ENRIQUETA VERCHER GONZÁLEZ

TEORÍA DE COLAS

UNIVERSITAT DE VALÈNCIA

Índice general

Capítulo 1

Concepto de proceso estocástico

TEMA I   CONCEPTO DE PROCESO ESTOCASTICO

Introducción

La Teoría de Procesos Estocásticos puede considerarse como la parte dinámica de la Teoría de la Probabilidad. Los modelos matemáticos de los sistemas que varían en el tiempo de modo aleatorio son conocidos como procesos estocásticos.

Veamos algunos ejemplos que nos servirán para ilustrar algunas de las ideas y problemas que se plantean en esta teoría.

Ejemplo: Camino Aleatorio simple.

Sea X_n una v.a. que denota la posición en el instante n de una partícula en movimiento, $n=0,1,2\ldots$ Inicialmente la partícula está en el origen $X_0=0$, en el instante siguiente $n=1$ da un salto de una posición, hacia arriba con probabilidad $\frac{1}{2}$ o hacia abajo con la misma probabilidad. En $n=2$ da un nuevo salto, otra vez de una posición, con la misma probabilidad de ir hacia arriba o hacia abajo siendo los saltos en los instantes $n=1,2$ independientes. En general,

$$X_n = X_{n-1} + Z_n \qquad (1)$$

donde Z_n es el salto en la n-ésima etapa, de modo que $\{Z_1, Z_2 \ldots\}$ son mutuamente independientes y equidistribuidas:

$$P(Z_n=1) = P(Z_n=-1) = \frac{1}{2} \qquad n=1,2\ldots$$

Dado que $X_0=0$, tenemos que (1) se convierte en:

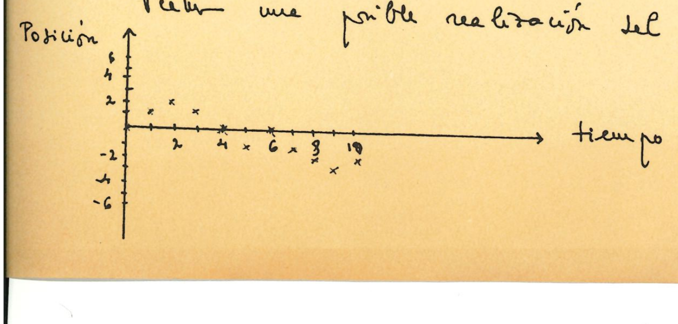
$$X_n = Z_1 + Z_2 + \ldots + Z_n \qquad (2)$$

lo que nos recuerda la distribución binomial. Para n intentos la probabilidad de que haya habido j saltos hacia abajo, y por consiguiente $n-j$ hacia arriba será: $\binom{n}{j}\frac{1}{2^n}$ Luego,

$$P(X_n = n-2j) = \binom{n}{j}\frac{1}{2^n}$$

Para n par, las posibles posiciones corresponden a los números $\{-n, \ldots, -2, 0, 2 \ldots n\}$ mientras que para los n impares, las posibles posiciones son $\{-n, \ldots, -1, 1, \ldots n\}$. Para un solo valor de n podríamos plantearnos el problema como uno típico del Cálculo de Probabilidades, ahora bien si consideramos todo el sistema como un proceso estocástico, la posición de la partícula es una función aleatoria de la variable temporal discreta n: $X_n \equiv X(n)$

Veamos una posible realización del proceso:



1. Definición de proceso estocástico

Definición

Un proceso estocástico es una familia de variables aleatorias $\{X(t),\ t\in T\}$, cada una de las cuales está definida sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$ donde T es un subconjunto de $\mathbb{R}$.

Al conjunto T lo llamaremos conjunto de índices o parámetros. En la mayor parte de las aplicaciones t será el parámetro tiempo y T el intervalo de tiempo en que se estudia el proceso. Según que el conjunto T sea discreto o continuo, tendremos respectivamente un proceso estocástico de parámetro discreto o continuo.

Un proceso estocástico es, en realidad, una función real de dos variables $\{X(t,\omega) : t\in T,\ \omega\in\Omega\}$. Para un suceso elemental $\omega_0\in\Omega$, $X(t,\omega_0)$ es una posible realización del proceso en función del tiempo. Para $t_0\in T$, $X(t_0,\omega)$ es una variable aleatoria. Finalmente $X(t_0,\omega_0)$ es una realización posible de esta variable aleatoria. En general, escribiremos $\{X(t),\ t\in T\}$ para designar un proceso estocástico, sin aludir explícitamente a su dependencia de ω.

Un proceso estocástico está estadísticamente determinado cuando se especifica la ley de probabilidad conjunta de las n variables aleatorias $X(t_1)\ X(t_2) - X(t_n)$ para todo n y $(t_1 - t_n)\in T^n$ con $t_i\neq t_j$, mediante:

a) la función de distribución conjunta

$$F_{X(t_1) - X(t_n)}(x_1, x_2 - x_n) = P(X(t_1)\leq x_1, \text{—} X(t_n)\leq x_n)$$

b) la función característica conjunta

$$\varphi_{X(t_1) - X(t_n)}(u_1, u_2 - u_n) = E\{e^{i(u_1 X(t_1) + - + u_n X(t_n))}\}$$

Otro procedimiento para describir procesos estocásticos es expresar $X(t)$ como una función de un proceso estocástico conocido. Por ejemplo, para T discreto sea el proceso $\{e(t),\ t\in T\}$, una sucesión de v.a. $N(0,1)$ independientes, podemos introducir el proceso $\{X(t),\ t\in T\}$ del siguiente modo:

$$X(t) = e(t) + c_1\, e(t-1) + c_2\, e(t-2) + \text{—} + c_n\, e(t-n) \qquad t\in T \qquad (\text{MA de orden } n)$$

En cualquier caso, para un estudio profundo de un proceso estocástico es conveniente conocer las siguientes funciones:

1. <u>Esperanza</u> o valor medio:

$$\mu(t) = E[X(t)] \qquad t\in T$$

2. <u>Autocovarianza</u> o momento mixto central:

$$\mathrm{cov}(t,s) = E\{[X(t)-\mu(t)][X(s)-\mu(s)]\} = r_X(t,s) \qquad (t,s)\in T^2$$

3. Momento mixto de 2º orden con respecto al origen:

$R(t,s) = E\{X(t)\,X(s)\}$ $\qquad (t,s)\in T^2$

Es inmediato que: $\operatorname{cov}(t,s) = R(t,s) - \mu(t)\mu(s)$

$\operatorname{Var}[X(t)] = \operatorname{cov}(t,t)$

Definición

Un proceso estocástico $\{X(t),\ t\in T\}$ se dice que es de segundo orden si $E[X^2(t)]<+\infty$ para todo $t\in T$.

Para estos procesos sabemos que existe la función media y la autocovarianza.

Teorema

Sea $\{X(t),\ t\in T\}$ un proceso estocástico de 2º orden, cuya función de autocovarianza es $\Gamma_x(s,t)$, $s\in T$ y $t\in T$. Se cumple:

(i) $\Gamma_x(s,t) = \Gamma_x(t,s)$

(ii) la forma cuadrática en $z_i\in\mathbb{R}$: $\sum_{i,j=1}^{n} z_i z_j \Gamma_x(t_i,t_j)$ para todo $n\in\mathbb{Z}$ y $t_i\in T$, $i=1\ldots n$ es definida no negativa.

(iii) $|\Gamma_x(s,t)|^2 \le \Gamma_x(s,s)\,\Gamma_x(t,t)$

Demostración

(i) Es inmediato de la definición: $\Gamma_x(s,t) = \operatorname{cov}(s,t) = R(t,s) - \mu(t)\mu(s) = \Gamma_x(t,s)$

(ii) Supongamos que $\mu(t)=0$, para todo $t\in T$, entonces:

$$\left[\sum_{i=1}^{n} z_i X(t_i)\right]^2 = \sum_{i=1}^{n} z_i^2 X^2(t_i) + 2\sum_{\substack{i,j=1\\ j>i}}^{n} z_i z_j X(t_i)X(t_j) = \sum_{i,j=1}^{n} z_i z_j X(t_i)X(t_j) \qquad z_i\in\mathbb{R},\ i=1\ldots n$$

luego

$$0 \le E\left[\left(\sum_{i=1}^{n} z_i X(t_i)\right)^2\right] = \sum_{i,j=1}^{n} z_i z_j E[X(t_i)X(t_j)] = \sum_{i,j=1}^{n} z_i z_j \Gamma_x(t_i,t_j)$$

En cualquier caso, y dado que:

$$\operatorname{Var}\left[\sum_{i=1}^{n} z_i X(t_i)\right] \underset{(*)}{=} \sum_{i,j=1}^{n} z_i z_j \Gamma_x(t_i,t_j) \ge 0$$

se sigue que la función de autocovarianza es definida no-negativa.

(iii) Recordemos la desigualdad de Schwarz para variables aleatorias con momentos de 2º orden finito: $|E(XY)|^2 \le E(X^2)E(Y^2)$. Si aplicamos esta desigualdad a las variables aleatorias: $Y = X(t)-\mu(t)$ y $S = X(s)-\mu(s)$ tendremos que:

$|\operatorname{cov}(t,s)|^2 \le \operatorname{Var}[X(t)]\operatorname{Var}[X(s)]$, es decir: $|\Gamma_x(s,t)|^2 \le \Gamma_x(t,t)\,\Gamma_x(s,s)$

Ejemplo. Consideremos el proceso estocástico de parámetro continuo $\{X(t),\ t\ge 0\}$ definido por:

$$X(t) = A\cos\omega t + B\operatorname{sen}\omega t$$

siendo ω una constante positiva conocida y A, B v.a. independientes $N(0,1)$.

Tenemos, entonces que:

— $X(t)$ se distribuye normal, por ser combinación lineal de v.a. normales, independ.

— $\mu(t) = E(A)\cos\omega t + E(B)\,\text{sen}\,\omega t = 0$

— $cov(t,s) = R(t,s) = E(A^2)\cos\omega t\cos\omega s + \overset{=0}{\cancel{E(AB)}}\{\cos\omega t\,\text{sen}\,\omega s + \text{sen}\,\omega t\cos\omega s\} +$
$+ E(B^2)\,\text{sen}\,\omega t\,\text{sen}\,\omega s = \cos\omega t\cos\omega s + \text{sen}\,\omega t\,\text{sen}\,\omega s = \cos\omega(t-s)$

De ahí que: $Var[X(t)] = \cos\omega(0) = 1$

En consecuencia: $X(t) \overset{D}{=} N(0,1)$.

__Definición__

Un proceso $\{X(t), t\in T\}$ diremos que tiene __incrementos independientes__ si para todo $t_1 \dots t_n \in T$ y $\forall n$, con $t_1 < t_2 < \dots < t_n$, las variables aleatorias $X(t_2)-X(t_1)$, $X(t_3)-X(t_2)$, ..., $X(t_n)-X(t_{n-1})$ son independientes.

Diremos que el proceso tiene __incrementos independientes estacionarios__ si además, con $t_1<t_2$, y para $h>0$, la v.a $X(t_2+h)-X(t_1+h)$ tiene la misma distribución $X(t_2)-X(t_1)$. Es decir, la distribución de probabilidad del incremento depende solamente de la diferencia t_2-t_1.

2. EL PROCESO DE WIENER

El proceso de Wiener es un proceso normal que ha sido muy importante para el desarrollo de la teoría de procesos estocásticos. Entre otras aplicaciones, el proceso de Wiener proporciona un modelo para el "movimiento browniano" o movimiento de una partícula sumergida en un fluido.

__Definición__

Diremos que el proceso estocástico $\{W(t), t\geq 0\}$ es un proceso de Wiener si:

1.- el proceso tiene incrementos independientes estacionarios

2.- $\forall t>0$, $W(t)$ sigue una distribución normal

3.- $\forall t>0$, $E[W(t)]=0$

4.- $W(0)=0$

Puesto que se trata de un proceso normal puede ser caracterizado mediante la función media y la función de autocovarianza.

Ecuación funcional de Cauchy (1821)

__Teorema__

Sea $\{W(t), t\geq 0\}$ un proceso estocástico con incrementos independientes estacionarios con $W(0)=0$ y $E[W(t)]=0$, $\forall t\geq 0$. Entonces, existe una constante positiva α tal que $Var[W(t)] = \alpha t$

Demostración

Sean s,t tales que $0\leq s\leq t$, tenemos:

$$0 \leq Var[W(t)-W(s)] = Var[W(t)] + Var[W(s)] - 2cov(t,s) = Var[W(t)] + Var[W(s)] - 2E[W(t)W(s)]$$

Ecuación funcional de Cauchy (1821)

Sea $f: \mathbb{R} \longrightarrow \mathbb{R}$ una función que satisface

$$f(x+y) = f(x) + f(y) \qquad [1]$$

se trata de determinar la expresión de $f(x)$.

Por inducción finita, a partir de [1], se sigue que:

$$f(x_1 + \dots + x_n) = f(x_1) + \dots + f(x_n) \qquad [2]$$

si llamamos $x_k = x \quad k = 1 \dots n$: $\quad f(nx) = n\, f(x) \qquad [3]$

Así, pues, si $x = \frac{m}{n} t$ $\forall t \in \mathbb{R}$ resulta que: $nx = mt$ y

$$f(nx) = f(mt) = n\, f(x) = m\, f(t)$$

luego $\quad f\left(\frac{m}{n} t\right) = \frac{m}{n} f(t) \qquad [4]$

En consecuencia, si $t = 1$: $f(1) = c$ y se tendrá para todo racional positivo x:

$$f(x) = f(x.1) = x\, f(1) = cx \qquad [5]$$

Para $x = 0$: $f(0) = 0$ (que puede deducirse directamente de [1]). Así, pues, [4] y [5] serán también válidas para $\frac{m}{n} = 0$ y $x = 0$, respectivamente.

Para x negativo, haciendo $y = -x$, se deduce de [1] que:

$$f(-x) = f(0) - f(x) = -f(x) \longrightarrow f(+x) = -f(-x) \qquad [6]$$

de [4] y [6] se llega a:

$$f(rt) = r\, f(t) \qquad \forall t \in \mathbb{R} \text{ y } \forall \text{ racional } r \qquad [7]$$

y si hacemos $t = 1$: $\quad f(r) = cr \quad \forall$ racional r.

Si asumimos que $f(x)$ es no-negativa (no-positiva) para valores positivos y suficientemente pequeños de x, se obtiene [5] como solución general de la ecuación de Cauchy (Darboux, 1880).

Para $y > 0$ suficientemente pequeño, si $f(y) \geq 0$ y de [1] se sigue que:

$$f(x+y) = f(x) + f(y) \geq f(x)$$

por lo que $f(\cdot)$ es monótona creciente (no necesariamente de forma estricta). Además, se ha visto [7] que para $x = r$ racional $f(r) = cr$.

Sea $\{r_n\}$ una sucesión creciente y $\{R_n\}$ una sucesión decreciente de racionales tales que convergen a x. Entonces, para todo n : $r_n < x < R_n$ y $c r_n = f(r_n) \leq f(x) \leq f(R_n) = c R_n$

y tomando límites $\quad f(x) = cx$.

Por ser el proceso de incrementos independientes, las v.a. $w(t)-w(s)$ y $w(s)-w(0)$ son independientes y:

$$E[w(t)w(s)] = E\{w(s)[w(t)-w(s)+w(s)]\} = E\{w(s)[w(t)-w(s)]\} + \mathrm{Var}[w(s)] =$$

$$= \overset{0}{E[w(s)]}\, E[w(t)-w(s)] + \mathrm{Var}[w(s)] = \mathrm{Var}[w(s)] \qquad (4)$$

luego: $\mathrm{Var}[w(t)] - \mathrm{Var}[w(s)] \geq 0$.

Si, para simplificar, introducimos la notación $f(t) = \mathrm{Var}[w(t)]$ resulta que $f(t) - f(s) \geq 0$, $t \geq s \iff f(\cdot)$ es una función no decreciente.

Como, además, los incrementos son estacionarios, $w(t)-w(s)$ y $w(t-s)-w(0) = w(t-s)$ están equidistribuidas y, sustituyendo en (3) tenemos que:

$$f(t-s) = f(t) - f(s) \qquad (5)$$

En consecuencia, para λ y μ arbitrarios: $\lambda>0$, $\mu>0$

$$\underbrace{f(t+\lambda+\mu) - f(t)}_{f(\lambda+\mu)} = \underbrace{[f(t+\lambda+\mu) - f(t+\lambda)]}_{f(\mu)} + \underbrace{[f(t+\lambda) - f(t)]}_{f(\lambda)}$$

Luego, $f(\cdot)$ es una función lineal no decreciente sobre la semirrecta $[0,+\infty[$, y $f(t) = \alpha t$ con $\alpha>0$ (si $\alpha=0$, tenemos un proceso trivial). ◆

Nota. La fórmula de la autocovarianza se sigue de (4):

$$\mathrm{cov}(t,s) = E[w(t)w(s)] = \alpha \min\{t,s\}$$

Comentario

El movimiento de una partícula sumergida en un líquido o gas se denomina movimiento browniano en honor del botánico inglés R. Brown quien observó en 1857 que estas partículas tienen un movimiento irregular. La explicación de este fenómeno fue uno de los mayores éxitos de la mecánica estadística y la teoría cinética. En 1905, Einstein demostró que se podía explicar este movimiento suponiendo que las partículas están sujetas al bombardeo continuo de las moléculas del medio que las rodea, y desarrolló un modelo matemático del fenómeno.

Un modo de explicar el movimiento browniano sería el siguiente: Consideremos una partícula específica sumergida en un fluido, representando por $w(t)$ el desplazamiento, en una determinada dirección (proyección), después de transcurrido un tiempo t, y suponiendo que el origen de la escala coincide con la posición inicial $w(0)=0$, $w(t)$ será una variable aleatoria, cuyas características son las siguientes:

a) el desplazamiento de la partícula durante el intervalo de tiempo $[0,t]$, grande comparado con el tiempo que transcurre entre dos impactos sucesivos, se puede considerar como la suma de un gran número de desplazamientos pequeños.

consecuencia, y según el Teorema Central del Límite, parece razonable reconocer que la distribución de $w(t)$ sea normal.

(b) También es razonable suponer que la distribución de probabilidad del desplazamiento $w(t)-w(s)$ sería la misma que la del $w(t+h)-w(s+h)$, $\forall h>0$, ya que se supone que el medio está en equilibrio. $E[w(t)]=0$ La ley de probabilidad del desplazamiento de la partícula a lo largo de un intervalo sólo dependerá de la amplitud del mismo, y no del instante en que comienza la observación.

(c) Finalmente, podemos afirmar que el proceso $w(t)$ tiene incrementos independientes, ya que el número y magnitud de los impactos sufridos por la partícula en intervalos no solapados son independientes.

Lo que justifica los axiomas del proceso de Wiener, llamado así en honor de N. Wiener que fue uno de los primeros en estudiar el movimiento browniano (1923, 1930). Este proceso se conoce también con el nombre de Wiener-Lévy.

3. OBTENCION DEL PROCESO DE WIENER COMO LIMITE DE UN CAMINO ALEATORIO

Considérese el modelo siguiente para el movimiento de una partícula sobre una línea, determinado por los numerosos impactos con las moléculas del fluido. En el instante inicial la partícula se encuentra en el origen : $X(0)=0$. Supóngase que entre cada par de impactos sucesivos transcurre un tiempo T, y que el efecto del impacto supone un desplazamiento de s ó $-s$, con idéntica probabilidad. La posición respecto del origen, $X(t)$, es una v.a., y tendríamos un proceso estocástico $\{X(t),\ t\geq 0\}$ con discontinuidades en los puntos $nT=t$ $n\in\mathbb{N}$ (camino aleatorio).

Supongamos que en los n primeros impactos se han dado k avances (dirección positiva de $X(t)$), y $n-k$ retrocesos, tendremos que:

$$X(nT) = ks-(n-k)s = (2k-n)s$$

$$P[X(nT)=(2k-n)s] = P[k \text{ avances}] = \binom{n}{k}\frac{1}{2^n}$$

Lema

Para el camino aleatorio anterior se cumple que: $E[X(nT)]=0$ y $\mathrm{Var}[X(nT)]=ns^2$ (6)

Demostración

Sea Z_i una v.a. que toma los valores s y $-s$ con probabilidad iguales, según que el iésimo sea dado hacia arriba o hacia abajo. Entonces

$$E(Z_i)=0 \qquad E(Z_i^2)=\mathrm{Var}[Z_i]=s^2$$

Ahora bien, $X(nT)=Z_1+\dots+Z_n$, siendo las v.a Z_i independientes, luego

$$E[X(nT)]=0 \quad \text{y} \quad E[X(nT)^2]=\mathrm{Var}[X(nT)]=\sum_{i=1}^{n}\mathrm{Var}\,Z_i = ns^2. \qquad \blacklozenge$$

Nota. Por otra parte, por el teorema de Moivre, para n grande, tenemos que:

$$P[k \text{ avances}] \cong \frac{1}{\sqrt{2\pi n p q}}\, e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{npq}}\right)^2} dk$$

si definimos $r = 2k-n$, tenemos que $k=\frac{n+r}{2}$ y $dk=\frac{dr}{2}$, luego $p=q=\frac{1}{2}$

$$P[k \text{ avances}] = P[X(nT)=rs] \cong \frac{1}{2}\,\frac{1}{\sqrt{\frac{n\pi}{2}}}\, e^{-\frac{r^2}{2n}}\, dr \qquad (7)$$

y haciendo $t=nT$, por (6) resulta que $E[X(t)]=0$ y $Var[X(t)] = t\frac{s^2}{T}$.

