S_1 converge si y sólo si $\rho < 1$. La segunda condición (S_2) se satisface si $\rho \leq 1$.

Luego, la condición necesaria y suficiente para que un modelo *M/M/1* tenga solución de equilibrio, es que $\rho<1$, que es la condición de estabilidad.

Por tanto, la probabilidad de que el canal esté ocupado es

$$P(\text{canal esté ocupado}) = 1 - \pi_0 = 1 - (1 - \rho) = \rho$$

La probabilidad de encontrar al menos *n* de clientes en el sistema

$$P(N \geq n) = \sum_{k=n}^{\infty} \pi_k = \sum_{k=n}^{\infty} (1-\rho)\rho^k = (1-\rho)\rho^n \sum_{k=n}^{\infty} \rho^{k-n}$$

Medidas de rendimiento

Comenzando por el número medio de clientes en el sistema, *L*, y en la cola, L_q. Se tiene

$$\begin{aligned} L = E(N) &= \sum_{n=0}^{\infty} n\pi_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n = (1-\rho)\sum_{n=0}^{\infty} n\rho^n \\ &= (1-\rho)\rho\sum_{n=1}^{\infty} n\rho^{n-1} = (1-\rho)\rho\frac{d}{d\rho}\left(\sum_{n=0}^{\infty}\rho^n\right) \\ &= (1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) = \frac{(1-\rho)\rho}{(1-\rho)^2} = \frac{\rho}{1-\rho}. \end{aligned}$$

La sexta igualdad se debe a que las operaciones de suma y diferenciación pueden intercambiarse cuando las funciones implicadas se comportan lo suficientemente bien.

Otra expresión equivalente, en función de λ y μ, es $L = \dfrac{\lambda}{\mu - \lambda}$ ya que $\rho = \lambda/\mu$.

L también podía haberse deducido directamente por tener *N* distribución geométrica. Nótese que *L*, como función de *ρ*, tiene una asíntota vertical en *ρ*=1, lo que indica el dramático comportamiento del número medio de clientes en el sistema según nos acercamos hacia la violación de la condición de estabilidad.

Aparte de la media que acabamos de calcular de la variable *N*, podemos obtener su varianza, a partir de la distribución geométrica

$$\sigma_N^2 = \sum_{n=0}^{\infty}(n-L)^2\pi_n = \frac{\rho}{(1-\rho)^2}$$

$$= L - (1-\pi_0) = L - \rho = \frac{\rho^2}{1-\rho},$$

que en términos de λ y μ es $L_q = \dfrac{\lambda^2}{\mu(\mu-\lambda)}$

Nótese que la igualdad $L_q = L - (1 - \pi_0)$ es general para cualquier cola con un servidor y dando servicio de uno en uno, ya que para obtenerla no se ha utilizado el tipo de distribuciones de los tiempos entre llegadas o de servicio.

Otra relación entre L y L_q es $L_q = \dfrac{\rho}{1-\rho}\rho = L\rho = L\dfrac{\lambda}{\mu} \Longrightarrow \dfrac{L_q}{\lambda} = \dfrac{L}{\mu}$

Recordemos que siempre $N = N_q + N_s$. Pero en el modelo que estamos tratando, si $N \geq 1$, entonces $N = N_q + 1$, mientras que en general $L \neq L_q+1$, ya que L y L_q son medias y hay momentos en los que el servidor está desocupado.
Además, por las fórmulas de Little el número medio de clientes en el servidor $L_s = \lambda W_s = \rho$.

Calculamos el tamaño esperado de la cola cuando hay cola, denotado como $L_q' = E(N_q | N_q>0)$. Como la probabilidad condicionada de n clientes en el sistema dado que la cola no está vacía,

$\pi_n' = P(n$ clientes en el sistema $| n \geq 2) =$
$P(n$ clientes en el sistema $, n \geq 2)/P(n \geq 2) = \pi_n / (1- \pi_0- \pi_1) = \pi_n / \rho^2$,

para $n \geq 2$, se llega a

$$L_q' = E(N_q \mid N_q > 0) = \sum_{n=2}^{\infty}(n-1)\pi_n' = \sum_{n=2}^{\infty} n\frac{\pi_n}{\rho^2} - \sum_{n=2}^{\infty}\frac{\pi_n}{\rho^2}$$
$$= \frac{(L-\pi_1) - (1-\pi_0-\pi_1)}{\rho^2}$$
$$= (\rho/(1-\rho)) - (1-\pi_0) \quad = 1$$

de donde $E(N_q \mid N_q > 0) = \dfrac{L_q}{P(N_q>0)} = \dfrac{L_q}{P(N\geq 2)} = \dfrac{L_q}{\rho^2} = \dfrac{1}{1-\rho}$.

Tiempos de espera

Sólo resta estudiar los tiempos de espera del modelo *M/M/1*. Obtendremos no sólo las medias W y W_q, sino también las distribuciones de probabilidad de las v.a. w y q.

Las medias se calculan fácilmente por las fórmulas de Little:

$$W = E(w) = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{E(s)}{1-\rho}$$
$$W_q = E(q) = \frac{L_q}{\lambda} = \frac{\rho^2}{\lambda(1-\rho)} = \frac{\rho E(s)}{1-\rho}$$

Nótese que como dijimos para L, W tiene también un comportamiento dramático según ρ tiende a 1.

Ahora queremos calcular el tiempo medio de espera en cola para aquellos clientes que deben esperar.

Como

$$\begin{aligned} W_q &= E(q) = E(q \mid q = 0)P(q = 0) + E(q \mid q > 0)P(q > 0) \\ &= 0(1-\rho) + E(q \mid q > 0)\rho, \end{aligned}$$

se tiene

$$E(q \mid q > 0) = \frac{W_q}{\rho} = \frac{E(s)}{1-\rho} = W$$

más.

Para hallar la distribución de la variable aleatoria q, nótese primero que tiene un punto (t = 0) con probabilidad positiva: $P(q = 0) = P(N = 0) = 1 - \rho$.

Por otra parte, si al llegar el cliente encuentra n personas en el sistema (una de ellas en el servidor), tendrá que esperar a que todas se sirvan. Así, el tiempo de espera en cola es la suma de las variables aleatorias "tiempo de servicio del cliente i ", i =1,...,n,

$$q = s_1+\ldots+s_n$$

en donde s_i son independientes e idénticamente distribuidas según una exponencial de parámetro μ.

Debemos recordar que por la pérdida de memoria de la distribución exponencial, no hace falta tener en cuenta el tiempo de servicio ya consumido por el cliente que actualmente está sirviéndose.

Por la reproductividad de la distribución gamma, $q \mid N=n$ sigue una distribución gamma de parámetros $p = n$, a = μ (en este caso es Erlang, al ser p entero).

Tema 2. Modelos de Colas Poissonianos Probabilidades y Estadística II

Por el teorema de la probabilidad total se tiene

$$\begin{aligned} P(0 < q \leq t) &= \sum_{n=1}^{\infty} P(q \leq t \mid N = n)P(N = n) \\ &= \sum_{n=1}^{\infty} \int_0^t \mu e^{-\mu x} \frac{(\mu x)^{n-1}}{(n-1)!} \rho^n (1-\rho) dx \\ &= \int_0^t (1-\rho) e^{-\mu x} \sum_{n=1}^{\infty} \frac{(\mu \rho x)^{n-1}}{(n-1)!} \rho \mu dx \end{aligned}$$

Esta expresión es válida para todo t, aunque q sea discreta en el origen y continua para $t > 0$.

Análogamente, calculamos la distribución de la v.a. *w*. Si cuando llega un cliente ya hay *n* en el sistema, éste tendrá que estar en el sistema un tiempo total igual a la suma de *n* + 1 v.a.i.i.d. según una ley exponencial de parámetro μ.

Así, la distribución de *w* será una gamma de parámetros *p* = *n*+1, *a* = μ. Variando *n*, por el teorema de la probabilidad total:

$$F_w(t) = P(w \leq t) = \sum_{n=0}^{\infty} P(w \leq t \mid N = n)P(N = n)$$
$$= \sum_{n=0}^{\infty} \int_0^t \mu e^{-\mu x} \frac{(\mu x)^n}{n!} \rho^n (1-\rho) dx = \int_0^t \mu(1-\rho)e^{-\mu x} \sum_{n=0}^{\infty} \frac{(\mu \rho x)^n}{n!} dx$$
$$= \int_0^t \mu(1-\rho)e^{-\mu x} e^{\mu \rho x} dx = 1 - e^{-\mu(1-\rho)t} = 1 - e^{-t/W} \quad (t \geq 0).$$

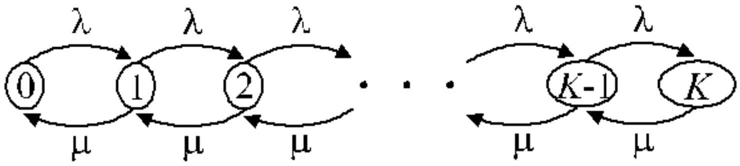
Es decir, *w* sigue una distribución exponencial de parámetro μ(1 – *ρ*) = μ –λ = 1/W.