Supongamos, ahora, que t permanece constante, pero s y T tienden a cero (desplazamiento infinitesimal e impactos casi continuos). En este caso la varianza de $X(t)$ permanecerá finita y distinta de cero solamente si s y $\sqrt{T}$ son infinitésimos del mismo orden, i.e. si $s^2=\alpha T$, $\alpha>0$, en cuyo caso $E[X(t)^2]=\alpha t$.

En estas condiciones definimos el proceso estocástico: $W(t) = \lim_{T\to 0} X(t)$ (8)

Teorema

El proceso estocástico $\{W(t),\ t\geq 0\}$, con $t=nT$ definido por (8) es el proceso de Wiener.

Demostración

Tenemos que probar los cuatro axiomas dados por la definición. El axioma 3, $E[W(t)]=0$ $\forall t\geq 0$, es evidente por ser $E[X(t)]=0$. Además, se cumple $Var[W(t)]=\alpha t$. También es inmediato que $W(0)=0$ (axioma 4), por ser $X(0)=0$, cualquiera que sea T.

Veamos que $W(t)$ se distribuye normalmente (axioma 2):

$$P[X(nT)=rs] \cong \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sqrt{n}}\, e^{-\frac{r^2}{2n}}\, dr$$

en consecuencia,

$$P[X(t)=rs] \cong \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sqrt{n}}\, e^{-\frac{(rs)^2}{2ns^2}}\,\frac{1}{s}\, d(rs)$$

si $sr=x \longrightarrow P[X(\underset{nT}{t})=x] \cong \frac{1}{\sqrt{2\pi}}\,\frac{1}{s}\,\frac{1}{\sqrt{n}}\, e^{-\frac{x^2}{2ns^2}}\, dx$

Si $\left.\begin{array}{l} nT=t \quad n=\frac{t}{T} \\ n\to+\infty,\ T\to 0 \end{array}\right\} \longrightarrow P[X(t)\leq x] = \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sqrt{t}\sqrt{\frac{s^2}{T}}}\int_{-\infty}^{x} e^{-\frac{z^2}{2t\, s^2/T}}\, dz$

luego

$$P[X(t)\leq x] = \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sqrt{\alpha t}}\int_{-\infty}^{x} e^{-\frac{z^2}{2\alpha t}}\, dz \longrightarrow W(t) \overset{D}{=} N(0,\alpha t)$$

Sean $t_1<t_2<t_3$, los incrementos $W(t_3)-W(t_2)$ y $W(t_2)-W(t_1)$ son independientes, pues los impactos en dos intervalos que no se solapan son independientes. En este caso, y según se vió en un teorema anterior (pág 5): $R(t,s)=cov(t,s)=\alpha\min\{t,s\}$. Falta probar que los incrementos son estacionarios (axioma 1).

Desde luego $W(t_2)-W(t_1)$ sigue una distribución normal con media cero y varianza:

$$(1\ \ -1)\begin{Bmatrix} \alpha t_2 & \alpha t_1 \\ \alpha t_1 & \alpha t_1 \end{Bmatrix}\begin{pmatrix}1\\-1\end{pmatrix} = (\alpha(t_2-t_1)\quad 0)\begin{pmatrix}1\\-1\end{pmatrix} = \alpha(t_2-t_1)$$

matriz de covarianzas del vector $[W(t_2), W(t_1)]$

Luego $W(t_2) - W(t_1) \stackrel{d}{=} N(0, \alpha(t_2 - t_1))$, con lo que la distribución del incremento sólo depende de la diferencia $(t_2 - t_1)$. ◆

<u>Comentario</u>

Un proceso estocástico que aparece también relacionado con el proceso de Wiener es el que se conoce con el nombre de "ruido blanco" o proceso aleatorio puro. Se trata de un proceso estocástico que tiene varianza infinita y que se obtiene como derivada (en media cuadrática) del proceso de Wiener.

El ruido blanco sirve como aproximación a ciertos procesos que se presentan en física y en determinadas ramas de ingeniería, además sirve para definir determinadas ecuaciones estocásticas diferenciales.

Capítulo 2

Procesos de Poisson

TEMA II : PROCESOS DE POISSON

1. DEDUCCION AXIOMATICA

Definición

Un proceso estocástico $\{X(t),\ t\geq 0\}$, de valores enteros no negativos, es un proceso de Poisson con un número medio de ocurrencia λ si, y sólo si,

(i) el proceso tiene incrementos independientes y estacionarios

(ii) el número de sucesos ocurridos en el intervalo de tiempo $]s,t]$, i.e. $X(t)-X(s)$ tiene una distribución de Poisson de parámetro $\lambda(t-s)$:

$$P\{X(t)-X(s)=k\} = \frac{[\lambda(t-s)]^k}{k!}e^{-\lambda(t-s)} \qquad k=0,1,2\ldots$$

Nota. La propia definición pone de manifiesto que los incrementos son estacionarios, pues la distribución de $X(t)-X(s)$ depende exclusivamente de la diferencia $t-s$.

Definición

Un proceso de contaje es un proceso, con valores enteros no negativos, $\{X(t),\ t\geq 0\}$ que cuenta el número de sucesos (llegadas de clientes a una cola, desintegraciones radiactivas) ocurridos en el intervalo $]0,t]$, $t>0$.

Teorema

Si un proceso de contaje $\{X(t),\ t\geq 0\}$ satisface los siguientes axiomas:

A1.- $X(0)=0$

A2.- los incrementos son estacionarios e independientes

A3.- existe un $\lambda>0$ tal que para h positivo haciéndose próximo a cero:

$$P(X(t+h)-X(t)=1) = \lambda h + o(h)$$

donde $o(h)$ es un infinitésimo de orden mayor que h. Además:

$$P(X(t+h)-X(t)>1) = o(h).$$

Entonces, el proceso es de Poisson con un número medio de ocurrencia λ.

Demostración

Sólo necesitamos demostrar que para $s<t$: $X(t)-X(s) \stackrel{D}{=} P_o[\lambda(t-s)]$, o lo que es lo mismo:

$$P_m(t-s) \stackrel{(*)}{=} P(X(t)-X(s)=m) = \frac{[\lambda(t-s)]^m}{m!}e^{-\lambda(t-s)} \qquad m=0,1,2\ldots$$

Para $h\to 0$, A3 nos permite escribir:

$$\sum_{m=2}^{\infty} p_m(h) = o(h)$$

Consideremos la probabilidad de que en un intervalo cualquiera de longitud h ocurra al menos un suceso:

$$p(h) = \sum_{m=1}^{\infty} p_m(h) = p_1(h) + \sum_{m=2}^{\infty} p_m(h) = [\lambda h + o(h)] + o(h) = \lambda h + o(h) \quad \text{cuando } h\to 0$$

Entonces, para todo t, y dado que los incrementos son independientes:

$$p_0(t+h) = p_0(t)\, p_0(h) = p_0(t)\,[1-p(h)] \longrightarrow \frac{p_0(t+h)-p_0(t)}{h} = -p_0(t)\,\frac{p(h)}{h}$$

y para $h \to 0$ resulta: $p_0'(t) = -\lambda\, p_0(t)$ (1)

La ecuación diferencial (1) tiene como solución: $p_0(t) = C\,e^{-\lambda t}$, y teniendo en cuenta la condición inicial $p_0(0) = 1 \to C = 1$, resulta que: $p_0(t) = e^{-\lambda t}$ (2)

Vamos a calcular $p_m(t)$, para $m = 1, 2, \ldots$. Se tiene que

$$p_m(t+h) = p_m(t)\,p_0(h) + p_{m-1}(t)\,p_1(h) + \sum_{i=2}^{m} p_{m-i}(t)\,p_i(h)$$

siendo $0 \le \sum_{i=2}^{m} \underbrace{p_{m-i}(t)}_{\le 1}\, p_i(h) \le \sum_{i=2}^{m} p_i(h) = o(h)$ cuando $h \downarrow 0$

luego
$$p_m(t+h) - p_m(t) = p_m(t)\,[\,p_0(h) - 1\,] + p_{m-1}(t)\,p_1(h) + \sum_{i=2}^{m} p_{m-i}(t)\,p_i(h) =$$
$$= -p_m(t)\,p(h) + p_{m-1}(t)\,p_1(h) + o(h) =$$
$$= -\lambda\, p_m(t)\,h + \lambda\, p_{m-1}(t)\,h + o(h) \qquad h \downarrow 0$$

y
$$\frac{p_m(t-h) - p_m(t)}{h} = -\lambda\, p_m(t) + \lambda\, p_{m-1}(t) + \frac{o(h)}{h}$$

y al tomar límites: $p_m'(t) = -\lambda\, p_m(t) + \lambda\, p_{m-1}(t) \qquad m = 1, 2, \ldots$ (3)

Hacemos la siguiente hipótesis de inducción: $p_{m-1}(t) = \frac{(\lambda t)^{m-1}}{(m-1)!}\, e^{-\lambda t}$, válida para $m = 1$, según se ha visto en (2). Por otra parte, (3) es equivalente a:

$$p_m'(t) + \lambda\, p_m(t) = \frac{\lambda^m\, t^{m-1}}{(m-1)!}\, e^{-\lambda t} \qquad (3)$$

Para resolver la ecuación diferencial (3) utilizaremos el factor integrante $e^{\lambda t}$

$$p_m'(t)\, e^{\lambda t} + \lambda\, e^{\lambda t}\, p_m(t) = \frac{d}{dt}\left[\, p_m(t)\, e^{\lambda t} \right] = \frac{\lambda^m\, t^{m-1}}{(m-1)!}$$

de donde: $p_m(t)\, e^{\lambda t} = \frac{\lambda^m\, t^m}{m!} + C \longrightarrow p_m(t) = \frac{(\lambda t)^m}{m!}\, e^{-\lambda t} + C\, e^{-\lambda t}$

Si $t = 0$, $p_m(0) = 0 \longrightarrow C = 0$, luego: $m = 0, 1, 2, \ldots$

$$p_m(t) = \frac{(\lambda t)^m}{m!}\, e^{-\lambda t} \longleftrightarrow X(t) \sim P_o(\lambda t)$$

Por ser los incrementos estacionarios: $X(t) - X(s) = X(t-s) - X(0) = X(t-s) \sim P_o[\lambda(t-s)]$

Nota. Para el proceso de Poisson vamos a obtener las funciones de autocovarianza y autocorrelación. Sabemos que:

$$E[X(t_1)] = \lambda t_1 \quad \text{y} \quad E[X(t_2) - X(t_1)] = \lambda(t_2 - t_1) \qquad 0 < t_1 < t_2$$

$$E\{[X(t_1)]^2\} = E^2[X(t_1)] + Var[X(t_1)] = \lambda^2 t_1^2 + \lambda t_1$$

Las variables aleatorias $X(t_2)-X(t_1)$ y $X(t_1)-X(0)=X(t_1)$ son independientes, para $0<t_1<t_2$, luego:

$$E\{[X(t_2)-X(t_1)]\,X(t_1)\} = E\{X(t_2)-X(t_1)\}\,E\{X(t_1)\} = \lambda(t_2-t_1)\lambda t_1$$

$$= E\{X(t_1)X(t_2)\} - E\{[X(t_1)]^2\} = E\{X(t_1)X(t_2)\} - \lambda^2 t_1^2 - \lambda t_1$$

en consecuencia: $E\{X(t_1)X(t_2)\} = \lambda^2 t_1 t_2 - \lambda^2 t_1^2 + \lambda^2 t_1^2 + \lambda t_1 = \lambda t_1(\lambda t_2 + 1)$. Luego

$$\text{cov}(t_1,t_2) = E\{X(t_1)X(t_2)\} - \lambda^2 t_1 t_2 = \lambda t_1 \equiv \lambda \min\{t_1, t_2\} \qquad (4)$$

El coeficiente de correlación entre $X(t_1)$ y $X(t_2)$ es:

$$\rho = \frac{\lambda t_1}{\sqrt{\lambda t_1}\sqrt{\lambda t_2}} = \sqrt{\frac{t_1}{t_2}}$$

obsérvese que si $t_1=t_2$, entonces $\rho=1$; mientras que para $t_2 \ggg t_1$ el valor de ρ se hace muy pequeño.

2. TIEMPOS ENTRE LLEGADAS Y TIEMPO DE ESPERA

Definición

Se llama tiempo de espera hasta el suceso n-ésimo a la v.a. W_n que representa el tiempo transcurrido hasta que se registran n sucesos.

Los tiempos entre llegadas sucesivas se definen de la siguiente forma: T_1 es el tiempo transcurrido hasta que ocurre el primer suceso; T_j, $j>1$, es el tiempo que transcurre desde el $(j-1)$-ésimo suceso hasta el j-ésimo.

En función de los tiempos de espera $\{W_j\}$, los tiempos entre llegadas $\{T_j\}$ viene dado por:

$$T_1 = W_1, \quad T_2 = W_2 - W_1 \; \text{—} \; T_n = W_n - W_{n-1} \longrightarrow W_n = T_1 + T_2 + \cdots + T_n, \quad n \geq 1$$

Existe una relación básica entre el proceso de conteo $\{X(t),\ t\geq 0\}$ y los tiempos de espera:

$$\left.\begin{array}{ll} X(t) \leq n & \text{sii } W_{n+1} > t \\ \text{y} \quad X(t) = n & \text{sii } W_n \leq t \ \text{y} \ W_{n+1} > t \end{array}\right\} \qquad (5)$$

De (5) pueden deducirse importantes relaciones probabilísticas:

$$P\{X(t)\leq n\} = P\{W_{n+1} > t\}, \quad n=0,1,2\ldots \qquad (6)$$

y $P\{X(t)=n\} = P\{(W_n\leq t)(W_{n+1}>t)\} \underset{(*)}{=} P\{W_n \leq t\} - P\{(W_n\leq t)(W_{n+1}\leq t)\}$, es decir

$$P\{X(t)=n\} = P\{W_n\leq t\} - P\{W_{n+1}\leq t\} \qquad n=1,2\ldots \qquad (7)$$

Finalmente tendríamos, como condición límite:

$$P\{X(t)=0\} = 1 - P\{W_1 \leq t\} \qquad t>0 \qquad (8)$$

Se pueden establecer estas relaciones mediante la función de distribución de los tiempos de espera:

$$\left.\begin{array}{ll} F_{X(t)}(n) = 1 - F_{W_{n+1}}(t) & n = 0, 1, 2 \ldots \\ P\{X(t) = n\} = F_{W_n}(t) - F_{W_{n+1}}(t) & n = 1, 2 \ldots \\ P\{X(t) = 0\} = 1 - F_{W_1}(t) & \end{array}\right\} \quad (1)$$

Teorema

Si el proceso de conteo $\{X(t), t \geq 0\}$ es un proceso de Poisson con número medio de ocurrencia λ, entonces: $W_n \sim \Gamma(n, \lambda)$.

Demostración

Hay que probar que la función de densidad de W_n es de la forma:

$$f_{W_n}(t) = \frac{\lambda^n e^{-\lambda t} t^{n-1}}{(n-1)!} \qquad t > 0$$

Ahora bien, por (1): $1 - F_{W_n}(t) = F_{X(t)}(n-1) = P\{X(t) \leq n-1\} = \sum_{k=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}$, luego

$$F_{W_n}(t) = \begin{cases} 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} & t > 0 \\ 0 & t \leq 0 \end{cases}$$

y derivando con respecto a t, para $t > 0$, se obtiene

$$f_{W_n}(t) = \lambda e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} - e^{-\lambda t} \sum_{k=1}^{n-1} k \frac{(\lambda t)^{k-1}}{k!} \lambda = \lambda e^{-\lambda t} \left\{ \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} - \sum_{k=0}^{n-2} \frac{(\lambda t)^k}{k!} \right\} =$$

$$= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \qquad \blacklozenge$$

Teorema

Los tiempos sucesivos entre llegadas $T_1, T_2 \ldots$ de un proceso de Poisson con intensidad λ son v.a. independientes y equidistribuidas que obedecen una ley exponencial con media $1/\lambda$.

Demostración

Hay que probar que para todo n y $t_1, t_2 \ldots t_n \in \mathbb{R}$:

$$F_{T_1 T_2 \ldots T_n}(t_1, \ldots t_n) = P\{T_1 \leq t_1, T_2 \leq t_2 \ldots T_n \leq t_n\} = (1 - e^{-\lambda t_1})(1 - e^{-\lambda t_2}) \ldots (1 - e^{-\lambda t_n}) = \prod_{i=1}^{n} F_{T_i}(t_i)$$

La relación anterior se cumple trivialmente para $n = 1$:

$$P(T_1 \leq t_1) = 1 - P(T_1 > t_1) = 1 - P\{X(t_1) = 0\} = 1 - e^{-\lambda t_1}$$

Para $n = 2$ podemos establecer:

$$P\{T_2 > y \,/\, T_1 \leq x\} = P\{X(T_1 + y) - X(T_1) = 0 \,/\, T_1 \leq x\} \underset{(*)}{=} P\{X(y) - X(0) \equiv X(y) = 0\} = e^{-\lambda y}$$

(*) como el proceso de Poisson tiene incrementos estacionarios e independientes

Por lo tanto,

$$P\{T_1 \le t_1, T_2 \le t_2\} = (1 - P\{T_2 > t_2 / T_1 \le t_1\})\, P(T_1 \le t_1) = (1-e^{-\lambda t_2})(1-e^{-\lambda t_1}) \qquad (10)$$

Siguiendo el proceso de inducción, validaríamos la expresión (10) para cualquier n.

<u>Distribución condicional de los tiempos de espera</u>

Supongamos que en el intervalo $[0,t]$ ha tenido lugar un único suceso, y nos interesa determinar la distribución del instante de tiempo en el que ocurrió tal suceso. Ahora bien, debido a la estacionariedad e independencia de los incrementos, parece razonable que los subintervalos en $[0,t]$ de igual amplitud tengan la misma probabilidad de contener al suceso. En otras palabras, la distribución de dicho instante de tiempo debería ser una distribución uniforme sobre $[0,t]$. Comprobemos que esta conjetura es cierta para $s \le t$:

$$P\{W_1 \le s / X(t)=1\} = \frac{P\{X(s)=1;\ X(t)-X(s)=0\}}{P\{X(t)=1\}} = \frac{P\{X(s)=1\}\, P\{X(t)-X(s)=0\}}{P\{X(t)=1\}} = \frac{e^{-\lambda s}\lambda s\, e^{-\lambda(t-s)}}{e^{-\lambda t}\lambda t} =$$

Este resultado puede ser generalizado, pero antes de hacerlo recordemos el concepto de estadístico de orden.

<u>Definición</u>

Sean $Y_1, Y_2, \dots Y_n$ v.a. Diremos que $[Y_{(1)}, Y_{(2)}, \dots Y_{(n)}]$ es el <u>estadístico de orden</u> correspondiente al vector aleatorio $[Y_1, Y_2, \dots Y_n]$, si $Y_{(k)}$ es el k-ésimo valor más pequeño entre asumidos por las variables $Y_1, \dots Y_n$.

Si dichas v.a. son independientes y absolutamente continuas, con idéntica distribución definida por la función de densidad $f(\cdot)$, se tendrá:

$$f_{[Y_{(1)} \dots Y_{(n)}]}(y_1, y_2 \dots y_n) = n!\, f(y_1) \dots f(y_n) \qquad \text{para } y_1 < y_2 \dots < y_n$$

(Cada una de las $n!$ permutaciones de $Y_1 \dots Y_n$ conduce al mismo valor del estadístico de

<u>Teorema</u>

Dados $X(t)=n$, los n tiempos de espera $W_1, W_2 \dots W_n$ se distribuyen conjuntamente de forma idéntica que los estadísticos de orden correspondientes a $\underline{n}$ v.a. independientes, equidistribuidas uniformemente en el intervalo $(0,t)$.

Demostración

Sean $0 < t_1 < t_2 < \dots < t_n < t_{n+1} = t$ y sea $h_i > 0$, suficientemente pequeño para que $t_i + h_i < t_{i+1}$. Entonces:

$$P\{t_i \le W_i \le t_i + h_i,\ i=1,2 \dots n / X(t) = n\} =$$

$$= \frac{P\{\text{exactamente un suceso en } [t_i, t_i+h_i]\ i=1,2\dots n;\ \text{ningún suceso en el resto de } [0,t]\}}{P\{X(t)=n\}} =$$

$$= \frac{(\lambda h_1 e^{-\lambda h_1})(\lambda h_2 e^{-\lambda h_2}) \cdots (\lambda h_n e^{-\lambda h_n})\, e^{-\lambda (t-h_1-h_2-\cdots-h_n)}}{e^{-\lambda t}\,\frac{(\lambda t)^n}{n!}} = \frac{n!}{t^n}\, h_1 h_2 \cdots h_n$$

luego $$\frac{P\{t_i \le w_i \le t_i+h_i,\ i=1\ldots n \,/\, X(t)=n\}}{h_1 h_2 \cdots h_n} = \frac{n!}{t^n}$$

y haciendo $h_i \to 0$, $i=1\ldots n$ se tendrá que:

$$f_{[w_1 \ldots w_n]}(t_1, t_2, \ldots t_n \,/\, X(t)=n) = \frac{n!}{t^n} \qquad \text{siendo } 0<t_1<t_2<\ldots<t_n<t$$ ◆

<u>Nota.</u> Intuitivamente suele decirse que bajo la condición de que n sucesos han ocurrido $(0,t)$, los tiempos de espera $\tilde{w}_1, \tilde{w}_2 \ldots \tilde{w}_n$ (en que los sucesos se presentan) considerados como v.a. no ordenada (i.e. $\tilde{w}_1$ representa el tiempo de ocurrencia de un suceso al azar no del primero) se distribuyen como variables independientes y uniformes en $(0,t)$.

<u>Ejemplo.</u> <u>Cola de Poisson con infinitos servidores</u> $M|G|\infty$

Supongamos que los clientes llegan a una estación de servicio de acuerdo con un proceso de Poisson de parámetro λ. Después de su llegada cada cliente es inmediatamente servido por uno de los infinitos servidores disponibles. Los tiempos de servicio se suponen independientes, con distribución común G.

Sea $X(t)$ el número de clientes en el sistema (atendidos) en el instante t. Determinaremos la distribución de $X(t)$ condicionándola a $N(t)$, número de usuarios llegados durante el intervalo de tiempo $[0,t]$:

$$P\{X(t)=j\} = \sum_{n=0}^{\infty} P\{X(t)=j \,/\, N(t)=n\}\, e^{-\lambda t}\frac{(\lambda t)^n}{n!} \qquad (11)$$

La probabilidad de que un usuario que llegó en el instante x esté todavía en la cola en el instante t es: $1-G(t-x)$.

Dado que $N(t)=n$, se sigue del teorema anterior que la probabilidad de que uno al azar de estos clientes esté todavía recibiendo servicio en t, viene dada por:

$$p = \int_0^t [1-G(t-x)]\,\frac{1}{t}\,dx = \int_0^t [1-G(x)]\,\frac{1}{t}\,dx$$

independientemente de los otros. De ahí se deduce que:

$$P\{X(t)=j \,/\, N(t)=n\} = \begin{cases} 0 & \text{si } j>n \\ \binom{n}{j} p^j (1-p)^{n-j} & j=0,1\ldots n \end{cases}$$

y, así de (11):

$$P\{X(t)=j\} = \sum_{n=j}^{\infty} \binom{n}{j} p^j (1-p)^{n-j} e^{-\lambda t}\frac{(\lambda t)^n}{n!} = e^{-\lambda t}\frac{(\lambda t p)^j}{j!}\sum_{n=j}^{\infty}\frac{[\lambda t(1-p)]^{n-j}}{(n-j)!} = e^{-\lambda t}\frac{(\lambda t p)^j}{j!}\, e^{\lambda t (1-p)}$$

$$= e^{-\lambda t p}\frac{(\lambda t p)^j}{j!}$$

En consecuencia, $X(t)$ tiene una distribución de Poisson con media $\lambda \int_0^t [1-G(x)]\,dx$

3. PROCESO COMPUESTO DE POISSON

Definición

Un proceso $\{Z(t),\ t \geq 0\}$ es un proceso compuesto de Poisson si se puede representar como

$$Z(t) = \sum_{n=1}^{X(t)} Y_n \qquad (12)$$

donde $\{X(t),\ t\geq 0\}$ es un proceso de Poisson con intensidad λ e $\{Y_n,\ n=1,2,\dots\}$ es una familia de v.a. independientes, equidistribuidas como una v.a Y, e independientes del proceso $X(t)$.

Ejemplo:

Supongamos que los poseedores de una póliza de una determinada compañía de seguros vida mueren en instantes $w_1 \leq w_2 \leq \dots \leq w_n \leq \dots$ Se supone que los fallecimientos son sucesos de Poisson con parámetro λ. El asegurado que muere en el instante i lleva una póliza por una cantidad Y_n, que se abonará a sus beneficiarios en momento de su fallecimiento. La compañía de seguros estará interesada en conocer $Z(t)$, que representa, de acuerdo con (12), la cantidad que deberá pagar en $(0,t)$, a fin de determinar la reserva que ha de tener disponible para hacer frente a sus compromisos.

Teorema

Un proceso de Poisson compuesto $\{Z(t),\ t\geq 0\}$ tiene incrementos estacionarios e independientes, y su función característica es:

$$\varphi_{Z(t)}(u) = e^{\lambda t\,[\varphi_Y(u)-1]} \qquad (13)$$

Demostración

Para $t_0 < t_1 < t_2$: $Z(t_1) - Z(t_0) = \sum_{n=X(t_0)+1}^{X(t_1)} Y_n$ y $Z(t_2) - Z(t_1) = \sum_{n=X(t_1)+1}^{X(t_2)} Y_n$, que son v.a. independientes, al ser las Y_n involucradas en ambos sumatorios.

Para demostrar que $\{Z(t),\ t\geq 0\}$ tiene incrementos estacionarios y que se cumple (13) teniendo en cuenta que $Z(0)=0$, bastará con probar que, para $t > s \geq 0$:

$$\varphi_{Z(t)-Z(s)}(u) = e^{\lambda(t-s)\,[\varphi_Y(u)-1]}$$

con lo que la distribución de $Z(t)-Z(s)$ depende sólo de $(t-s)$.

Tenemos que:

$$\varphi_{Z(t)-Z(s)}(u) = E\left\{e^{iu[Z(t)-Z(s)]}\right\} \overset{(*)}{=} \sum_{n=0}^{\infty} E\left\{e^{iu[Z(t)-Z(s)]} / X(t)-X(s)=n\right\} P\{X(t)-X(s)=$$

ahora bien, para $n=0,1,2\dots$

$$E\left\{e^{iu[Z(t)-Z(s)]} / X(t)-X(s)=n\right\} = \varphi_{\sum_{k=X(s)+1}^{X(t)} Y_k}\left(u / X(t)-X(s)=n\right) \underset{\text{indep y equidist}}{=} [\varphi_Y(u)]^n$$

de ahí que:

$$\varphi_{Z(t)-Z(s)}(u) = \sum_{n=0}^{\infty} [\varphi_Y(u)]^n \, P\{X(t)-X(s)=n\} = \sum_{n=0}^{\infty} [\varphi_Y(u)]^n \, e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^n}{n!} =$$

$$= e^{-\lambda(t-s)} \sum_{n=0}^{\infty} \frac{[\lambda(t-s)\varphi_Y(u)]^n}{n!} = e^{-\lambda(t-s)} e^{\lambda(t-s)\varphi_Y(u)} = e^{\lambda(t-s)[\varphi_Y(u)-1]}$$

Nota. Además, y de (12), se tiene que:

$$E[Z(t)] = E_{X(t)}[E\{Z(t)/X(t)=n\}] = \sum_{n=0}^{\infty} E\{Z(t)/X(t)=n\} \, P\{X(t)=n\} \underset{(Z(0)=0)}{=}$$

$$= \sum_{n=1}^{\infty} n E(Y) \, P(X(t)=n) = E(Y) \sum_{n=1}^{\infty} n P\{X(t)=n\} = E(Y) \, E[X(t)] = \lambda t \, E(Y)$$

$$Var[Z(t)] = E\{[Z(t)]^2\} - [E\{Z(t)\}]^2 = E\{[Z(t)]^2\} - \lambda^2 t^2 E^2(Y)$$

$$E\{[Z(t)]^2\} = \sum_{n=0}^{\infty} E\{[Z(t)]^2/X(t)=n\} . P\{X(t)=n\} = \sum_{n=0}^{\infty} \{Var[Z(t)/X(t)=n] + (E[Z(t)/X(t)=n])^2\} P(\ldots$$

$$= \sum_{n=0}^{\infty} \{n \, Var(Y) + n^2 E^2(Y)\} P\{X(t)=n\} = Var(Y)\lambda t + E^2(Y) \, E\{[X(t)]^2\} =$$

$$= E(Y^2)\lambda t - \cancel{E^2(Y)\lambda t} + \cancel{E^2(Y)\lambda t} + E^2(Y)\lambda^2 t^2$$

luego $Var[Z(t)] = \lambda t \, E(Y^2)$

4. PROCESOS DE POISSON NO-HOMOGENEOS

Definición

Un proceso de contaje $\{X(t), t \geq 0\}$ es un proceso de Poisson no homogeneo o con incrementos no estacionarios si, y sólo si, las v.a $X(t)$ siguen una distribución de Poisson con tasa instantanea de ocurrencia $\frac{d}{dt} E[X(t)]$ no constante.

Nota. En los procesos de Poisson homogeneos la tasa de ocurrencia es igual a λ

Vamos a deducir axiomáticamente la probabilidad de un proceso de este tipo.

Teorema

Si un proceso de contaje $\{X(t), t \geq 0\}$ satisface los axiomas:

A1.- $X(0)=0$

A2.- Los incrementos son independientes

A3.- $p_1(h/t) = P\{X(t+h)-X(t)=1\} = \lambda(t)h + o(h)$ y $\sum_{n=2}^{\infty} p_n(h/t) = o(h)$

siendo $h \simeq 0$ y $\lambda(t)$ una función continua.

Entonces $X(t) \doteq P_o(\mu(t))$ con $\mu(t) = \int_0^t \lambda(s)\,ds$, i.e. $\{X(t), t \geq 0\}$ es un proceso de Poisson no-homogeneo.

Demostración

El procedimiento de la demostración será idéntico al del Teorema de la pág. 10. Obtendremos:

$$p_m'(t) + \lambda(t)\, p_m(t) = \lambda(t)\, p_{m-1}(t) \qquad (14)$$

con la particularidad de que $p_m(t)$ representa exclusivamente la probabilidad de que se registren m sucesos en $(0,t]$, que no tiene porque coincidir con la de que se produzcan m sucesos en $(T, t+T]$ (no hemos exigido incrementos estacionarios).

Si $\mu(t) = \int_0^t \lambda(s)\, ds$, el factor integrante en esta ocasión será $e^{\mu(t)}$:

$$p_m'(t)\, e^{\mu(t)} + \lambda(t)\, e^{\mu(t)}\, p_m(t) = \frac{d}{dt}\left[p_m(t)\, e^{\mu(t)} \right] = \lambda(t)\, e^{\mu(t)}\, p_{m-1}(t)$$

Ahora la hipótesis de inducción será:

$$p_{m-1}(t) = e^{-\mu(t)} \frac{[\mu(t)]^{m-1}}{(m-1)!}$$

(válida para $m=1$, siguiendo el razonamiento usado en (2)), luego:

$$\frac{d}{dt}\left[p_m(t)\, e^{\mu(t)} \right] = \lambda(t) \frac{[\mu(t)]^{m-1}}{(m-1)!} = \frac{d}{dt}\left[\frac{[\mu(t)]^m}{m!} + C \right]$$

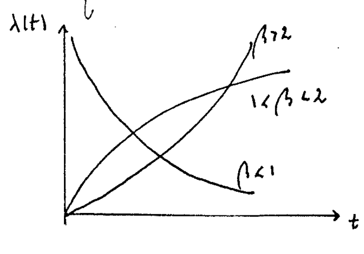
$$\left.\begin{array}{ll} \text{con lo que} & p_m(t) = e^{-\mu(t)} \dfrac{[\mu(t)]^m}{m!} + C\, e^{-\mu(t)} \\ \text{como} & p_m(0) = 0 = C \end{array}\right\} \quad p_m(t) = e^{-\mu(t)} \frac{[\mu(t)]^m}{m!}$$

lo que confirma la hipótesis de inducción. ◆

Comentario. En Teoría de la Fiabilidad aparecen este tipo de procesos, siendo $\lambda(t) = \frac{d}{dt} E[X(t)] \left(= \frac{d}{dt}\mu(t) \right)$ la tasa de avería que usualmente aumenta con el t a medida que el equipo envejece.

Una posible expresión para la tasa de avería es la ley de Weibull: $\lambda(t) = \beta\lambda\, t^{\beta-1}$. En este caso, $p_m(t) = e^{-\lambda t^{\beta}} \dfrac{(\lambda t^{\beta})^m}{m!}$

Con $\beta > 2$, $\lambda(t)$ es una función creciente-convexa, mientras que para $1<\beta<2$ la función es creciente-cóncava. Si $\beta<1$, $\lambda(t)$ es decreciente.



Ley de Weibull

Capítulo 3

Procesos de Markov

TEMA III : PROCESOS DE MARKOV

1. PROCESOS DE MARKOV

Definición

Un proceso estocástico $\{X(t),\ t\in T\}$ es un proceso de Markov si, y sólo si, $\forall n$ y $\forall t_1<t_2 \ldots < t_n$ en T y $x_1, x_2 \ldots x_n$ de $\mathbb{R}$ se tiene que:

$$P\{X(t_n)\le x_n / X(t_1)=x_1, \ldots X(t_{n-1})=x_{n-1}\} = P\{X(t_n)\le x_n / X(t_{n-1})=x_{n-1}\} \qquad (1)$$

Intuitivamente se interpreta diciendo que dado el presente del proceso el futuro es independiente del pasado.

Definición

Un número real x es un valor posible o un estado del proceso estocástico $\{X(t)$ si, y sólo si, existe $t\in T$ tal que:

$$P\{x-h < X(t) < x+h\} > 0 \qquad \forall h>0$$

El espacio de estados S no es más que el conjunto de valores posibles del proceso.

Los procesos de Markov se clasifican de acuerdo con la naturaleza de T y S:

		Espacio de estados : S — DISCRETO	CONTINUO
Naturaleza del parámetro : T	DISCRETO	Cadena de Markov de parámetro discreto	Proceso de Markov de parámetro discreto
	CONTINUO	Cadena de Markov de parámetro continuo	Proceso de Markov de parámetro continuo

Definición

Sea A un intervalo de $\mathbb{R}$. La función $P(x,s;A,t) = P\{X(t)\in A / X(s)=x\}$, para t se llama probabilidad de transición.

Comentario.

En este tema y los siguientes nos dedicaremos al estudio de las cadenas de Markov que es importante desde 2 puntos de vista. Primero, porque las cadenas de Markov tienen propiedades teóricas muy ricas que pueden ser estudiadas a un nivel elemental. Segundo, porque sirven de modelo para una gran cantidad de fenómenos naturales lo que las hace útiles para muchas y variadas aplicaciones.

En lo que se refiere al estudio teórico de las cadenas de Markov nos ocuparemos fundamentalmente de:

(i) el comportamiento dependiente del tiempo de una cadena de Markov. Hallaremos la función de probabilidad de transición resolviendo las ecuaciones en diferencias finitas que satisfacen.

(ii) el comportamiento a largo plazo de tales procesos. En particular determinaremos condiciones bajo las cuales existe una probabilidad π tal que:

$$\pi(A) = \lim_{t \to \infty} P(x,s\,;A,t)$$

con independencia de x y s.

(iii) la naturaleza de ciertos tiempos de transición y de ocupación de un estado.

2. CADENAS DE MARKOV DE PARAMETRO DISCRETO

Definición

Una cadena de Markov discreta $\{X_n,\ n=0,1,2,\ldots\}$ es un proceso de Markov cuyo espacio de estados S es discreto, así como el parámetro.

A efectos de simplificar la notación S suele tomarse igual a N (o como un subconjunto de N si S es finito), y se dice que X_n está en el estado i si $X_n = i$.

Para especificar la ley de probabilidad de una cadena de Markov $\{X_n,\ n \in N\}$, basta establecer $\forall n,m$ con $n > m \geq 0$ y $\forall i,j \in S$, la función de cuantía:

$$p_i(n) = P\{X_n = i\} \qquad (2)$$

y la función de cuantía condicional:

$$p_{ij}(m,n) = P\{X_n = j \,/\, X_m = i\} \qquad (3)$$

llamada probabilidad de transición.

Teorema

La ley de probabilidad de una cadena de Markov se determina mediante las funciones de cuantía (2) y (3), y $\forall k$, $\forall n_1, n_2 \ldots n_k$ tales que $n_k > n_{k-1} > \ldots > n_2 > n_1$, y $\forall i_1, i_2 \ldots i_k$

$$P\{X_{n_1} = i_1,\ X_{n_2} = i_2 \ldots X_{n_k} = i_k\} = p_{i_1}(n_1)\, p_{i_1 i_2}(n_1, n_2) \ldots p_{i_{k-1} i_k}(n_{k-1}, n_k) \qquad (4)$$

Demostración

Aplicando la ley multiplicativa de la probabilidad se tiene que:

$$P\{X_{n_1} = i_1,\ X_{n_2} = i_2 \ldots X_{n_k} = i_k\} = P\{X_{n_k} = i_k \,/\, X_{n_1} = i_1 \ldots X_{n_{k-1}} = i_{k-1}\} \cdot P\{X_{n_1} = i_1 \ldots X_{n_{k-1}} = i_{k-1}\}$$

$$\overset{(1)}{=} P\{X_{n_k} = i_k \,/\, X_{n_{k-1}} = i_{k-1}\}\, P\{X_{n_1} = i_1, \ldots X_{n_{k-1}} = i_{k-1}\} = p_{i_{k-1} i_k}(n_{k-1}, n_k)\, P\{X_{n_1} = i_1 \ldots X_{n_{k-1}} = i_{k-1}\}$$

de nuevo:

$$P\{X_{n_1} = i_1 \ldots X_{n_{k-1}} = i_{k-1}\} = P\{X_{n_{k-1}} = i_{k-1} \,/\, X_{n_1} = i_1, \ldots X_{n_{k-2}} = i_{k-2}\}\, P\{X_{n_1} = i_1 \ldots X_{n_{k-2}} = i_{k-2}\}$$

$$\overset{(1)}{=} p_{i_{k-2} i_{k-1}}(n_{k-2}, n_{k-1})\, P\{X_{n_1} = i_1, \ldots X_{n_{k-2}} = i_{k-2}\}$$

y así sucesivamente.

Definición

Una cadena de Markov es homogénea o tiene probabilidades de transición estacionarias si $p_{ij}(m,n)$ sólo depende de $n-m$. Entonces

$p_{ij}(n) = P\{X_{t+n}=j / X_t = i\}$ $\forall t \in N$ es la probabilidad de transición en n pasos.

La probabilidad de transición de un paso $p_{ij}(1)$ se escribirá simplemente p_{ij}.

Cadena de Markov de dos estados (CM-2E)

Constituye el modelo más sencillo de cadena de Markov, sin embargo corresponde a situaciones prácticas. Consideremos, por ejemplo, dos almacenes de ultramarinos en un mismo mercado. Debido a los mecanismos de competencia y a las ofertas especiales los clientes alternan sus compras (aleatoriamente). Si un cliente compra en el almacén 1 (respect. 2) una determinada semana, la semana siguiente cambia al almacén 2 (respect. 1) con probabilidad p (respect. q). Entonces, la matriz de probabilidades de transición de un paso es:

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \left\{ \begin{array}{cc} 1-p & p \\ q & 1-q \end{array} \right\} \end{array} \qquad 0<p<1 \qquad 0<q<1$$

Vamos a calcular la probabilidad de ocupación: $p(n+1) = [p_1(n+1), p_2(n+1)]$ sabemos que

$$p_1(n+1) = P\{X_{n+1}=1\} = P\{X_{n+1}=1, X_n=1\} + P\{X_{n+1}=1, X_n=2\} = P\{X_n=1\}\, P\{X_{n+1}=1 / X_n=1\}$$

$$+ P\{X_n=2\}\, P\{X_{n+1}=1 / X_n=2\} = p_1(n)(1-p) + p_2(n)\, q = p_1(n)(1-p) + [1-p_1(n)]\, q$$

$$= (1-p-q)\, p_1(n) + q$$

Por recurrencia:

$$p_1(n+1) = q + (1-p-q)\, p_1(n) = q[1+(1-q-p)] + (1-p-q)^2 p_1(n-1) =$$

$$= q \sum_{k=0}^{2} (1-p-q)^k + (1-p-q)^3 p_1(n-2) = \ldots =$$

$$= q \sum_{k=0}^{n} (1-p-q)^k + (1-p-q)^{n+1} p_1(0) =$$

$$= q\, \frac{(1-p-q)^{n+1}-1}{(1-p-q)-1} + (1-p-q)^{n+1} p_1(0) = \frac{q}{p+q} + (1-p-q)^{n+1}\left[p_1(0) - \frac{q}{p+q}\right]$$

Ahora, como $\forall n$, $n=0,1,2\ldots$ $p_2(n) = 1 - p_1(n)$, resulta:

$$p_2(n+1) = \frac{p}{p+q} + (1-p-q)^{n+1}\left[p_2(0) - \frac{p}{p+q}\right]$$

Es evidente que para determinar la probabilidad de ocupación $p(n)$ es suficien el conocimiento de la probabilidad incondicional iniciales $p(0) = [p_1(0), p_2(0)]$.

Vamos a calcular ahora la probabilidad de transición en n pasos: $p_{ij}($ que representaremos matricialmente mediante $P(n) = [p_{ij}(n)]$.