3. Modelo *M/M/1/K*: Capacidad *K* finita del sistema

Se admite a lo sumo un número *K* de clientes en el sistema, de forma que no se permiten más entradas en el sistema si se alcanza tal cota, siendo rechazadas. Así, las tasas de nacimiento y muerte dependen del número de clientes en el sistema

$$\lambda_n = \begin{cases} \lambda, & \text{si } n = 0, 1, ..., K-1 \\ 0, & \text{si } n \geq K \end{cases}$$

$$\mu_n = \begin{cases} \mu, & \text{si } n = 1, 2, ..., K \end{cases}$$



El sistema de ecuaciones en equilibrio es:

Estado	Tasa entrada	=	Tasa salida
0	$\mu\pi_1$	$=$	$\lambda\pi_0$
$1 \leq n \leq K-1$	$\lambda\pi_{n-1} + \mu\pi_{n+1}$	$=$	$(\lambda+\mu)\pi_n$
K	$\lambda\pi_{K-1}$	$=$	$\mu\pi_K$

Resolviéndolo se obtiene:

$$\lambda\pi_0 = \mu\pi_1 \quad \Rightarrow \pi_1 = \frac{\lambda}{\mu}\pi_0$$

$$\lambda\pi_1 = \mu\pi_2 \quad \Rightarrow \pi_2 = \frac{\lambda}{\mu}\pi_1 = \left(\frac{\lambda}{\mu}\right)^2 \pi_0$$

$$\cdots$$

$$\lambda\pi_{K-1} = \mu\pi_K \quad \Rightarrow \pi_K = \left(\frac{\lambda}{\mu}\right)^K \pi_0$$

Tema 2. Modelos de Colas Poissonianos Probabilidades y Estadística II

Sumando todas las ecuaciones, como $\sum_{n=0}^{K} \pi_n = 1$, pues $\pi_n = 0$ para $n > K$, se tiene que

$$1 = \pi_0 \sum_{n=0}^{K} \left(\frac{\lambda}{\mu}\right)^n = \pi_0 \frac{1-\left(\frac{\lambda}{\mu}\right)^{K+1}}{1-\frac{\lambda}{\mu}},$$

Por tanto, si $\lambda \neq \mu$,

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \frac{1-\frac{\lambda}{\mu}}{1-\left(\frac{\lambda}{\mu}\right)^{K+1}}, \quad n = 0, 1, 2, \ldots, K.$$

desborda ni crece indefinidamente al rechazar a los clientes que llegan cuando está lleno (la cadena de Markov asociada es irreducible y finita y, por tanto, ergódica).

Si $\lambda = \mu$, la distribución de probabilidad del número de clientes en el sistema es uniforme

$$\pi_n = \frac{1}{K+1}, \quad n = 0, 1, 2, ..., K.$$

Si eliminásemos el truncamiento, es decir $K \to \infty$, cuando $\lambda < \mu$ las expresiones de π_n se convierten en las obtenidas para el modelo *M/M/1*.

Medidas de rendimiento

Comenzamos con el número medio de clientes en el sistema. Para $\lambda = \mu$,

$$L = E(N) = \sum_{n=0}^{K} n\pi_n = \frac{1}{K+1}\sum_{n=0}^{K} n = \frac{K(K+1)}{2(K+1)} = \frac{K}{2},$$

que ya esperábamos por ser uniforme.

$$= \frac{\lambda[1-(K+1)(\lambda/\mu)^K + K(\lambda/\mu)^{K+1}]}{(\mu-\lambda)[1-(\lambda/\mu)^{K+1}]},$$

que también puede expresarse como

$$\frac{u}{1-u} - \frac{(K+1)u^{K+1}}{1-u^{K+1}}$$

donde el primer sumando es la expresión de *L* del modelo *M/M/1*. Por tanto, el número esperado de clientes en el sistema *M/M/1/K* es siempre menor que en el *M/M/1*, haciéndolo más eficiente.