Sea $p_1(0) = 1$, entonces:

$p_{11}(n) = P\{X_n = 1 / X_0 = 1\}$

se calculará teniendo en cuenta la relación: $p_1(n) = p_{11}(n)\, p_1(0) + p_{21}(n)\, p_2(0)$. E este caso:

$$p_{11}(n) = p_1(n) = \frac{q}{p+q} + (1-p-q)^n \left[1 - \frac{q}{p+q}\right] = \frac{q}{p+q} + (1-p-q)^n \frac{p}{p+q}$$

Como $p_{12}(n) = 1 - p_{11}(n)$, se tiene que

$$p_{12}(n) = \frac{p}{p+q} - (1-p-q)^n \frac{p}{p+q}$$

Análogamente, si $p_2(0) = 1$, tendremos que:

$$p_{21}(n) = p_1(n) = \frac{q}{p+q} - (1-p-q)^n \frac{q}{p+q}$$

y como $p_{22}(n) = 1 - p_{21}(n)$: $\quad p_{22}(n) = \dfrac{p}{p+q} + (1-p-q)^n \dfrac{q}{p+q}$

En definitiva:

$$P(n) = \frac{1}{p+q}\begin{bmatrix} q & p \\ q & p \end{bmatrix} + \frac{(1-p-q)^n}{p+q}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix}$$

y puede comprobarse que $P(n) = P^n$.

Diagonalización

3. ECUACION DE CHAPMAN-KOLMOGOROV

Teorema

En una cadena de Markov de parámetro discreto, para todo $n > u > m \geq 0$ y $\forall i,j \in S$ cumple:

$$p_{ij}(m,n) = \sum_{k \in S} p_{ik}(m,u)\, p_{kj}(u,n) \qquad (5)$$

Demostración

Realmente la ecuación (5) sólo tiene sentido si $p_i(m) > 0$, dado que si esta p bilidad es nula las correspondientes probabilidades del tipo $p_{i\cdot}(m,\cdot)$ son irrel vantes (nunca se producirá una tal transición). En este caso:

$$p_{ij}(m,n) \underset{(3)}{=} P\{X_n = j / X_m = i\} = \frac{P\{X_m = i, X_n = j\}}{P\{X_m = i\}} = \frac{1}{p_i(m)} \sum_{k \in S} P\{X_m = i, X_u = k, X_n = j\} =$$

$$= \frac{1}{p_i(m)} \sum_{k\in S} P\{X_n=j \,/\, X_m=i,\ X_u=k\} \cdot P\{X_m=i,\ X_u=k\} =$$

$$= \frac{1}{p_i(m)} \sum_{k\in S} P\{X_n=j \,/\, X_u=k\}\ P\{X_u=k \,/\, X_m=i\}\ P\{X_m=i\} = \sum_{k\in S} p_{kj}(u,n)\, p_{ik}(m,u)$$

La probabilidad de transición de una cadena de Markov se representan de forma matricial:

$$P(m,n) = \begin{pmatrix} p_{00}(m,n) & p_{01}(m,n) & \cdots & p_{0j}(m,n) & \cdots \\ p_{10}(m,n) & p_{11}(m,n) & \cdots & p_{1j}(m,n) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{i0}(m,n) & p_{i1}(m,n) & & p_{ij}(m,n) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

Esta matriz de probabilidades de transición es una matriz estocástica: sus elementos son no-negativos y los que hay en cada fila suman 1.

En forma matricial la ecuación de Chapman-Kolmogorov (5), se convierte en

$$P(m,n) = P(m,u)\, P(u,n) \qquad (6)$$

y de (6) se deduce:

$$P(m,n) = P(m,m+1)\, P(m+1,m+2) \cdots P(n-2,n-1)\, P(n-1,n) \qquad (7)$$

En consecuencia, para conocer las probabilidades de transición $P(m,n)$ $\forall n,m$, basta con conocer las matrices de probabilidades de transición de un paso.

Por otra parte, sea $p(n) = [p_0(n)\ \ p_1(n) - p_i(n) -]$ $\forall n$, se tendrá la relación

$$p(n) = p(0)\, P(0,n) \qquad (8)$$

Nota. En el caso de una cadena de Markov homogénea, si llamamos $P(n) = [p_{ij}(n)]$ y $P = [p_{ij}]$, tendríamos las ecuaciones:

$$P(n) = P^n \qquad (7)'$$

y

$$p(n) = p(0)\, P^n \qquad (8)'$$

es decir, la ley de probabilidad de una cadena de Markov homogénea queda determinada por la matriz de probabilidad de transición de un paso P y la probabilidad incondicional inicial $p(0)$.

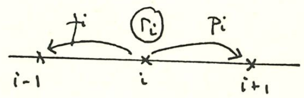
Limitaremos nuestro estudio a las cadenas homogéneas que sirven como modelo en la mayor parte de los problemas reales.

Veremos a continuación algunos ejemplos de cadenas de Markov, que constituyen una clase especial de procesos de ramificación.

4. PROCESOS DE RAMIFICACION

1. Camino aleatorio de una dimensión

Un camino aleatorio es una cadena de Markov en un espacio de estados $S \subset \mathbb{Z}$ con la propiedad de que si el sistema está en un estado i, en una transición el sistema o permanece en i o se desplaza a uno de los estados contiguos al i:



Así, por ejemplo, el camino aleatorio con espacio de estados $\mathbb{N}$ tiene una matriz de probabilidades de transición:

$$P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & & 0 & 0 & 0 \\ q_1 & r_1 & p_1 & 0 & & 0 & 0 & 0 \\ 0 & q_2 & r_2 & p_2 & & 0 & 0 & 0 \\ - & - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & \dots & q_i & r_i & p_i & \dots \\ - & - & - & - & - & - & - & - \end{pmatrix}$$

con $p_i > 0$, $q_i > 0$ y $r_i \geq 0$ y $q_i + r_i + p_i = 1$ $\forall i \geq 1$. Además, $p_0 \geq 0$, $r_0 \geq 0$, $r_0 + p_0 = 1$

Ejemplo.

La fortuna de un jugador comprometido en una serie de partidas se puede representar mediante un camino aleatorio. Supongamos que un jugador A juega contra un adversario B infinitamente rico (un casino, p.e.), y que cuando su fortuna es k tiene una probabilidad p_k de ganar una unidad, una probabilidad q_k de perderla y otra r_k de conservar fijos sus recursos.

Un caso especial es el de partidas repetidas e independientes de un juego en el que A tiene la probabilidad p de ganar. En estos casos: $p_k = p$, $q_k = 1-p$ y $r_k = 0$.

En general, respecto a r_0 y p_0, suelen hacerse dos hipótesis:

(1) si $r_0 = 1$ y $p_0 = 0$, se dice que el camino aleatorio tiene una barrera absorbente 0. En este caso, el adversario de A juega para obtener su fortuna y deja de jugar cuando lo ha conseguido.

(2) si $r_0 > 0$ y $p_0 > 0$, entonces hay una barrera reflectante en el estado 0. El jugador permite que A siga jugando cuando no dispone de dinero.

Otro caso de interés es cuando B comienza también con una fortuna limitada y entonces si A comienza con un capital x, se tiene $x + y = a$.

Así, sea $\{X_n, n \in \mathbb{N}\}$ es una cadena de Markov que representa la fortuna de A, cuyo espacio de estados es $S = \{0, 1, 2, \dots, a\}$. Si admitimos que el juego termina cuando uno de los dos jugadores queda arruinado, los estados 0 y a son absorbentes, y en este caso:

$$P=\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & q_2 & r_2 & p_2 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & q_{a-1} & r_{a-1} & p_{a-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

2. Modelo de crecimiento de una población

Supongase que se está estudiando una población homogenea cuyo mecanismo de crecimiento permanece invariante en el tiempo. X_n denotará el número de individuos que componen la población en el período n. La probabilidad de que un individuo de n tenga j "descendientes" en $n+1$ es $f_j \geq 0$. Se cumplirá $\sum_{j=0}^{\infty} f_j = 1$

Tenemos la siguiente relación:

$$p_{ij} = \sum_{j_1+j_2+\cdots+j_i=j} f_{j_1} f_{j_2} \cdots f_{j_i} = f_j^{(i*)} \qquad (*)$$

si llamamos f al vector $f=(f_0, f_1, \ldots)$, y sea $G(s)$ la función generatriz de probabilidad de f, i.e.

$$G(s) = \sum_{j=0}^{\infty} s^j f_j$$

es evidente que $f_j^{(i*)}$ es el coeficiente de s^j en el desarrollo en serie de potencias $[G(s)]^i$.

Supongamos que se comienza con un individuo en $n=0$. La probabilidad de tener j individuos en $n=1$ es f_j. Determinemos la probabilidad de que la población esté integrada por j individuos en $n=2$. La probabilidad de tener i individuos en $n=1$ y j individuos en $n=2$ es: $f_i f_j^{(i*)}$, entonces la prob. de tener j elementos en $n=2$ se obtendrá sumando sobre i:

$$p_j(2) = \sum_{i=0}^{\infty} f_i f_j^{(i*)}$$

que es el coeficiente de s^j en el desarrollo de $G^{(2)}(s) = G[G(s)] = \sum_{i=0}^{\infty} f_i [G(s)]^i$
en efecto:

$$\sum_{i=0}^{\infty} f_i [G(s)]^i = \sum_{i=0}^{\infty} f_i \left(\sum_{j=0}^{\infty} f_j^{(i*)} s^j \right) \underset{(*)}{=} \sum_{j=0}^{\infty} \Big(\underbrace{\sum_{i=0}^{\infty} f_i f_j^{(i*)}}_{p_j(2)} \Big) s^j$$

Ahora tenemos que la probabilidad de tener i individuos en $n=1$ y j en $n=3$ es: $f_i\, p_j^{(i*)}(2)$ $(**)$ (siendo $p_j^{(i*)}(2) = p_{ij}(1,3)$). Y la probabilidad de que haya j individuos en $n=3$ será:

$$p_j(3) = \sum_{i=0}^{\infty} f_i\, p_j^{(i*)}(2)$$

que coincide con el coeficiente de s^j en el desarrollo de $G^{(3)}(s) = G[G^{(2)}(s)] = \sum_{i=0}^{\infty} f_i [G^{(2)}$

(*) Dada la convergencia absoluta de la serie, se pueden intercambiar los sumatorios ya que:

$$\sum_{i=0}^{\infty} p_i\,[G^{(2)}(s)]^i = \sum_{i=0}^{\infty} p_i \left(\sum_{j=0}^{\infty} p_j(2)\, s^j\right)^i = \sum_{i=0}^{\infty} p_i \left(\sum_{j=0}^{\infty} p_j^{(i*)}(2)\, s^j\right) = \sum_{j=0}^{\infty} \Bigg(\underbrace{\sum_{i=0}^{\infty} p_i\, p_j^{(i*)}(2)}_{p_j(3)}\Bigg) s^j$$

En general, la probabilidad de que la población esté constituida por j individuos en la generación n-ésima, $p_j(n)$, viene dado por el coeficiente de s^j en el desarrollo de $G^{(n)}(s) = G[G^{(n-1)}(s)]$.

Ciertos aspectos del crecimiento o extinción de la población pueden ser descritos estudiando esta función $G^{(n)}(s)$. Veamos:

Sea e_n la probabilidad de que la población se extinga antes o en la n-ésima generación, se tendrá que: $\forall n$ $e_n \le e_{n+1}$, y $e_n = P(X_n = 0) = G^{(n)}(0)$

Supongamos que se está interesado en la probabilidad de una eventual extinción de la población: $e = \lim_{n\to\infty} e_n$

Puede comprobarse que la función generatriz de probabilidad $G(s)$ es una función continua, no decreciente, convexa y no negativa sobre $0 \le s \le 1$, siendo $G(1) = 1$.

La probabilidad de extinción e en un proceso de ramificación es una solución de la ecuación $G(s) = s$, puesto que:

$$e = \lim_{n\to\infty} e_{n+1} = \lim_{n\to\infty} G[G^{(n)}(0)] \underset{\text{cont. de } G}{=} G[\lim_{n\to\infty} G^{(n)}(0)] = G(e)$$

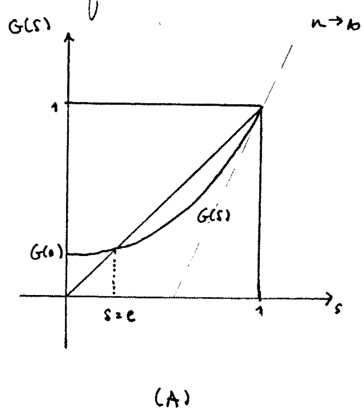
Además, por la convexidad de $G(s)$, e es la solución más pequeña de $G(s)=s$ en el intervalo $[0,1]$, la otra solución es 1. Veámoslo:

Sea $s_0 \in [0,1]$ una raíz de $G(s) = s$. Por el carácter no decreciente de G:

$$0 \le s_0 \longrightarrow e_1 = G(0) \le G(s_0) = s_0$$

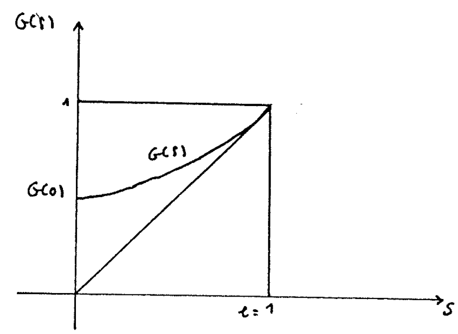
Supongamos que $e_n \le s_0$, entonces: $e_{n+1} = G^{(n+1)}(0) = G[G^{(n)}(0)] = G(e_n) \le G(s_0)$

En definitiva: $e = \lim_{n\to\infty} e_n \le s_0$.



(A)

$G(0) = s_0$



(B)

Comentarios. Nótese que si $g_0 = 0$, la probabilidad de extinción es nula, lógicamente.

También puede ocurrir que la primera solución de $G(s)=s$ sea $s=1$, lo cual significa que la población se extinguirá eventualmente con probabilidad 1 (Fig. B)

En el caso (A): $G'(1) > 1$, y en el (B): $G'(1) \le 1$; pero, sabemos que $G'(1) = \sum_{k=0}^{\infty} k\, q_k = \mu$ es el número esperado de individuos producidos en una generación a partir de un individuo de la generación precedente. Entonces, si $\mu \le 1$ la población se extinguirá eventualmente.

Si en una población hay varios individuos inicialmente, por ejemplo r, la probabilidad de una extinción eventual será e^r, que es pequeña para $e<1$, incluso con un r moderado (hemos supuesto que los r individuos actúan independientemente).

Nota. Si nos interesa determinar el número esperado de individuos que constituyen la n-ésima generación, es decir,

$$E(X_n) = \sum_{j=0}^{\infty} j\, P(X_n = j) = \sum_{j=0}^{\infty} j\, p_j(n)$$

podemos calcularlo a partir, asimismo de $G(s)$, en efecto:

$$G^{(n)}(s) = \sum_{j=0}^{\infty} p_j(n)\, s^j \longrightarrow \frac{d}{ds} G^{(n)}(s)\Big|_{s=1} = \sum_{j=0}^{\infty} j\, p_j(n)\, s^{j-1}\Big|_{s=1} = \sum_{j=0}^{\infty} j\, p_j(n)$$

entonces

$$E(X_n) = \frac{d}{ds} G^{(n)}(s)\Big|_{s=1} = \frac{d}{ds} G\left[G^{(n-1)}(s)\right]\Big|_{s=1} = \frac{d\, G\left[G^{(n-1)}(s)\right]}{d\, G^{(n-1)}(s)}\Big|_{s=1} \frac{d\, G^{(n-1)}(s)}{ds}\Big|_{s}$$

$(G^{(n-1)}(1)=1)$

$$= \frac{d\, G(t)}{dt}\Big|_{t=1} \frac{d\, G^{(n-1)}(s)}{ds}\Big|_{s=1} = G'(1)\, \frac{d\, G^{(n-1)}(s)}{ds}\Big|_{s=1} = \mu^n \quad \text{(por inducción)}$$

3. Fila de espera discreta

Los clientes llegan para ser servidos y toman su turno en una fila de espera. Supongamos que los tiempos de servicio a los clientes sucesivos son v.a. independientes y equidistribuidas. Para $n \in N$, $n \ge 1$, sea X_n el número de personas que esperan en la cola en el momento en que la n-ésima persona acaba de ser servida. Veamos que $\{X_n, n \in N\}$ es una cadena de Markov.

En efecto, sea U_n el número de clientes que llegan a la cola en el intervalo de tiempo en el que se está dando servicio al cliente n-ésimo. Entonces:

$$X_{n+1} = X_n - \delta(X_n) + U_{n+1} \qquad \text{siendo} \quad \delta(x) = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases}$$

es decir, el número de personas que esperan servicio cuando el cliente n+1 se va, una vez servido, depende de si el cliente n+1 estaba en la cola o no cuando terminó el n-ésimo. Por lo tanto,

$$X_{n+1} = \begin{cases} \eta_{n+1} & \text{si } X_n = 0 \\ \eta_{n+1} + X_n - 1 & \text{si } X_n > 0 \end{cases}$$

Como η_{n+1} es independiente de $X_1 \ldots X_n$, dado X_n, no es necesario conocer $X_1, X_2 \ldots X_{n-1}$ para determinar la distribución condicionada de X_{n+1}.

Esta cadena de Markov $\{X_n, n \in \mathbb{N}\}$ es una cadena de Markov inmersa ya que corresponde a la observación del proceso estocástico $\{N(t), t \geq 0\}$ con N representando el número de clientes que hay en la cola en el instante t, en una sucesión de tiempos $\{t_n\}$ relativos a los momentos en que los clientes sucesivos acaban de recibir servicio, i.e. $X_n = N(t_n)$.

Supongamos que el tiempo de servicio es prácticamente constante y el nº de clientes que llegan a la cola durante el tiempo en que se está dando servicio a un cliente se distribuye según la siguiente función de cuantía
$P(\eta_n = k) = a_k$, $k \in N$ $\forall n$ siendo $a_k \geq 0$ y $\sum_{k=0}^{\infty} a_k = 1$

Entonces, la cadena de Markov correspondiente es homogénea con probabilidades de transición de un paso:

$$p_{ij} = P\{X_{n+1} = j / X_n = i\} = P\{X_n - \delta(X_n) + \eta_{n+1} = j / X_n = i\} = P\{\eta_{n+1} = j - i + \delta(X_n) / X_n = i\}$$

Si $i = 0$: $p_{0j} = P\{\eta_{n+1} = j\} = a_j$

Si $i \neq 0$: $p_{ij} = P\{\eta_{n+1} = j - i + 1\} = \begin{cases} a_{j-i+1} & \text{si } j+1 \geq i \\ 0 & \text{si } j+1 < i \end{cases}$

Por lo tanto, la matriz de probabilidades de transición de un paso es:

$$P = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ 0 & 0 & 0 & a_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

Intuitivamente, queda claro que si el nº medio de clientes que llegan durante el tiempo de servicio: $\mu = \sum_{k=0}^{\infty} k a_k$, es tal que $\mu > 1$, la longitud de la cola aumentará indefinidamente. Recíprocamente, se verá que si $\mu < 1$, la longitud de la fila de espera tiende hacia un equilibrio. Si $\mu = 1$, se manifiesta una inestabilidad.

Nota. El uso de la cadena de Markov inversa constituye una técnica útil para estudiar las propiedades de los procesos estocásticos de parámetro continuo

5. CADENA DE NACIMIENTO-MUERTE

Esta cadena constituye una versión discreta del proceso de nacimiento-muerte cuya versión continua es mucho más realista (tema 6).

Definición

Una cadena de Markov $\{X_n, n \in T\}$, con espacio de estados $S = \{0,1,2 \dots\}$ se llama **cadena de nacimiento-muerte** si sus probabilidades de transición son:

$$p(x,x+1) = p_x \,, \quad p(x,x-1) = q_x \,, \quad p(x,x) = r_x$$

donde $0 \le p_x, q_x, r_x \le 1$, $p_x + q_x + r_x = 1$, $\forall x \in S$.

Se trata, pues, de un camino aleatorio de una dimensión.

Representaremos por t_x el número de transiciones que tienen de producirse hasta que se presenta, por primera vez, el estado x (suponiendo que la cadena arranca de un estado distinto).

Para $b \in S$, $b \neq 0$, pretendemos calcular $f(x) = P\{t_0 < t_b / X_0 = x\}$, $0 < x <$

Dado que:

$$f(x) = P\{t_0 < t_b, X_1 = x-1 / X_0 = x\} + P\{t_0 < t_b, X_1 = x, X_0 = x\} + P\{t_0 < t_b, X_1 = x+1 / X_0 =$$

$$= P\{t_0 < t_b / X_1 = x-1\} \, P\{X_1 = x-1 / X_0 = x\} + P\{t_0 < t_b / X_1 = x\} \, . \, P\{X_1 = x / X_0 =$$

$$+ P\{t_0 < t_b / X_1 = x+1\} \, P\{X_1 = x+1 / X_0 = x\} = f(x-1)\, q_x + f(x)\, r_x + f(x+1)\, p_x$$

luego $f(x) = f(x-1)\, q_x + f(x)\, r_x + f(x+1)\, p_x \qquad 0 < x < b$

con las condiciones límite : $f(0) = 1$, $f(b) = 0$

Ahora bien : $r_x = 1 - p_x - q_x$. $f(x) - f(x)\, r_x = f(x)\,(p_x + q_x) = f(x-1)\, q_x + f(x+1)$

entonces $\{f(x+1) - f(x)\}\, p_x = \{f(x) - f(x-1)\}\, q_x$

Si $p_x > 0$, $q_x > 0$ $\forall x \in S$: $f(x+1) - f(x) = \dfrac{q_x}{p_x} \{f(x) - f(x-1)\}$ y por recurrencia

$$f(x+1) - f(x) = \frac{q_x \, q_{x-1} \cdots q_1}{p_x \, p_{x-1} \cdots p_1} \underline{\{f(1) - f(0)\}}.$$

Si llamamos $a_x = \dfrac{q_x \, q_{x-1} \cdots q_1}{p_x \, p_{x-1} \cdots p_1}$ y convenimos que $a_0 = 1$, tendremos

$$f(x) - f(x+1) = a_x \{f(0) - f(1)\} \qquad 0 \le x < b$$

Si sumamos sobre $x=0,1\ldots b-1$, tenemos que: $1=f(0)-f(b)=[f(0)-f(1)]\sum_{x=0}^{b-1}a$

luego $f(0)-f(1)=\dfrac{1}{\sum_{x=0}^{b-1}a_x}$, y sustituyendo: $f(x)-f(x+1)=\dfrac{a_x}{\sum_{y=0}^{b-1}a_y}$ $\quad 0\le x<b$

Podemos escribir: $f(x)=\sum_{y=x}^{b-1}[f(y)-f(y+1)]=\dfrac{\sum_{y=x}^{b-1}a_y}{\sum_{y=0}^{b-1}a_y}$

En consecuencia: $P\{t_0<t_b/X_0=x\}=\dfrac{\sum_{y=x}^{b-1}a_y}{\sum_{y=0}^{b-1}a_y}$ $\quad 0\le x<b$

Tiempo de absorción en un camino aleatorio

Sea $\{X_n, n\in N\}$ una cadena nacimiento-muerte tal que $S=\{0,1\ldots b\}$, siendo 0 y b estados absorbentes, siendo $p_x=p$, $q_x=q$, $r_x=r$ $\forall x\in S$, con $p>0$, $q>0$. Definimos el tiempo de absorción o tiempo de ruina como:

$T=\min\{n / X_n=0 \text{ ó } X_n=b\}$

Teorema

Sea T el tiempo de absorción en un camino aleatorio con 0 y b como estados absorbentes, entonces se cumple:

(i) $P(T<\infty)=1$

(ii) $E(T)<+\infty$

Demostración

(i) Sea A el suceso $\{T=\infty\}=\{0<X_n<b,\ \forall n\}$, veremos que $P(A)=0$.

Definimos $A_n=\{0<X_k<b,\ 0\le k\le n\}$, es evidente que $A=\bigcap_{n\ge 0}A_n$, luego $\forall n\ge 0$ $P(A)\le P(A_n)$.

Dado que $X_n-X_0=Z_1+\ldots+Z_n$, $n\ge 1$, siendo las v.a Z_i independientes equidistribuidas, tales que:

$$Z_i=\begin{cases}1 & p\\ -1 & q\\ 0 & r\end{cases}\qquad \text{siendo } p>0,\ q>0$$

Las variables, para $k\ge 1$: $Y_k=X_{kb}-X_{(k-1)b}=Z_{(k-1)b+1}+\ldots+Z_{kb}$ (*) son también independientes y equidistribuidas (nótese que las $\{Y_k, k\ge 1\}$ tienen todas el mismo número b de sumandos, que no se solapan). Y se tendrá que:

$d\equiv P(Y_k=b \text{ ó } Y_k=-b)=p^b+q^b<(p+q)^b\le 1^b=1$

Entonces: $P(A_{nb})=P\{\bigcap_{k=0}^{nb}\{0<X_k<b\}\}\le P\{\bigcap_{k=1}^{n}\{0<X_{kb}<b\}\}\le P\{\bigcap_{k=1}^{n}\{-b<Y_k<$

$= (1-d)^n \equiv s^n \qquad 0<s<1 \qquad \longrightarrow \qquad P(A_{nb}) \le s^n \quad \forall n .$

Luego : $0 \le P(A) \le P(A_{nb}) \le s^n \quad \forall n$, y tomando $n \to \infty$, se concluye que $P(A)$

(ii) Para todo $n \ge 0$, existirá $k \ge 1$ tal que $(k-1)b \le n < kb$. Entonces

$P(T>n) = P(A_n) \le P(A_{(k-1)b}) \le s^{k-1}$

Por otra parte :

$$E(T) = \sum_{n=1}^{\infty} n P(T=n) = \sum_{n=1}^{\infty} n \{ P(T \ge n) - P(T \ge n+1) \} = \sum_{n=1}^{\infty} P(T \ge n) = \sum_{n=0}^{\infty} P(T>n) \overset{(**)}{\le}$$

$$b \sum_{k=0}^{\infty} s^k = \frac{b}{1-s} < +\infty \qquad \blacklozenge$$

Ejemplo : Problema de la ruina de un jugador

Consideremos el caso en que el camino aleatorio corresponde a una sucesión de partidas independientes de un mismo juego entre Pedro y Juan, que disponen de fortuna inicial a y b, respectivamente. La v.a X_n denotará la fortuna de Pedro después de realizada la n-ésima partida. El espacio de estados $S = \{0,1,2, \dots, a+b\}$, siendo 0 y $a+b$ estados absorbentes (corresponden a la ruina de Pedro y Juan, respectivamente). En este caso : $p_x = p$, $q_x = q$, $r_x = r$, $\forall x \in S$ con $p>0$, $q>0$. Ahora $P\{t_0 < t_{a+b} / X_0 = a\}$ constituye la probabilidad de que Pedro se arruine antes que Juan, y su valor es :

$$P(t_0 < t_{a+b} / X_0 = a) = \frac{\sum_{y=a}^{a+b-1} \left(\frac{q}{p}\right)^y}{\sum_{y=0}^{a+b-1} \left(\frac{q}{p}\right)^y} = \frac{\left(\frac{q}{p}\right)^{a+b} - \left(\frac{q}{p}\right)^a}{\left(\frac{q}{p}\right)^{a+b} - 1} = \frac{c^{a+b} - c^a}{c^{a+b} - 1} \qquad \text{para } c = \frac{q}{p} \neq$$

Si $\frac{p}{q} = 1 \longrightarrow P(t_0 < t_{a+b} / X_0 = a) = \dfrac{(a+b-1) - a + 1}{(a+b-1) - 0 + 1} = \dfrac{b}{a+b}$

Vamos a calcular ahora la fortuna esperada de Pedro, en el instante T en que se produce la ruina de uno de los jugadores :

Si $E(Z_k) = p - q \equiv \mu$ y dado que $X_0 = a$

$$E\{X_T\} = E\{a + \sum_{k=1}^{T} Z_k\} = a + E\{\sum_{k=1}^{T} Z_k\} \overset{(*)}{=} a + \sum_{n=1}^{\infty} E\{\sum_{k=1}^{T} Z_k / T=n\} \cdot P(T=n) =$$

$$= a + \sum_{n=1}^{\infty} n\mu P(T=n) = a + \mu E(T)$$

Nota : La relación $E\{X_T\} = a + \mu E(T)$ se conoce con el nombre de 1ª identidad de WALD.

Vamos a aplicar esta relación para calcular $P(t_0 < t_{a+b} / X_0 = a) = P(X_T = 0)$. Lo haremos del siguiente modo:

$$E\{X_T\} = 0 \cdot P(X_T = 0) + (a+b)\, P(X_T = a+b)$$

Si $p = q$, $\mu = 0$ y la fórmula de Wald nos dice que:

$$E\{X_T\} = a = (a+b)\, P\{X_T = a+b\} \longrightarrow P(X_T = a+b) = \frac{a}{a+b}$$

en consecuencia $P(X_T = 0) = \frac{b}{a+b}$ (resultado que ya habíamos obtenido).

Si $p \neq q$, $\mu \neq 0$ y la fórmula de Wald no puede ser utilizada, dado que se conoce el valor exacto de $E(T)$.

<u>Duración esperada del camino aleatorio simple</u>

Se ha demostrado que $E(T) < \infty$. Vamos a intentar obtener una expresión para dicha duración esperada.

De la 1ª fórmula de Wald, si $\mu = p - q \neq 0$:

$$E(T) = \frac{1}{\mu}\left(E(X_T) - a\right) = \frac{1}{p-q}\left[(a+b)\frac{1-c^a}{1-c^{a+b}} - a\right] \qquad \text{siendo } c = \frac{q}{p}$$

En el caso en que $\mu = 0$ no puede usarse la 1ª identidad de Wald, por lo que procedemos a deducir la llamada 2ª identidad de Wald.

Tenemos que $Var\{Z_k\} = p + q = 1 - r \equiv \sigma^2$, $k \geq 1$ (dado que $\mu = 0$), entonces:

$$Var\{X_T\} = E\{X_T^2\} - a^2 = E\Big\{\Big(\sum_{k=1}^{T} Z_k\Big)^2\Big\} = \sum_{n=1}^{\infty} E\Big\{\Big(\sum_{k=1}^{T} Z_k\Big)^2 / T = n\Big\}\, P(T=n) = \sigma^2 \sum_{n=1}^{\infty} n\, P(T=n) =$$

$$= \sigma^2 E(T) = (1-r)\, E(T)$$

Ahora bien:

$$Var\{X_T\} = 0^2\, P(X_T = 0) + (a+b)^2 P(X_T = a+b) - a^2 = (a+b)^2 \frac{a}{a+b} - a^2 = ab$$

en consecuencia: $E(T) = \frac{ab}{1-r}$. Siendo $Var(X_T) = \underbrace{(1-r)}_{\sigma^2} E(T)$ la <u>2ª identidad de Wald</u>.

<u>Un camino aleatorio en carcinogénesis</u>

Un agente inductor de cáncer se conoce como carcinógeno. En el estudio de la carcinogénesis una "etapa" corresponderá a una interacción entre el carcinógeno y la célula normal. La conversión de una célula normal a una célula cancerosa no se producirá necesariamente en una única etapa. El número de etapas para que se produzca tal transición será el número de "mutaciones" intermedias que determinan tal cambio. El nº de estadios intermedios que atravesará una célula sana (estado 0) hasta convertirse en maligna (estado N), constituirá, junto con los dos estados límite (0 y N) el espacio de estados: $S = \{0, 1, 2 \ldots N\}$.

La cadena $\{X_n, n \geq 0\}$, donde X_n es el estadio en que se encuentra la cé-la después de n etapas, constituye una descripción del proceso de mutaci[ón]. Las transmisiones, en cada etapa, acontecen con las siguientes probabilidades:

$$p(x, x+1) = p_x = \frac{x}{N} \qquad x = 1, 2 \ldots N-1$$

$$p(x, x-1) = q_x = \frac{N-x}{N} \qquad x = 1, 2 \ldots N-1$$

$$p(x, x) = 0 \qquad x = 1, 2 \ldots N-1$$

$$p(N, N) = p(0, 0) = 1 \implies \text{los estados } 0 \text{ y } N \text{ son absorbentes.}$$

Por π_x representamos la probabilidad de absorción en el estado N (cél. maligna) si inicialmente la célula se encontraba en el estado x ($X_0 = x$). E[stas] probabilidades estaran definidas para $x \in S$, siendo condiciones límite: $\pi_0 = 0$, π

Definimos la función generatriz de la probabilidad π_x: $G(s) = \sum_{x=0}^{N} \pi_x$

Se deduce que: (método de descomposición mediante la primera etapa)

$$\pi_x = \frac{x}{N}\,\pi_{x+1} + \left(1 - \frac{x}{N}\right)\pi_{x-1} \qquad x = 1, 2 \ldots N-1$$

y de ahí $\quad \pi_x\left[\frac{x}{N} + \left(1 - \frac{x}{N}\right)\right] = \frac{x}{N}\,\pi_{x+1} + \left(1 - \frac{x}{N}\right)\pi_{x-1}$

por lo que $\quad \pi_{x+1} - \pi_x = \frac{N-x}{x}\left(\pi_x - \pi_{x-1}\right) = \frac{(N-x)(N-x+1)}{x(x-1)}\left(\pi_{x-1} - \pi_{x-2}\right) = \cdots =$

$$= \frac{(N-x)(N-x+1)(N-x+2)\ldots(N-1)}{x(x-1)(x-2)\ldots 1}\left(\pi_1 - \pi_0\right) = a_x \pi_1$$

siendo $a_x = \binom{N-1}{x}$.

Por otra parte la relación $\pi_{x+1} - \pi_x = a_x[\pi_1 - \pi_0]$ es válida tambien para $x = 0$, tomando $a_0 = 1 = \binom{N-1}{0}$. Sumando:

$$\sum_{y=0}^{N-1}[\pi_{y+1} - \pi_y] = \pi_N - \pi_0 = 1 - 0 = 1 = \pi_1 \sum_{y=0}^{N-1} a_y = \pi_1 \sum_{y=0}^{N-1}\binom{N-1}{y} = \pi_1\, 2^{N-1} \longrightarrow \pi_1 = 2^{1-N}$$

Sustituyendo: $\quad \pi_{x+1} - \pi_x = a_x\, 2^{1-N}$. Ahora bien:

$$\pi_x = \sum_{y=0}^{x-1}[\pi_{y+1} - \pi_y] = 2^{1-N}\sum_{y=0}^{x-1} a_y \longrightarrow \pi_x = 2^{1-N}\sum_{y=0}^{x-1}\binom{N-1}{y} \qquad x = 1, 2 \ldots N.$$

siendo $\pi_0 = 0$

Capítulo 4

Clasificación de los estados de una cadena de Markov

TEMA 4: CLASIFICACION DE LOS ESTADOS DE UNA CADENA DE MARKOV

1. Clases comunicantes y cerradas

Definición

Sea $p_{ij}(0) = \begin{cases} 1 & \text{si } i=j \\ 0 & \text{si } i\neq j \end{cases}$. Un estado k es accesible desde un estado j si, sólo si, $\exists n\geq 0$ tal que $p_{jk}(n)>0$. Lo que se representa mediante $j \to k$. Se dice que j y k se comunican o que son estados comunicantes si, y sólo si, $j\to k$ y $k\to$ representándose por $j\leftrightarrow k$.

Teorema

La relación binaria definida entre estados por la condición de comunicantes es una relación binaria de equivalencia.

Demostración

La relación tiene las siguientes propiedades:

a) reflexiva: $i\leftrightarrow i$, $\forall i \in S$, puesto que $p_{ii}(0)=1$

b) simétrica: por definición

c) transitiva: probaremos primero la transitividad de la accesibilidad: $\left.\begin{array}{l} i\to j \\ j\to k\end{array}\right\} i\to$

$$\left.\begin{array}{l} i\to j \text{ sii } \exists n\geq 0 \,/\, p_{ij}(n)>0 \\ j\to k \text{ sii } \exists m\geq 0 \,/\, p_{jk}(m)>0 \end{array}\right\}$$

Por la ecuación de Chapman-Kolmogorov:

$$p_{ik}(n+m) = \sum_{h\in S} p_{ih}(n)\, p_{hk}(m) \geq p_{ij}(n)\, p_{jk}(m) > 0$$

Luego:

$$\begin{array}{l} i\leftrightarrow j \text{ sii } \left\{\begin{array}{l} i\to j \\ j\to i \end{array}\right. \\ \\ j\leftrightarrow k \text{ sii } \left\{\begin{array}{l} j\to k \\ k\to j \end{array}\right. \end{array} \longrightarrow \left.\begin{array}{l} i\to k \\ k\to i \end{array}\right\} \text{ sii } i\leftrightarrow k.$$

Esta relación binaria particiona S en clases comunicantes $C(j) = \{k\in S \,/\, k\leftrightarrow j\}$.

Nota. Puede haber una probabilidad positiva de que, partiendo de una clase, se pase a otra distinta (accesibilidad), pero sin embargo es evidente que no se puede volver a la clase inicial.

Definición

Un conjunto C de estados es cerrado si no se puede llegar a ningún estado fuera de C desde C, i.e. $\forall j\in C$ y $\forall k\notin C$ $p_{jk}(n)=0$ $\forall n\in N$.

Cuando un conjunto cerrado consta de un solo elemento se dice que este es absorbente.

Una vez que el proceso ingresa en una clase cerrada permanece en ella para siempre. Una clase cerrada debe contener por lo menos una clase comunicante.

Definición

Una cadena de Markov es irreducible si, y sólo si, existe sólo una clase comunicante.

Teorema

Si en la matriz $P(n) = P^n$ de probabilidades de transición de n pasos se eliminan las filas y columnas correspondientes a los estados que están fuera de la clase comunicante cerrada C, la matriz resultante sigue siendo estocástica, verificándose en C la ecuación de Chapman-Kolmogorov.

Demostración

Sea P la matriz de probabilidades de transición de un paso. Si escribimos los estados de C en primer lugar, tendremos:

$$P = \begin{Bmatrix} P_C & 0 \\ * & * \end{Bmatrix}$$

siendo P_C la submatriz de P formada por aquellos elementos cuyos dos subíndices corresponden a estados de C. P_C ha de ser, por lo tanto, una matriz estocástica. Además, multiplicando por bloques:

$$P(n) = P^n = \begin{Bmatrix} (P_C)^n & 0 \\ * & * \end{Bmatrix}$$

Es decir, cuando la cadena se llega a C podemos prescindir del resto de los estados restringiendo nuestro estudio a la cadena de Markov inducida en C (que naturalmente verificará la ecuación de Chapman-Kolmogorov). ◆

Ejemplos:

Consideremos cuatro cadenas de Markov cada una con espacio de estados $S = \{1,2,3,4\}$ y con matrices de probabilidad de transición de un paso:

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix} \quad P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix} \quad P_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix} \quad P_4 = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{bmatrix}$$

Vamos a determinar la clase comunicante de cada una de estas cadenas:

$$(1) \quad \left.\begin{array}{llll} 1 \to 1 & 2 \to 1 & 3 \to 1 & 4 \to 1 \\ & 2 \to 2 & 3 \to 2 & 4 \to 2 \\ & & 3 \to 4 & 4 \to 3 \\ & & & 4 \to 4 \end{array}\right\} \quad S = \overset{\text{estado absorbente}}{\{1\}} \cup \underbrace{\{2\} \cup \{3,4\}}_{\text{clase no cerrada}}$$

$$(2) \quad \left.\begin{array}{llll} 1 \to 1 & 2 \to 1 & 3 \to 1 & 4 \to 1 \\ & & 3 \to 2 & 4 \to 2 \\ & & & 4 \to 3 \end{array}\right\} \quad S = \overset{\text{estado abs.}}{\{1\}} \cup \underbrace{\{2\} \cup \{3\} \cup \{4\}}_{\text{clase no cerrada}}$$

(3) $1 \to 3$ $\quad 2 \to 1$ $\quad \begin{array}{l} 3 \to 1 \\ 3 \to 2 \end{array}$ $\quad \left.\begin{array}{l} 4 \to 1 \\ 4 \to 2 \\ 4 \to 3 \end{array}\right\}$ $\quad S = \underbrace{\{1,2,3\}}_{\text{clase cerrada } (P_3C)} \cup \{4\}$

(4) $\begin{array}{l} 1 \to 1 \\ 1 \to 2 \end{array}$ $\quad \begin{array}{l} 2 \to 1 \\ 2 \to 2 \\ 2 \to 3 \end{array}$ $\quad \begin{array}{l} 3 \to 1 \\ 3 \to 2 \\ 3 \to 3 \\ 3 \to 4 \end{array}$ $\quad \left.\begin{array}{l} 4 \to 1 \\ 4 \to 2 \\ 4 \to 3 \\ 4 \to 4 \end{array}\right\}$ $\quad S = \{1,2,3,4\} \equiv$ cadena irreducible.

Definición

El estado $i \in S$ tiene período $d(i)$ si, y sólo si, $d(i) = \text{M.C.D}\{n \geq 1 \mid p_{ii}(n) > 0\}$. En caso de que $p_{ii}(n) = 0 \ \forall n \geq 1$, se conviene $d(i) = 0$.

Nota. Si en un camino aleatorio $p_i > 0, q_i > 0 \ \forall i \geq 1$, $p_0 > 0$ y $r_i = 0 \ \forall i$, cada estado tiene período 2. En cambio, si $\exists i_0$ tal que $r_{i_0} > 0$, todos los estados tienen período 1.

En efecto, en el primer caso:

$$P = \begin{Bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 1-p_1 & 0 & p_1 & 0 & 0 & \\ 0 & 1-p_2 & 0 & p_2 & 0 & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{Bmatrix}$$

Observemos que $i \begin{array}{l} \nearrow i-1 \\ \searrow i+1 \end{array}$ ($i \equiv$ estado inicial), $i+1 \begin{array}{l} \nearrow i \\ \searrow i+2 \end{array}$, $i-1 \begin{array}{l} \nearrow i \\ \searrow i-2 \end{array}$, luego $i \begin{array}{l} \nearrow i-1 \to i \\ \to i+1 \to i \\ \searrow i-1 \to i-2 \to i-1 \to i \\ \searrow i+1 \to i+2 \to i+1 \to i \end{array}$

en consecuencia $d(i) = 2$.

Por otra parte, si $r_{i_0} > 0$, $\left. i_0 \begin{array}{l} \nearrow i_0 - 1 \\ \to i_0 \\ \searrow i_0 + 1 \end{array} \right\} \to d(i_0) = 1$

Para cualquier otro estado $j \in S$ el sistema siempre puede alcanzar el estado i_0, permanecer en él tanto como se quiera y volver a j, de forma que el nº total de transiciones sea primo.

Teorema

Los estados comunicantes tienen el mismo período, i.e. si $k \leftrightarrow j$ entonces $d(k) = d(j)$

Demostración

Bastará con ver que si $j \to k$, entonces $d(j)$ divide a $d(k)$. Entonces, como $d(j)$ y $d(k)$ se dividirán mutuamente, tendremos que concluir que son iguales.

En primer lugar elegiremos M y N tales que $p_{jk}(M)\, p_{kj}(N) > 0$ (j y k son mutuamente accesibles).