Como para todo λ y μ se tiene

$$\begin{aligned} L_s = E(N_s) &= E(N_s \mid N=0)P(N=0) + E(N_s \mid N>0)P(N>0) \\ &= 0\pi_0 + 1(1-\pi_0) = 1-\pi_0, \end{aligned}$$

entonces $L_q = L - L_s = L - (1 - \pi_0)$.

En este modelo se rechaza a los clientes que llegan cuando ya hay *K* en el sistema (*K*-1 en la cola), lo que ocurre con probabilidad π_K.

Luego, la probabilidad de que al llegar un cliente entre en el sistema es $1-\pi_K$, representando la proporción de tiempo que el sistema no está saturado o la proporción de clientes que llegan que realmente entran en el sistema.

Así, la tasa media de entradas al sistema o paso a través del sistema, λ_e= **λ**, se define como

$$\lambda_e = \lambda(1-\pi_K).$$

La utilización verdadera del servidor, *ρ*, probabilidad de que el servidor esté

Tiempos de espera

Entendiendo por clientes en el sistema aquellos que entran en el sistema, podemos aplicar las fórmulas de Little para conseguir los tiempos medios en el sistema y en la cola

$$W = L / \lambda_e \qquad W_q = L_q / \lambda_e$$

Para obtener el tiempo medio de espera en cola para aquellos clientes que deben esperar, hacemos como en el modelo *M/M/1*,

$$E(q \mid q > 0) = W_q/\rho = W_q/(1 - \pi_0).$$

El proceso de obtención de las distribuciones de los tiempos *q* y *w* es más complejo que en el modelo *M/M/1*. Como hicimos entonces, utilizaremos el teorema de la probabilidad total, pero ahora condicionando a la v.a. N_e, que cuenta el número de clientes en el sistema cuando entra un cliente en él.

Denotamos con $q_n = P(N_e = n)$, $n = 0,1,2,...,K-1$, la probabilidad de que un cliente que entra en el sistema encuentre *n* clientes en él, que por el teorema de Bayes es

$$q_n = \frac{\pi_n}{1 - \pi_K}, \qquad n = 0, 1, 2, ..., K - 1.$$

Nótese que en este modelo la entrada no es una verdadera Poisson, $p_n \neq q_n$, y $\lambda_n = \lambda$ para $n \leq K-1$ pero $\lambda_n = 0$ para $n \geq K$, a diferencia de lo que ocurría en el *M/M/1*. Así, para $t \geq 0$,

$$= 1 - \sum_{n=0}^{K-1} q_n \left[\sum_{i=0}^{n} \frac{(\mu t)^i e^{-\mu t}}{i!} \right] = 1 - \sum_{n=0}^{K-1} q_n F_{\mathcal{P}(\mu t)}(n),$$

donde $$F_{\mathcal{P}(\mu t)}(n) = \sum_{i=0}^{n} \frac{(\mu t)^i e^{-\mu t}}{i!} = \int_t^{\infty} \mu \frac{\mu^n x^n e^{-\mu x}}{n!} dx$$

(igualdad debida a la relación entre las distribuciones de Erlang y Poisson) es la función de distribución de Poisson de parámetro μt en el punto n, que puede obtenerse a partir de las tablas de dicha distribución.
De forma similar

$$
\begin{aligned}
F_q(t) &= P(q=0) + P(0 < q \leq t) = q_0 + \sum_{n=1} P(q \leq t \mid N_e = n) P(N_e = n) \\
&= q_0 + \sum_{n=1}^{K-1} \left[\int_0^t \mu e^{-\mu x} \frac{(\mu x)^{n-1}}{(n-1)!} dx \right] q_n = q_0 + \sum_{n=0}^{K-2} q_{n+1} \left[1 - \int_t^{\infty} \mu e^{-\mu x} \frac{(\mu x)^n}{n!} dx \right] \\
&= q_0 + \sum_{n=0}^{K-2} q_{n+1} - \sum_{n=0}^{K-2} q_{n+1} \left[\int_t^{\infty} \mu e^{-\mu x} \frac{(\mu x)^n}{n!} dx \right] = 1 - \sum_{n=0}^{K-2} q_{n+1} \left[\sum_{i=0}^{n} \frac{(\mu t)^i e^{-\mu t}}{i!} \right] \\
&= 1 - \sum_{n=0}^{K-2} q_{n+1} F_{\mathcal{P}(\mu t)}(n)
\end{aligned}
$$