Sea $n \geq 0$ tal que $p_{kk}(n) > 0$. Veamos que $p_{jj}(M+n+N) > 0$. Dado que por Chapman podemos poner:

$$p_{jj}(M+n+N) \geq p_{jk}(M)\, p_{kk}(n)\, p_{kj}(N) > 0$$

Análogamente: $p_{jj}(M+2n+N)>0$, dado que $p_{jj}(M+2n+N) \geq p_{jk}(M)\, p_{kk}(n)\, p_{kk}(n)\, p_{kj}(N)>0$
En consecuencia $d(j)$ dividirá a la vez a $M+2n+N$ y a $M+n+N$, y por lo tanto a su diferencia que es n. Como $d(j)$ divide a todo n tal que $p_{kk}(n)>0$, dividirá también al MCD de todos ellos, i.e., a $d(k)$. ◆

Definición

Una cadena de Markov en la cual todo estado es de período 1 se llama aperiódica.

2. TIEMPOS DE PRIMER PASO

Definición

Sea $f_{ij}(n)$ la probabilidad (condicional) de que partiendo de i se encuentre por primera vez j después de n transiciones:

$$f_{ij}(n) = P\{X_n=j,\ X_\nu\neq j\quad \nu=1,\dots,n-1 \,/\, X_0=i\}$$

Se llama tiempo de primer paso a la v.a. n que indica las transiciones que se han producido hasta alcanzar j.

Se tienen trivialmente las relaciones: $f_{ii}(0)=0$, $f_{ij}(0)=0$, $f_{ij}(1)=p_{ij}$

Lema

Para todo $i,j\in S$ y $\forall n\in N$: $p_{ij}(n)=\sum_{\nu=0}^{n} f_{ij}(\nu)\, p_{jj}(n-\nu)$ (1)

Demostración

$$p_{ij}(n) = P(X_n=j/X_0=i) = \sum_{\nu=1}^{n} P(X_n=j,\ X_\nu=j,\ X_m\neq j\quad m=1,2,\dots,\nu-1 \,/\, X_0=i) \overset{(*)}{=} (\text{prop. Markov})$$

$$= \sum_{\nu=1}^{n} P(X_n=j/X_\nu=j)\, P(X_\nu=j,\ X_m\neq j,\ m=1,2,\dots,\nu-1 \,/\, X_0=i) = \sum_{\nu=1}^{n} p_{jj}(n-\nu)\, f_{ij}(\nu) =$$

$$= \sum_{\nu=0}^{n} f_{ij}(\nu)\, p_{jj}(n-\nu), \text{ dado que } f_{ij}(0)=0.$$ ◆

Nota. Siempre que se utilice la identidad (1) se dirá que se ha usado el método de primera entrada.

La función generatriz de las probabilidades $\{p_{ij}(n)\}$ y $\{f_{ij}(n)\}$ son, respectivamente:

$$P_{ij}(s)=\sum_{n=0}^{\infty} p_{ij}(n)\, s^n \qquad \text{y} \qquad F_{ij}(s)=\sum_{n=0}^{\infty} f_{ij}(n)\, s^n \qquad \text{con } |s|<1$$

Teorema

Dada una cadena de Markov $\{X_n,\ n\in N\}$, las funciones generatrices $P_{ij}(s)$ y $F_{ij}(s)$ para $|s|<1$, están relacionadas por:

(1) $P_{ij}(s) = F_{ij}(s)\ P_{jj}(s)$ con $i \neq j$

(2) $P_{ii}(s) - 1 = F_{ii}(s)\ P_{ii}(s)$

Demostración

Sabemos que con $i \neq j$ $\forall n \in \mathbb{N}$: $p_{ij}(n) = \sum_{\nu=0}^{n} f_{ij}(\nu)\, p_{jj}(n-\nu)$ (cuando $n=0$: $p_{ij}(0) = 0 =$ $= f_{ij}(0)\, p_{jj}(0)$).

Por lo tanto $P_{ij}(s) = \sum_{n=0}^{\infty} s^n p_{ij}(n) = \sum_{n=0}^{\infty} s^n \left(\sum_{\nu=0}^{n} f_{ij}(\nu)\, p_{jj}(n-\nu) \right) =$ (cambiando el orden del sumatorio)

$$= \sum_{\nu=0}^{\infty} \sum_{n=\nu}^{\infty} f_{ij}(\nu)\, p_{jj}(n-\nu)\ s^n = \sum_{\nu=0}^{\infty} f_{ij}(\nu)\, s^\nu \sum_{n=\nu}^{\infty} p_{jj}(n-\nu)\, s^{n-\nu} =$$

$$= F_{ij}(s)\ P_{jj}(s)$$

Analogamente, $\forall n \in \mathbb{N}$: $p_{ii}(n) = \sum_{\nu=0}^{n} f_{ii}(\nu)\, p_{ii}(n-\nu) \overset{(f_{ii}(0)=0)}{=} \sum_{\nu=1}^{n} f_{ii}(\nu)\, p_{ii}(n-\nu)$ (cuando n $p_{ii}(0) = 1$, en cambio $f_{ii}(0) = 0$). Luego :

$$\sum_{n=1}^{\infty} s^n p_{ii}(n) = \sum_{n=1}^{\infty} s^n \sum_{\nu=1}^{n} f_{ii}(\nu)\, p_{ii}(n-\nu) = \sum_{\nu=1}^{\infty} \sum_{n=\nu}^{\infty} s^n f_{ii}(\nu)\, p_{ii}(n-\nu) =$$

$$= \sum_{\nu=1}^{\infty} f_{ii}(\nu)\, s^\nu \sum_{n=\nu}^{\infty} p_{ii}(n-\nu)\, s^{n-\nu} \underset{(f_{ii}(0)=0)}{=} F_{ii}(s)\ P_{ii}(s)$$

Como $p_{ii}(0) = 1$, se deduce que $P_{ii}(s) - 1 = F_{ii}(s)\ P_{ii}(s)$ ◆

3. ESTADOS RECURRENTES Y TRANSITORIOS

Definición

Sea f_{ij}^* la probabilidad de que partiendo del estado i, el sistema entre en el estado j alguna vez : $f_{ij}^* = \sum_{n=1}^{\infty} f_{ij}(n)$. Se dice que un estado $i \in S$ es recurrente si, y solo si, $f_{ii}^* = 1$, y que es transitorio si, y solo si, $f_{ii}^* < 1$.

Vamos a establecer un resultado que caracteriza la recurrencia de un estado en función de las probabilidades de transición $p_{ii}(n)$, para ello necesitaremos el siguiente lema.

Lema de ABEL

a) Supongamos que $\sum c_n$ converge y llamemos $f(x) = \sum_{n=0}^{\infty} c_n x^n$, $-1 < x < 1$. Entonces:

$$\lim_{x \to 1^-} f(x) = \sum c_n$$

b) Si $c_n \geq 0$ y $\lim_{s \to 1^-} \sum_{n=0}^{\infty} c_n s^n = c \leq \infty$, entonces $\sum c_n = c$.

Demostración

(a) Sea $S_n = c_0 + c_1 + \cdots + c_n$ y $S_{-1} = 0$. Entonces:

$$\sum_{n=0}^{m} c_n x^n = \sum_{n=0}^{m} (S_n - S_{n-1}) x^n = \sum_{n=0}^{m} (S_n x^n - S_{n-1} x^n) = (1-x) \sum_{n=0}^{m-1} S_n x^n + S_m x^m$$

para $|x|<1$, hagamos $m \to \infty$ y obtendremos: $f(x) = (1-x) \sum_{n=0}^{\infty} S_n x^n = \sum_{n=0}^{\infty} c_n x^n$ (2)

Sea $S = \lim_{n\to\infty} S_n$. Para $\varepsilon > 0$ $\exists N$ / $n > N$: $|S - S_n| < \varepsilon/2$

Como $(1-x) \sum_{n=0}^{\infty} x^n = 1$ para $|x|<1$, a partir de (2) obtendremos:

$$|f(x) - S| = |(1-x)\sum_{n=0}^{\infty} (S_n - S) x^n| \leq (1-x) \sum_{n=0}^{N} |S_n - S|\,|x|^n + \frac{\varepsilon}{2} < \varepsilon \quad (*)$$

tomando $1-x$ suficientemente pequeño.

(b) Al ser $c_n \geq 0$, tenemos que $\sum_{n=0}^{\infty} c_n s^n \leq \sum_{n=0}^{\infty} c_n$ para $0 \leq s \leq 1$ (3)

El caso $c=\infty$, es pues evidente. Si $c < \infty$, $\forall N$: $\sum_{n=0}^{N} c_n s^n \leq \sum_{n=0}^{\infty} c_n s^n$ luego:

$$\lim_{s\to 1^-} \sum_{n=0}^{N} c_n s^n = \sum_{n=0}^{N} c_n \leq \lim_{s\to 1^-} \sum_{n=0}^{\infty} c_n s^n = c. \quad 0 \leq s \leq 1$$

Por lo tanto $\{\sum_{n=0}^{N} c_n, N = 1,2,\dots\}$ es una sucesión monótona creciente, acotada por c, y su límite no puede ser menor que c, por (3). ◆ Nota (back)

Teorema

Un estado $i \in S$ es recurrente si, y sólo si, $\sum_{n=0}^{\infty} p_{ii}(n) = \infty$

Demostración

i es recurrente si, y sólo si, $f_{ii}^* = \sum_{n=0}^{+\infty} f_{ii}(n) = 1 = \lim_{s\to 1^-} F_{ii}(s)$, dado que $F_{ii}(1) = \sum_{n=0}^{\infty} f_{ii}(n) = f_{ii}^* = 1$ es convergente, podemos aplicar el lema de Abel (a) y resulta que: $\lim_{s\to 1^-} F_{ii}(s) = \sum_{n=0}^{\infty} f_{ii}(n)$.

Por otra parte, y puesto que $P_{ii}(s) - 1 = F_{ii}(s) P_{ii}(s)$, resultará que:

$$\lim_{s\to 1^-} F_{ii}(s) = 1 \iff \lim_{s\to 1^-} P_{ii}(s) = +\infty$$

y por lema de Abel (b) : $\lim_{s\to 1^-} P_{ii}(s) = \sum_{n=0}^{\infty} p_{ii}(n) = +\infty$ ◆

Teorema

Un estado $i \in S$ es recurrente si, y sólo si, $\sum_{n=0}^{\infty} p_{ji}(n) = \infty$ $\forall j \in S$ tal que i es accesible desde j ($j \to i$).

Demostración

Si $i \in S$ es accesible desde j $\exists N$ tal que $p_{ji}(N) > 0$, luego $\exists n$: $f_{ji}(n) > 0$. En consecuencia: $\sum_{n=0}^{\infty} f_{ji}(n) > 0$, pero por el lema de Abel (a) : $\lim_{s\to 1^-} F_{ji}(s) = \sum_{n=0}^{\infty} f_{ji}(n)$

Ahora bien, si $\sum f_{ji}(n) > 0$ $\exists n$ tal que $f_{ji}(n) > 0$.

Por el teorema anterior, $i \in S$ es recurrente si, y sólo si, $\lim_{s\to 1^-} P_{ii}(s) = +\infty$. pero dado que:

$$P_{ji}(s) = F_{ji}(s)\,[\,1 + F_{ii}(s)\,P_{ii}(s)\,]$$

Esto se cumple si y sólo si,

$$\lim_{s\to 1^-} P_{ji}(s) = \underbrace{\left(\lim_{s\to 1^-} F_{ji}(s)\right)}_{>0}\left[1+\underbrace{\left(\lim_{s\to 1^-} F_{ii}(s)\right)}_{=1}\underbrace{\left(\lim_{s\to 1^-} P_{ii}(s)\right)}_{+\infty}\right] = +\infty$$

pues, por el lema de Abel (b):

$$\lim_{s\to 1^-} P_{ji}(s) = \sum_{n=0}^{\infty} p_{ji}(n) = +\infty \quad . \quad ◆$$

Teorema

Si i y j son estados comunicantes, e i es un estado recurrente entonces j también lo será.

Demostración

$i \leftrightarrow j$ si y sólo si, $\exists\, m,n \geq 1$ tales que $p_{ij}(n)>0$ y $p_{ji}(m)>0$.

Sea $\nu \geq 0$. Por Chapman-Kolmogorov se deduce que:

$$p_{jj}(m+\nu+n) \geq p_{ji}(m)\, p_{ii}(\nu)\, p_{ij}(n)$$

y por tanto:

$$\sum_{\nu=0}^{\infty} p_{jj}(m+\nu+n) \geq p_{ji}(m)\, p_{ij}(n) \sum_{\nu=0}^{\infty} p_{ii}(\nu) = +\infty \quad (i \text{ es recurrente})$$

luego $\sum_{\eta=0}^{\infty} p_{jj}(\eta) = +\infty$, y j es un estado recurrente. ◆

Comentario. Es decir, la recurrencia, al igual que la periodicidad, es una propiedad de clase comunicante.

Definición

Sea g_{ij} la probabilidad de que partiendo de i la cadena haga infinitas visitas al estado j.

Teorema

El estado $i \in S$ es recurrente o transitorio según que g_{ii} sea, respectivamente, 1 ó 0.

Demostración.

Sea $g_{ii}(n) = P\{$partiendo el sistema del estado i vuelva a i al menos n veces$\}$. Utilizando el método de primera entrada se prueba:

$$g_{ii}(n) = \sum_{\nu=1}^{\infty} f_{ii}(\nu)\, g_{ii}(n-1) = g_{ii}(n-1)\, f_{ii}^*$$

por recurrencia: $g_{ii}(n) = (f_{ii}^*)^{n-1} g_{ii}(1) = (f_{ii}^*)^n \qquad (g_{ii}(1) = f_{ii}^*)$

Por otra parte: $g_{ii} = \lim_{n\to\infty} g_{ii}(n)$, y en consecuencia:

$$g_{ii} = \begin{cases} 1 \\ 0 \end{cases} \text{según sea } f_{ii}^* = \begin{cases} 1 \\ <1 \end{cases} \qquad ◆$$

Teorema

Si j es accesible desde i, $i \to j$, siendo i recurrente, entonces $f_{ji}^* = 1$ y $j \to i$.

Demostración

Por un teorema anterior tenemos que: (por ser i recurrente)

$$g_{ii} = 1 = \sum_{k \in S} p_{ik}(\nu)\, g_{ki} \qquad \forall \nu \geq 0$$

es decir: $0 = \sum_{k \in S} p_{ik}(\nu)(1 - g_{ki}) \Rightarrow p_{ik}(\nu)(1 - g_{ki}) = 0 \quad \forall \nu \geq 0 \quad \forall k \in S$

Por otra parte: $i \to j$ y $\exists N : p_{ij}(N) > 0 \Rightarrow g_{ji} = 1$ $(k = j,\ \nu = N)$

Además: $g_{ji}(n) = \sum_{\nu=1}^{\infty} f_{ji}(\nu)\, g_{ii}(n-1) = g_{ii}(n-1)\, f_{ji}^*$. Tomando límites, $n \to \infty$:

$g_{ji} = g_{ii}\, f_{ji}^*$ (4) luego $f_{ji}^* = 1$. Consecuentemente $j \to i$, siendo i y j estados comunicantes. ◆

Comentario. Los únicos estados accesibles desde un estado recurrente son aquellos estados que se comunican con él.

Corolario

Las clases comunicantes recurrentes son cerradas

Demostración

Evidente en virtud del teorema anterior. ◆

Nota. Por otra parte, si i y j pertenecen a una clase comunicante y recurrente, se cumplirá:

$$f_{ij}^* = f_{ji}^* = 1 \qquad (5)$$

Veamos un recíproco de la proposición anterior. ¡criterio de recurrencia!

Teorema

Sea $j \in \mathscr{C} \equiv$ clase comunicante cerrada.

$\mathscr{C}$ es recurrente si, y sólo si, $\forall i \in \mathscr{C},\ i \neq j : f_{ij}^* = 1$

Demostración

CN. Evidente, por el teorema anterior.

CS. Vamos a probar que $f_{jj}^* = 1$. Se tiene que:

$n = 1 \longrightarrow f_{jj}(1) = p_{jj}$

$n \geq 2 \longrightarrow f_{jj}(n) = \sum_{\substack{i \in \mathscr{C} \\ i \neq j}} p_{ji}\, f_{ij}(n-1)$, por ser $\mathscr{C}$ cerrada.

Sumando: $f_{jj}^* \equiv \sum_{n=1}^{\infty} f_{jj}(n) = p_{jj} + \sum_{n=2}^{\infty} \sum_{\substack{i \in \mathscr{C} \\ i \neq j}} p_{ji}\, f_{ij}(n-1) = p_{jj} + \sum_{\substack{i \in \mathscr{C} \\ i \neq j}} p_{ji} \sum_{n=2}^{\infty} f_{ij}(n-1) =$

$$= p_{jj} + \sum_{\substack{i \in \mathscr{C} \\ i \neq j}} p_{ji}\, \underbrace{f_{ij}^*}_{=1} = \sum_{i \in \mathscr{C}} p_{ji} = 1$$

◆

Teorema

Sea $\mathcal{C}$ = clase transitoria cerrada $\longrightarrow$ $\mathcal{C}$ tiene infinitos estados

Demostración

Por (4), y al ser $i \in \mathcal{C}$ transitorio, $g_{ii} = 0 \longrightarrow g_{ji} = 0$, es decir, i se visita un número finito de veces partiendo de cualquier estado j, con probabilidad 1. Y con probabilidad 1 la cadena abandonaría una clase $\mathcal{C}$ transitoria finita, luego la clase comunicante transitoria cerrada debe poseer infinitos estados. ◆

Comentario

De todos los resultados anteriores se deduce que si S es el conjunto de estados de una cadena de Markov $S = C_1 \cup C_2 \cup \dots \cup C'_1 \cup C'_2 \cup \dots \cup C''_1 \cup C''_2 \cup \dots$ siendo
C_r = clase recurrente cerrada, C'_r = cl. transitoria cerrada y C''_r = cl. transitoria no cerrada.

En una cadena de Markov finita no hay clases del tipo C'_r.

Como una clase comunicante cerrada se puede extraer y tratar por sí misma como una cadena de Markov (Teor. pág 34), para estudiar el comportamiento asintótico de una cadena de Markov basta restringirnos a las clases comunicantes cerradas (recurrentes y transitorias).

Ordenación típica de los estados de una cadena de Markov

$$
\begin{array}{cc|cccc|c|}
 & & \multicolumn{4}{c|}{\text{Recurrentes}} & \text{Transitorios} \\
 & & C_1 & C_2 & \cdots & C_r & \\
\hline
\text{Recurrentes} & C_1 & \boxed{P_1} & 0 & \cdots & 0 & \\
 & C_2 & 0 & \boxed{P_2} & \cdots & 0 & 0 \\
 & \vdots & \vdots & \vdots & & \vdots & \\
 & C_r & 0 & 0 & \cdots & \boxed{P_r} & \\
\hline
\text{Transitorios} & & \multicolumn{4}{c|}{A} & B \\
\hline
\end{array}
$$

Si el espacio de estados S es finito, entonces $A \neq \emptyset$.

Teorema

Sea $\mathcal{C}$ una clase comunicante cerrada $\mathcal{C}$, con periodo t.

Entonces $\mathcal{C}$ puede dividirse de manera única en t subconjuntos $\{D_0, D_1, \dots, D_{t-1}\}$ de tal manera que la evolución se produce cíclicamente pasando en cada etapa de D_r a D_{r+1} (en particular de D_{t-1} a D_0).

$$
\begin{array}{c|ccccccc|}
 & D_0 & D_1 & D_2 & \cdots & D_{t-2} & D_{t-1} \\
\hline
D_0 & 0 & \boxed{A_1} & 0 & \cdots & 0 & 0 \\
D_1 & 0 & 0 & \boxed{A_2} & \cdots & 0 & 0 \\
\vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
D_{t-2} & 0 & 0 & 0 & \cdots & 0 & \boxed{A_{t-1}} \\
D_{t-1} & \boxed{A_0} & 0 & 0 & \cdots & 0 & 0
\end{array}
$$

Demostración.

Fijemos un estado $i \in \mathcal{C}$, cualquiera, y definamos:

$D_0 = \{ j \in \mathcal{C} : \exists k \text{ tal que } p_{ji}(kt) > 0 \}$

$D_1 = \{ j \in \mathcal{C} : \exists k \text{ tal que } p_{ji}(kt-1) > 0 \}$

$\vdots$

$D_{t-1} = \{ j \in \mathcal{C} : \exists k \text{ tal que } p_{ji}(kt-t+1) > 0 \}$

Puesto que todos los estados intercomunican, es claro que cada estado $j \in \mathcal{C}$ pertenece por lo menos a uno de estos subconjuntos (en particular $i \in D_0$). $\bigcup_{\ell=1}^{t} D_{t-\ell} = \mathcal{C}$

Veamos que estos conjuntos son disjuntos. Efectivamente, si puede $j \in D_r \cap D_s$, existirían k y k' tales que:

$p_{ji}(kt-r) > 0$ y $p_{ji}(k't-s) > 0$

Ahora bien, como $i \to j$, existirá k_1 tal que $p_{ij}(k_1 t + s') > 0$. Donde luego, ha de cumplirse que $s = s'$, ya que por Chapman-Kolmogorov: $p_{ii}((k'+k_1)t + s' - s) > 0$, y si $s \neq$ el periodo de i no sería t.

Por lo tanto: $p_{jj}((k+k_1)t + s - r) > 0$ y ha de cumplirse $s = r$.

Finalmente, si $j \in D_r$ y $p_{jj'} > 0$. Sea $j' \in D_s$, es decir, $\exists k$ tal que $p_{j'i}(kt-s) >$ 0. Entonces $p_{ji}(kt-s+1) > p_{jj'} \, p_{j'i}(kt-s) > 0$, luego $j \in D_{s-1}$, i.e. $s-1 = r$ y $s = r+1$. ■

4. PROBABILIDADES DE PRIMER PASO Y ABSORCIÓN

Vamos a demostrar que en el caso de que $j \in S$ sea transitorio y $k \in S$ recurrente la probabilidad de primer paso se puede determinar como la solución de un sistema de ecuaciones lineales que involucran a todas estas probabilidades de primer paso.

Dada una cadena de Markov, llamaremos T al conjunto de estados transitorios.

Teorema

Sea $k \in S - T$ (estado recurrente), las probabilidades de primer paso $\{ f^*_{jk}, j \in T \}$ satisfacen:

$$f^*_{jk} = \sum_{i \in T} p_{ji} f^*_{ik} + \sum_{i \in C} p_{ji} \qquad \forall j \in T \qquad (6)$$

siendo $C \overset{(*)}{=} C(k)$.

Demostración.

$f_{jk}(n) = \sum_{\substack{i \in S \\ i \neq k}} p_{ji} f_{ik}(n-1)$ con $n > 1$ y $f_{jk}(1) = p_{jk}$. Entonces:

$$f^*_{jk} = \sum_{n=1}^{\infty} f_{jk}(n) \underset{(**)}{=} p_{jk} + \sum_{\substack{i \in S \\ i \neq k}} p_{ji} f^*_{ik} = \sum_{i \in S} p_{ji} f^*_{ik} \qquad (\text{puesto que } f^*_{kk} = 1)$$

Ahora bien, si $i \in C = C(k)$ entonces $f^*_{ik} = 1$ (por (5)). Si $i \notin C$ e $i \notin T$ $^{(*)}$, entonces $f^*_{ik} = 0$ (puesto que k no puede ser accesible desde i, en otro caso estarían en la misma clase (Teor. pág. 40)).

Sustituyendo:

$$f^*_{jk} = \sum_{i \in T} p_{ji} f^*_{ik} + \sum_{i \in C} p_{ji}$$ ◆

Definición

Sea j un estado transitorio, $j \in T$, si $\mathcal{C}$ es una clase recurrente se define $f^*_{j\mathcal{C}}$ como la probabilidad de que la cadena de Markov sea "absorbida" por la clase $\mathcal{C}$.

Evidentemente, una vez entra la cadena en $\mathcal{C}$, por ser cerrada, ya no saldrá.

Teorema

Sea $\mathcal{C}$ una clase recurrente. Las probabilidades de absorción $\{f^*_{j\mathcal{C}}, j \in T\}$ verifican

$$f^*_{j\mathcal{C}} = \sum_{i \in T} p_{ji} f^*_{i\mathcal{C}} + \sum_{i \in \mathcal{C}} p_{ji} \qquad \forall j \in T \qquad (7)$$

Demostración

Definimos $f_{j\mathcal{C}}(n) = P(X_n \in \mathcal{C}, X_\nu \notin \mathcal{C}, \nu = 1, 2, \dots, n-1 / X_0 = j)$, para $j \in T$. Entonces $f^*_{j\mathcal{C}} = \sum_{n=1}^{\infty} f_{j\mathcal{C}}(n) \leq 1$.

Ahora bien: $f_{j\mathcal{C}}(1) = \sum_{i \in \mathcal{C}} p_{ji}$

$n > 1$: $f_{j\mathcal{C}}(n) = \sum_{i \in T} p_{ji} f_{i\mathcal{C}}(n-1)$, porque si i pertenece a otra clase recurrente que no sea $\mathcal{C}$, entonces $f_{i\mathcal{C}}(n-1) =$

Sumando:

$$f^*_{j\mathcal{C}} = \sum_{i \in \mathcal{C}} p_{ji} + \sum_{i \in T} p_{ji} f^*_{i\mathcal{C}}$$ ◆

Nota. Para que (6) sea una fórmula útil para hallar la probabilidad de primer paso se ha de demostrar que el sistema (6) tiene una solución única con valores en $[0,1]$. Este sistema también se puede escribir con: $v_j = f^*_{jk}$ y $b_j = \sum_{i \in \mathcal{C}} p_{ji}$

$$v_j = \sum_{i \in T} p_{ji} v_i + b_j \qquad \forall j \in T \qquad (8)$$

Capítulo 5

Comportamiento límite de una cadena de Markov

TEMA 5: COMPORTAMIENTO LÍMITE DE UNA CADENA DE MARKOV

1. Distribución límite y distribución estacionaria

Definición

Sea X_n una cadena de Markov cuyo espacio de estados es S, con matriz de transición P. Sea $\pi = (\pi_i)_{i\in S}$ una distribución de probabilidad, si cumple que $\pi = \pi P$, es decir, $\pi_j = \sum_{k\in S} \pi_k p_{kj}$ $\forall j\in S$, entonces π es una distribución estacionaria.

Dado que $p(n) = p(n-1)P$, si se tiene que $p(0)=\pi$, entonces $p(1)=\pi$, $p(2)=\pi$ y así sucesivamente: $p(n)=\pi$ $\forall n$.

Recordemos del tema anterior que:

$f_{jk}(n) = P\{X_r \neq k,\ r=1,\dots,n-1;\ X_n = k / X_0 = j\}$

$f^*_{jk} = \sum_{n=1}^{\infty} f_{jk}(n)$ es la probabilidad de que el estado k sea eventualmente alcanzado.

Un estado $i\in S$ es recurrente si, y sólo si, $f^*_{ii}=1$ y transitorio si, y sólo si, $f^*_{ii}<1$. Para un estado i recurrente $\{f_{ii}(n),\ n=1,2,\dots\}$ es una distribución de probabilidad y la media de esta distribución

$$\mu_i = \sum_{n=1}^{\infty} n\, f_{ii}(n)$$

es el tiempo medio de recurrencia. Se tiene que $\mu_i = F'_{ii}(1)$.

Definición

Según que μ_i sea finito o infinito se dice que el estado i es recurrente positivo o recurrente nulo, respectivamente.

Por otra parte, si $f^*_{jk}=1$, $\sum_{n=1}^{\infty} n\, f_{jk}(n)$ es el tiempo medio de primer paso de j a k.

Definición

Un estado recurrente positivo y aperiódico se dice que es ergódico.

Comentario: cadenas finitas irreducibles

Sea una cadena de Markov irreducible (existe una sola clase comunicante), en este caso se puede hablar de cadena recurrente, de cadena aperiódica, etc. porque todas estas propiedades, según se ha probado en el tema anterior, son propiedades de clase comunicante. También se vió que es necesario que una cadena irreducible finita sea recurrente.

Vamos a estudiar la existencia de distribución estacionaria para cadenas irreducibles y recurrentes.

Nota.

Veremos que la recurrencia nula o positiva es una propiedad de clase comunicante. Para ello se establecerá una caracterización de la recurrencia nula o positiva basada en las funciones $F_{ii}(s)$ y $P_{ii}(s)$.

Recordemos que $F_{ii}(s) = \sum_{n=0}^{\infty} f_{ii}(n) s^n$, luego $\frac{d}{ds} F_{ii}(s)\Big|_{s=1} = \sum_{n=1}^{\infty} n f_{ii}(n) s^{n-1}\Big|_{s=1} = \mu_i$. Por otra parte:

$$\frac{d}{ds} F_{ii}(s)\Big|_{s=1} = \lim_{s \to 1^-} \frac{F_{ii}(1) - F_{ii}(s)}{1-s} = \lim_{s \to 1^-} \frac{1 - F_{ii}(s)}{1-s}$$

Así, pues:

(a) i es un estado recurrente positivo si, y sólo si, $\lim_{s \to 1^-} \frac{1-F_{ii}(s)}{1-s} < \infty$

(b) i es un estado recurrente nulo si, y sólo si, $F_{ii}'(1) = \infty$.

Pero, además, sabemos que $F_{ii}(s) = 1 - \frac{1}{P_{ii}(s)}$, en consecuencia $1 - F_{ii}(s) = \{P_{ii}(s)\}^{-1}$, luego:

(a) i es un estado recurrente positivo si, y sólo si, $\lim_{s \to 1^-} (1-s) P_{ii}(s) > 0$

(b) i es un estado recurrente nulo si, y sólo si, $\lim_{s \to 1^-} (1-s) P_{ii}(s) = 0$

Lema

La recurrencia positiva o nula es una propiedad de clase

Demostración

Supongamos que k es un estado recurrente positivo. Si $j \leftrightarrow k$, existen M y N tales que $p_{jk}(M) > 0$ y $p_{kj}(N) > 0$. Entonces, por Chapman-Kolmogorov, para $n > N+M$:

$$p_{kk}(n) \geq p_{kj}(N)\, p_{jj}(n-M-N)\, p_{jk}(M)$$

$$p_{jj}(n) \geq p_{jk}(M)\, p_{kk}(n-N-M)\, p_{kj}(N)$$

Además:

$$(1-s)\{P_{jj}(s)\} = (1-s)\Big\{1 + \sum_{n=1}^{\infty} p_{jj}(n) s^n\Big\} \geq (1-s)\Big\{1 + \sum_{n=1}^{N+M} p_{jj}(n) s^n + s^{N+M} p_{jk}(M)\, p_{kj}(N) \sum_{n=1}^{\infty} p_{kk}(n) s^n\Big\}$$

$$\geq (1-s)\{1 + s^{N+M} p_{jk}(M)\, p_{kj}(N)\,(P_{kk}(s) - 1)\} =$$

$$= (1-s)\{P_{kk}(s)\}\, s^{N+M} p_{jk}(M)\, p_{kj}(N) + (1-s)\{1 - s^{N+M} p_{jk}(M)\, p_{kj}(N)\}$$

luego

$$\lim_{s \to 1^-} (1-s)\{P_{jj}(s)\} \geq \lim_{s \to 1^-} (1-s)\{P_{kk}(s)\}\, p_{jk}(M)\, p_{kj}(N)$$

y como el segundo miembro es positivo, también lo será el primero. ◆

y ~~recurrente~~.

Teorema de Levinson (existencia de distribución estacionaria)

Para la cadena irreducible y recurrente P existe distribución estacionaria $\{\pi_j, j\in S\}$ si, y sólo si, la cadena es recurrente positiva. En tal caso, la única distribución estacionaria viene dada por:

$$\pi_j = \frac{1}{\mu_j} = \lim_{s\uparrow 1} (1-s)\, P_{jj}(s)$$

Demostración

CS Veamos que si la cadena es recurrente positiva, $\pi_j = \frac{1}{\mu_j}$ da lugar a una distribución estacionaria.

Desde luego, $\pi_j > 0 \quad \forall j\in S$ (los tiempos medios de recurrencia μ_j son finitos). Veamos ahora que $\sum_{j\in S} \pi_j \leq 1$. En efecto, para $s<1$:

$$(1-s)\sum_{j\in S} P_{ij}(s) = (1-s)\sum_{j\in S}\sum_{n=0}^{\infty} p_{ij}(n)\, s^n \overset{(*)}{=} (1-s)\sum_{n=0}^{\infty} s^n \underbrace{\sum_{j\in S} p_{ij}(n)}_{=1} = 1$$

Se sabe que:
$$\begin{cases} P_{ij}(s) = F_{ij}(s)\, P_{jj}(s) & j\neq i \\ P_{ii}(s) - 1 = F_{ii}(s)\, P_{ii}(s) \end{cases}$$

reemplazando:

$$(1-s)\Big\{\sum_{\substack{j\in S\\ j\neq i}} F_{ij}(s)\, P_{jj}(s) + F_{ii}(s)\, P_{ii}(s) + 1\Big\} = 1$$

o lo que es lo mismo: $(1-s)\sum_{j\in S} F_{ij}(s)\, P_{jj}(s) = s$

Tomando límites$^{(*)}$ y recordando que $\lim_{s\to 1^-} F_{ij}(s) = f_{ij}^* = 1$ (cadena recurrente):

$$\sum_{j\in S} \pi_j \leq 1$$

Veamos, además que $\pi = \pi P$, es decir: $\pi_j = \sum_{k\in S} \pi_k\, p_{kj}, \quad \forall j\in S$. En efecto:

$$P_{ij}(s) = p_{ij}(0) + s\sum_{n=1}^{\infty} p_{ij}(n)\, s^{n-1} = p_{ij}(0) + s\sum_{n=1}^{\infty}\Big(\sum_{k\in S} p_{ik}(n-1)\, p_{kj}\Big) s^{n-1} = p_{ij}(0) + s\sum_{k\in S} p_{kj}\sum_{n=1}^{\infty} p_{ik}(n-1)s^{n-1} =$$

$$= p_{ij}(0) + s\sum_{k\in S} p_{kj}\, P_{ik}(s)$$

luego $P_{ij}(s) - p_{ij}(0) = F_{ij}(s)\, P_{jj}(s) = s\Big\{\sum_{k\in S} p_{kj}\, F_{ik}(s)\, P_{kk}(s) + p_{ij}\Big\}$

De ahí que:

$$\frac{1-s}{s} F_{ij}(s)\, P_{jj}(s) = (1-s)\sum_{k\in S} p_{kj}\, F_{ik}(s)\, P_{kk}(s) + p_{ij}(1-s)$$

Cuando $s \uparrow 1$, tomando límite: $\pi_j \ge \sum_{k\in S} \pi_k p_{kj}$, y sumando para j:

$$\sum_{j\in S}\pi_j \ge \sum_{j\in S}\sum_{k\in S}\pi_k p_{kj} = \sum_{k\in S}\pi_k \underbrace{\sum_{j\in S} p_{kj}}_{1} = \sum_{k\in S}\pi_k$$

lo que nos obliga a que: $\pi_j = \sum_{k\in S}\pi_k p_{kj}$. Así que $\left\{ \pi_j \big/ \sum_{k\in S}\pi_k , j\in S\right\}$ es una distribución estacionaria.

CN Sea, ahora, la distribución estacionaria $\{v_j, j\in S\}$. Se cumple, naturalmente: $v_j \ge 0$, $\sum_{j\in S} v_j = 1$ y $v_j = \sum_{i\in S} v_i p_{ij} = \sum_{i\in S} v_i p_{ij}(n)$ $\forall n\ge 1$.

Tendremos, además:

$$(1-s)P_{jj}(s)\sum_{i\in S} v_i F_{ij}(s) = (1-s)\Big\{\sum_{i\in S} v_i P_{ij}(s) - v_j\Big\} = (1-s)\Big\{\sum_{i\in S} v_i \sum_{n=0}^{\infty} p_{ij}(n)s^n - v_j\Big\} =$$

$$= (1-s)\Big\{\sum_{n=0}^{\infty} s^n \underbrace{\sum_{i\in S} v_i p_{ij}(n)}_{v_j} - v_j\Big\} = (1-s)\sum_{n=1}^{\infty} s^n v_j = s v_j$$

tomando límite para $s\uparrow 1$: $\pi_j \lim_{s\uparrow 1}\sum_{i\in S} v_i F_{ij}(s) = v_j$.

Se tiene que: $F_{ij}(s)\le 1$ $\forall i$ y $\sum_{i\in S} v_i F_{ij}(s) \le \sum_{i\in S} v_i = 1$. Luego $\pi_j \ge v_j$, $\forall j\in S$. En definitiva, como alguna v_j tendrá que ser estrictamente positiva, la correspondiente π_j también lo será, siendo j recurrente positivo.

El razonamiento anterior prueba que si existe distribución estacionaria la cadena es recurrente positiva. Aplicando la conclusión obtenida en la primera parte de esta demostración, $\{\pi_j, j\in S\}$ satisface: $\sum_{j\in S}\pi_j \le 1$ y $\pi_j = \sum_{i\in S}\pi_i p_{ij}$, $j\in S$. Pero, como: $\pi_j \ge v_j$ $\forall j\in S$, ha de cumplirse $\sum_{j\in S}\pi_j = 1$, con lo que la distribución estacionaria única es $\{\pi_j, j\in S\}$.

Nota. Disponiendo del resultado anterior es fácil expresar para una cadena cualquiera la condición para que exista distribución estacionaria, aplicando el teorema de descomposición.

Teorema

Sea P una cadena de Markov arbitraria. Existe distribución estacionaria $\{\pi_j, j\in S\}$ si, y sólo si, existe alguna clase recurrente positiva.

Demostración

CS Sea, por ejemplo $C_1 \subset S$ recurrente positiva. Por el T. de Levinson hay distribución estacionaria respecto de P_1: $\pi_{C_1} = \{\pi_j, j\in C_1\}$

Evidentemente la distribución $\tilde{\pi} := \begin{cases} \tilde{\pi}_j = \pi_j & \text{si } j \in C_1 \\ \tilde{\pi}_j = 0 & \text{si } j \notin C_1 \end{cases}$, es estacionaria respecto de P

$$\tilde{\pi} P = (\pi_{C_1}, \vec{0})\, P = (\pi_{C_1} P_1 \ \vec{0}) = (\pi_{C_1}, \vec{0}) = \tilde{\pi}$$

c.r. Supongamos que no hay subcadena recurrente positiva. Consecuentemente no habrá distribución estacionaria respecto de $P_1 \dots P_r$.

Veamos ahora que si la cadena tiene estados transitorios T, ninguna distribución estacionaria puede ser tal que: $\sum_{j \in T} \pi_j > 0$.

Observemos que $\forall i \in S \sim T$ y $\forall j \in T$, $p_{ij} = 0$. Por consiguiente: $\pi_j = \sum_{i \in T} \pi_i p_{ij} + \sum_{i \in S \sim T} \pi_i p_{ij} = \sum_{i \in T} \pi_i p_{ij}$. Asimismo: $\pi_j = \sum_{i \in T} \pi_i p_{ij}(n)$, $j \in T$, $n \geq 0$.

Sumemos para n: $\sum_{n \geq 0} \pi_j = \sum_{n \geq 0} \sum_{i \in T} \pi_i p_{ij}(n) = \sum_{i \in T} \pi_i \sum_{n \geq 0} p_{ij}(n) = \sum_{i \in T} \pi_i \{ \lim_{s \uparrow 1} F_{ij}(s) P_{jj}(s) + p_{ij}(0) \}$

Si $\pi_j > 0$, para $j \in T$, el primer miembro es divergente mientras que el segundo es convergente.

Por otra parte, si π es una distribución estacionaria respecto de P, se cumple: $\pi_T = 0_{\text{card}\, T}$. Por lo tanto, la distribución estacionaria π se puede descomponer:

$$\pi = (\pi_{C_1}, \pi_{C_2} \dots \pi_{C_r}, 0_{\text{card}\, T})$$

cumpliendose $\pi_{C_i} P_i = \pi_{C_i}$. Sea $\pi_{C_{i_0}} \neq 0_{\text{card}\, C_{i_0}}$. dividiendo por la suma (no nula) de las componentes de $\pi_{C_{i_0}}$ resultaría una distribución estacionaria para P_{i_0}, y esto es imposible.

◆

Se ha visto que la existencia de distribución estacionaria está ligada a la propiedad de recurrencia positiva para la cadena de Markov. Veamos un caso particular de cadenas que tienen siempre esta propiedad.

Teorema

Una cadena de Markov irreducible con un número finito de estados es recurrente positiva

Demostración.

La recurrencia viene dada por la finitud.

Por otra parte, se ha visto que se verifica la identidad

$$(1-s) \sum_{j \in S} F_{ij}(s) P_{jj}(s) = s \qquad s \in [0,1)$$

Tomando límites para $s \uparrow 1$: (siendo card $S < \infty$)

$$\sum_{j \in S} \lim_{s \to 1^-} F_{ij}(s) \lim_{s \to 1^-} (1-s) P_{jj}(s) = \sum_{j \in S} \underbrace{f^*_{ij}}_{1} \frac{1}{\mu_j} = \sum_{j \in S} \frac{1}{\mu_j} = 1$$

Luego algún μ_j ha de ser finito, y por lo tanto todos lo son, al ser la recurrencia positiva una propiedad de clase.

◆

2. Teorema límite

Ya se ha visto cuáles son las cadenas de Markov que tienen distribución estacionaria y en qué casos esta distribución es única. Si existen en una cadena varias subcadenas recurrentes positivas cada una tiene su distribución estacionaria única, y la cadena cuenta de tantas distribuciones estacionarias como combinaciones lineales convexas de dichas distribuciones parciales.

Estudiaremos ahora en qué condiciones una distribución estacionaria π verifica que para $i \in S$,

$$\lim_{n\to\infty} p_{ji}(n) = \pi_i \quad , \quad \forall j \in S$$

En este caso, sea cual sea la distribución inicial de la cadena, la distribución de X_n tiende a π cuando $n \to \infty$, por lo que se dice que π es la distribución límite.

Veamos un lema previo, que es un resultado sobre divisibilidad.

Lema

Sea I un conjunto no vacío de enteros positivos tal que:

(i) $M.C.D.\{I\} = 1$

(ii) si $m, n \in I \longrightarrow m+n \in I$ (incluso para $n=m$)

Entonces existe un n_0 tal que $n \in I$ $\forall n \geq n_0$.

Demostración

Probaremos, en primer lugar, que I contiene dos enteros consecutivos. Si no sucede tal cosa, entonces existirá un entero $k \geq 2$ y un $n_1 \in I$ tal que $n_1 + k \in I$, y cualquier par de enteros distintos en I difieren por lo menos en k.

Se sigue de (i) que existe un $n \in I$ tal que k no es divisor de n. Podemos entonces escribir que: $n = mk + r$, para un entero no-negativo y $0 < r < k$.

A partir de (ii) se sigue que:

$(m+1)(n_1+k) \in I$ y $n + (m+1)n_1 \in I \rightarrow (m+1)(n_1+k) - n - (m+1)n_1 = k + mk - n = k - r$

la diferencia es un positivo menor que k. Esto contradice la definición de k.

Hemos probado que I contiene dos enteros consecutivos, a saber, n_1 y n_1+1. Sea $n \geq n_1^2$. Existirán m y r enteros no-negativos tales que

$$0 \leq r < n_1 \quad y \quad n - n_1^2 = mn_1 + r.$$

Así, pues, $n = r(n_1+1) + (n_1 - r + m)n_1 \in I$ (por (ii)). Luego $n \in I$ $\forall n \geq n_0 = n_1^2$. ◆

Teorema de convergencia (caso aperiódico)

Sea $\{X_n, n \geq 0\}$ una cadena de Markov irreducible, recurrente positiva y aperiódica. Sea $\pi = \{\pi_i, i \in S\}$ su distribución estacionaria.

probabilidad 1.

Independientemente de la distribución del estado inicial (X_0, Y_0) se tiene que

$P\{X_n=j \,/\, T\le n\} = P\{Y_n=j \,/\, T\le n\} \quad \forall j\in S$

puesto que las dos cadenas X_n e Y_n son indistinguibles, con $n\ge T$. Multiplicando por $P\{T\le n\}$ se deduce:

$$P\{X_n=j,\, T\le n\} = P\{Y_n=j,\, T\le n\} \qquad \forall j\in S \qquad (3)$$

La ecuación (3) implica:

$$P\{X_n=j\} = P\{X_n=j,\, T\le n\} + P\{X_n=j,\, T>n\} \underset{(3)}{=} P\{Y_n=j,\, T\le n\} + P\{X_n=j,\, T>n\} \le$$
$$\le P\{Y_n=j\} + P\{T>n\}$$

Analogamente: $P\{Y_n=j\} \le P\{X_n=j\} + P\{T>n\}$. En consecuencia:

$$|P\{X_n=j\} - P\{Y_n=j\}| \le P\{T>n\}$$

Como T es finito con probabilidad 1: $\lim_{n\to\infty} P\{T>n\} = P\{T=\infty\}=0$. Concluimos que

$$\lim_{n\to\infty} \{P\{X_n=j\} - P\{Y_n=j\}\} = 0 \qquad \forall j\in S \qquad (4)$$

Utilizaremos el resultado (4) para completar la demostración del teorema. Sea $i\in S$, concreto. Tomaremos una distribución inicial de (X_n, Y_n) tal que $P\{X_0=i\}=1$ y $P\{Y_0=j_0\} \underset{(*)}{=} \pi_{j_0}$. Resultará entonces:

$$P\{X_n=j\} = p_{ij}(n) \qquad j\in S$$
$$P\{Y_n=j\} = \pi_j \qquad j\in S$$

De estas dos relaciones se deduce: $\lim_{n\to\infty}\{p_{ij}(n) - \pi_j\} = \lim_{n\to\infty}\{P\{X_n=j\} - P\{Y_n=j\}\} = 0$ ◆

Comenzaremos por considerar una ligera extensión de este resultado de convergencia (caso aperiódico), que nos ayudará en la demostración de la convergencia para el caso periódico.

Lema 1

Sea $\{X_n, n\ge 0\}$ una cadena de Markov con espacio de estados S. Sea $j\in S$ un estado recurrente positivo ($\mu_j<+\infty$) y aperiódico. Entonces $\lim_{n\to\infty} p_{jj}(n) = \frac{1}{\mu_j}$

Demostración

Consideremos la clase $\mathcal{C}(j)$ = clase comunicante correspondiente a j. $\mathcal{C}(j)$ es recurrente positiva y aperiódica, además de cerrada (por ser recurrente).

Sea π la distribución estacionaria única con toda su masa probabilística concentrada en $\mathcal{C}(j)$. Sumergiéndonos en la subcadena $\mathcal{C}(j)$: $\lim_{n\to\infty} p_{jj}(n) = \pi_j = \frac{1}{\mu_j}$ ◆

Lema 2

Sea $\{X_n, n\geq 0\}$ una cadena de Markov con espacio de estados S. Sea $j\in S$ un estado recurrente positivo y aperiódico. Entonces

$$\lim_{n\to\infty} p_{ij}(n) = \frac{f^*_{ij}}{\mu_j}$$

Demostración

Sabemos que: $p_{ij}(n) = \sum_{m=1}^{n} f_{ij}(m)\, p_{jj}(n-m) = \sum_{m=1}^{\infty} f_{ij}(m)\, a_m(n)$

siendo $a_m(n) = \begin{cases} 0 & \text{si } m>n \\ p_{jj}(n-m) & \text{si } m\leq n \end{cases}$

Tenemos que: (Lema 1)

$$\lim_{n\to\infty} f_{ij}(m)\, a_m(n) = \frac{f_{ij}(m)}{\mu_j} \qquad \forall m$$

Por otra parte:

$$f_{ij}(m)\, a_m(n) \leq f_{ij}(m) \quad \forall n \qquad \text{y} \qquad \sum_{m=1}^{\infty} f_{ij}(m) = f^*_{ij}$$

Aplicando el Teorema de la Convergencia Dominada:

$$\lim_{n\to\infty} p_{ij}(n) = \frac{1}{\mu_j}\sum_{m=1}^{\infty} f_{ij}(m) = \frac{f^*_{ij}}{\mu_j}$$

◆

Nota. Los resultados establecidos en los Lemas 1 y 2 para estados recurrentes positivos y aperiódicos, son también ciertos para estados recurrentes y aperiódicos, aunque no los demostraremos.

Lema 3

Sea $\{X_n, n\geq 0\}$ una c. de Markov cuyo espacio de estados es S. Sea $j\in S$ un estado transitorio. Entonces $\lim_{n\to\infty} p_{ij}(n) = 0$

Demostración.

En realidad es evidente, puesto que:

a) $p_{ij}(n) = 0 \quad \forall n$ si j no es accesible desde i, ó

b) $\sum_{n=0}^{\infty} p_{ij}(n) < +\infty \quad \forall i\in S$ tal que j es accesible desde i (teorema pag. 38). El estado j sólo puede ser visitado un número finito de veces.

En consecuencia:

$$\lim_{n\to\infty} p_{ij}(n) = 0$$

◆

Teorema de convergencia (caso periódico)

Sea $\{X_n, n\geq 0\}$ una cadena de Markov irreducible, recurrente positiva y de periodo $d>1$. Se cumple:

$$\lim_{m\to\infty} p_{ij}(md+r) = d\pi_j = \frac{d}{\mu_j} \qquad 0\leq r<d \qquad i,j\in S$$

Demostración

Consideremos el proceso $Y_m = X_{md}$, $m\geq 0$, $\{Y_m, m\geq 0\}$ es una cadena de Markov con matriz de probabilidades de transición de un paso $Q=P^d$.

Elijamos $j\in S$. Se tiene que:

$$d_Q(j) = \text{M.C.D.}\{m : q_{jj}(m)>0\} = \text{MCD}\{m : p_{jj}(md)>0\} = \frac{1}{d}\text{MCD}\{md : p_{jj}(md)>0\} =$$

$$= \frac{1}{d}\text{MCD}\{n : p_{jj}(n)>0\} = \frac{1}{d}d = 1$$

Así, pues, todos los estados tienen periodo 1 con respecto a la cadena $\{Y_m\}$.

Supongamos que la cadena $\{X_n\}$, y en consecuencia la $\{Y_m\}$ arrancan del estado j. Puesto que la cadena $\{X_n\}$ retornará al estado j después de transcurrir un nº de etapas que sea un múltiplo de d, se sigue que el tiempo esperado de retorno a j en $\{Y_m\}$ será μ_j/d, siendo μ_j el tiempo medio de retorno en la cadena $\{X_n\}$. Por lo tanto, j será un estado recurrente positivo de periodo 1 en la nueva cadena, aplicando el lema anterior (1):

$$\lim_{m\to\infty} q_{jj}(m) = \frac{d}{\mu_j} = d\pi_j$$

y así, pues: $\displaystyle\lim_{m\to\infty} p_{jj}(md) = d\pi_j$, $j\in S$

Si i es otro estado cualquiera, sea $0\leq r<d$ tal que $p_{ij}(n)=0$ salvo que $n=md+r$ se tendrá que:

$$p_{ij}(md+r) = \sum_{k=0}^{m} f_{ij}(kd+r)\, p_{jj}((m-k)d) = \sum_{k=0}^{\infty} f_{ij}(kd+r)\, a_k(m)$$

siendo $a_k(m) = \begin{cases} p_{jj}((m-k)d) & \text{si } m\geq k \\ 0 & \text{si } m<k \end{cases}$

Por el teorema de la convergencia dominada de Lebesgue: (*)

$$\lim_{m\to\infty} p_{ij}(md+r) = \sum_{k=0}^{\infty} f_{ij}(kd+r)\lim_{m\to\infty} a_k(m) = d\pi_j \sum_{k=0}^{\infty} f_{ij}(kd+r) = d\pi_j \underbrace{f_{ij}^*}_{1} = d\pi_j$$

♦

Ejemplo.

Un poliedro tiene todas sus aristas con longitud común igual a 1 cm. Una hormiga camina por las aristas de forma que cuando llega a un vértice elige, con igual probabilidad, cualquier arista incidente para seguir caminando.

La longitud media del camino recorrido hasta volver al vértice de salida es 6 cm. para un tipo de vértices y 7.5 cm. para los restantes. Determinar la forma del poliedro.

Sean v = nº de vértices, a = nº de aristas y c = nº de caras. La fórmula de Euler establece: $c+v = a+2$.

Sea $V=\{1,2 \ldots v\}$ el conjunto de vértices del poliedro ($S=V$ finito). Sea a_j, $j\in$ el nº de aristas incidentes en el vértice j, está claro que:

$$\sum_{j=1}^{v} a_j = 2a$$

Sea p_{jk} la probabilidad de que la hormiga, estando en el vértice j, vaya en una etapa, al vértice k. Se tiene que:

$$p_{jk} = \begin{cases} 1/a_j & \text{si } k \text{ es contiguo a } j \\ 0 & \text{en caso contrario} \end{cases}$$

Evidentemente se trata de una cadena irreducible, y por ser finita, recurrente positiva: existirá, una, distribución estacionaria. La condición de estacionariedad

$$\pi_k = \sum_{j=1}^{v} \pi_j p_{jk} = \sum_{j \text{ incid. con } k} \pi_j \frac{1}{a_j}$$

La solución de este sistema es: $\pi_k = \dfrac{a_k}{2a}$ $\quad k=1,2 \ldots v$

Entonces: $\mu_k = \dfrac{1}{\pi_k} = 2a/a_k$ $\quad k=1,\ldots,v$.

Por hipótesis sólo hay dos valores diferentes para μ_k. Luego sólo habrá dos valores para a_k, que representaremos por a_1 y a_2:

$$\mu_1 = 2a/a_1 = 6 \longrightarrow 2a = 6a_1$$

$$\mu_2 = 2a/a_2 = 7.5 \longrightarrow 2a = 7.5a_2 \Big\} \; 4a = 12a_1 = 15a_2 \longrightarrow 4a_1 = 5a_2 = k' = 4.5k = 4/3\, a$$

luego $a_1 = 5k$

$a_2 = 4k$

$a = 15k$

Sean v_1 y v_2 los números de vértices que hay del 1º y 2º tipo, respectivamente. Se cumplirá:

$$v_1 a_1 + v_2 a_2 = 2a \longrightarrow 5v_1 + 4v_2 = 30$$

La única solución entera $\xrightarrow{(**)}$: $v_1 = 2$ $\quad v_2 = 5 \longrightarrow \underline{v=7}$ vértices

Donde luego $a \leq \binom{7}{2} = 21$ y $a = 15k$, tendríamos que $k=1$ y $\underline{a=15}$ aristas

De la fórmula de Euler: $c = a + 2 - v = \underline{10 \text{ caras}}$.

Sean c_i, $i \geq 3$, el nº de caras del poliedro con i lados, entonces:

$$3c_3 + 4c_4 + 5c_5 + \dots = 2a = 30 = 3(c_3+c_4+c_5+\dots) + c_4 + 2c_5 + \dots = \underbrace{3c}_{30} + c_4 + 2c_5 + \dots$$

luego $c_4 = c_5 = \dots = 0$ y $\underline{c_3 = 10}$. El poliedro tiene todas sus caras triangulares. Se trata de 2 pirámides pentagonales unidas por las bases.

3. Cadenas infinitas irreducibles

Para cadenas irreducibles con un nº infinito (numerable) de estados el comportamiento límite tiene posibilidades más ricas. Tales sistemas pueden ser recurrentes positivos, recurrentes nulos y transitorios. Los sistemas ergódicos (recurrentes positivos y aperiódicos) conservan la propiedad de tener una distribución de equilibrio propia, que constituye una distribución estacionaria, que es la única existente. Siendo, también en este caso las probabilidades de equilibrio las inversas de los tiempos medios de recurrencia. Veamos un procedimiento sencillo para determinar esta distribución estacionaria.

Teorema

Sea una cadena infinita (con $S = \{0,1,2 \dots\}$) irreducible, recurrente positiva y aperiódica (ergódica). Sea el sistema (*):

$$\sum_{j=0}^{\infty} x_j p_{jk} = x_k \qquad k = 0,1,2\dots \qquad (\text{condiciones de estacionariedad})$$

Toda solución absolutamente convergente del sistema (*) es, salvo una constante multiplicativa, la distribución estacionaria π.

Demostración

Supongamos que $x = (x_k, k \in S)$ es una solución de (*), verificándose que $\sum_{k=0}^{\infty} |x_k| < +\infty$. Vemos que se cumple: $xP = x$, en consecuencia:

$$xP^2 = xP = x \longrightarrow xP^n = x$$

es decir: $x_k = \sum_{j=0}^{\infty} x_j p_{jk}(n)$

De nuevo, por el Teorema de la Convergencia Dominada: (*)

$$x_k = \sum_{j=0}^{\infty} x_j \pi_k \longrightarrow \pi_k = x_k \Big/ \sum_{j=0}^{\infty} x_j$$

♦

Las ecuaciones de estacionariedad juegan un papel importante en la teoría de cadenas de Markov. Se ha probado que cuando la cadena es ergódica hay, salvo una constante multiplicativa, una única solución absolutamente convergente. Probaremos ahora que el recíproco es cierto, o sea, que si las ecuaciones de estacionariedad tienen una solución absolutamente convergente entonces la cadena es ergódica.

Consecuentemente, si para una cadena dada podemos obtener una distribución de probabilidad como solución de dicha ecuación, podemos estar seguros de que la cadena es ergódica, y la distribución obtenida es la de equilibrio.

Teorema

Un sistema (infinito) irreducible y aperiódico es ergódico (recurrente positivo) si las ecuaciones (de estacionariedad) $\sum_{j=0}^{\infty} x_j p_{jk} = x_k$, $k=0,1,2\ldots$ (*) tienen una solución no nula absolutamente convergente, y sólo si, toda solución no-negativa es absolutamente convergente.

Demostración

CS Sea $x=(x_k,\ k=0,1,2\ldots)$ una solución del sistema (*), no nula y absolutamente convergente. Sea $x_k \neq 0$, entonces

$$x_k = \sum_{j=0}^{\infty} x_j p_{jk} = \sum_{j=0}^{\infty} x_j p_{jk}(n) \longrightarrow x_k = \sum_{j=0}^{\infty} x_j \lim_{n\to\infty} p_{jk}(n) \quad \text{(dada la convergencia absoluta)}$$

Entonces, $\exists j \,/\, \lim_{n\to\infty} p_{jk}(n) > 0 \longrightarrow \sum_{n=0}^{\infty} p_{jk}(n) = \infty \longrightarrow k$ es un estado recurrente.

$$\Longleftrightarrow \sum_{n=0}^{\infty} p_{kk}(n) \geq \sum_{n=h}^{\infty} \underbrace{p_{kj}(h)}_{>0} p_{jk}(n-h) = +\infty$$

Luego, k es recurrente positivo.

CN Si el sistema es ergódico: $\lim_{n\to\infty} p_{jk}(n) = \pi_k > 0$

Si $\bar{x} = (x_j,\ j=0,1,\ldots)$ es una solución no-negativa de (*), entonces:

$$x_k = \sum_{j=0}^{+\infty} x_j p_{jk}(n) \geq \sum_{j=0}^{J} x_j p_{jk}(n) \xrightarrow[n\to\infty]{} x_k \geq \sum_{j=0}^{J} x_j \pi_k \xrightarrow[J\to\infty]{} \frac{x_k}{\pi_k} \geq \sum_{j=0}^{J} x_j$$

, con $x_k>0$. Luego la solución $\bar{x}$ es absolutamente convergente.

Comentario

Cuando tratamos con sistemas infinitos irreducibles (aperiódicos, por simplificar), un importante problema consiste en determinar si el sistema es ergódico, nulo o transitorio. Hemos visto que si el nº de estados es finito entonces la cadena es ergódica con una distribución de equilibrio, que es la única distribución estacionaria. Para cadenas con infinitos estados buscamos criterios de clasificación basados exclusivamente en las probabilidades de transición de un paso, pues la determinación de probabilidades límite es usualmente muy complicada. Estos criterios dependerán, como se ha visto en los resultados anteriores, de la existencia y natravés de la solución de cierto sistema de ecuaciones lineales relacionado con la cadena. Para establecer estos criterios haremos uso de las probabilidades-tabú.

<u>Definición</u>

Para una cadena de Markov arbitraria, sea H un subconjunto dado de estados. La probabilidad-tabú es:

$${}_H p_{jk}(n) = P\{X_m \notin H \quad m=0,1,\dots,n-1,\ X_n = k / X_0 = j\}$$

i.e., es la probabilidad condicional, dado el estado inicial j, de ingresar en el estado k en n transiciones, evitando el conjunto-tabú H en los tiempos $1,2,\dots,n-1$.

Nótese que ${}_k p_{jk}(n) = f_{jk}(n)$ [illegible]

<u>Teorema</u>

Una cadena infinita irreducible y aperiódica con espacio de estados $S=\{0,1,2,\dots\}$ es transitoria si, y sólo si, el sistema de ecuaciones

$$\sum_{k=1}^{\infty} p_{jk}\, y_k = y_j \qquad j=1,2,\dots \qquad (1)$$

tiene una solución no-nula y acotada.

<u>Demostración</u>

<u>CN</u> Sabemos que si el sistema es transitorio $f_{jk}^* = \sum_{n=1}^{\infty} f_{jk}(n) < 1$. Se cumple, además, la relación: $f_{j0}^* = p_{j0} + \sum_{k=1}^{\infty} p_{jk} f_{k0}^*$.

Sea $g_j = 1 - f_{j0}^*$, la probabilidad de no encontrar nunca el estado 0 a partir de j. Para una cadena transitoria $g_j > 0$ y $g_j = \sum_{k=1}^{\infty} p_{jk}\, g_k$, $j=1,2,\dots$. De ahí que el sistema (1) tenga una solución no-nula y acotada $y_j = g_j$, $j=1,2,\dots$.

<u>CS</u> Supongamos que las ecuaciones (1) tienen una solución no-nula y acotada: $(y_j)_j$. Podemos asumir $|y_j| \le 1$ (dividiendo por la cota del valor absoluto). Sea P_0 la matriz obtenida a partir de P prescindiendo de la fila y columna correspondiente al estado 0. Entonces, si $y = (y_j,\ j=1,2,\dots)$ podemos escribir (1) como:

$$y = P_0\, y \longrightarrow P_0^n\, y = y$$

Los elementos de P_0^n son de hecho las probabilidades-tabú ${}_0 p_{jk}(n)$, con lo que la expresión anterior puede escribirse:

$$y_j = \sum_{k=1}^{\infty} {}_0 p_{jk}(n)\, y_k \qquad j=1,2,\dots$$

y de ahí,

$$|y_j| \le \sum_{k=1}^{\infty} {}_0 p_{jk}(n)\, |y_k| \le \sum_{k=1}^{\infty} {}_0 p_{jk}(n) = g_j(n)$$

donde $g_j(n)$ es la prob de no entrar en 0 durante n transiciones, partiendo de j. Ahora: $1 - f^*_{j0} = g_j = \lim_{n\to\infty} g_j(n) \geq |y_j| > 0$, para algún j. Para este mismo j se cumple $f^*_{j0} < 1$ y j es un estado transitorio, siéndolo todos los demás por tratarse de una cadena irreducible. ◆

Una formulación alternativa de este resultado sería la siguiente:

Teorema

Una cadena infinita irreducible y aperiódica es transitoria sí, y sólo sí, el sistema de ecuaciones

$$\sum_{k=0}^{\infty} p_{jk}\, y_k = y_j \qquad j=1,2\ldots \qquad (2)$$

tiene una solución acotada no constante.

Demostración

CN Tenemos que $f^*_{j0} = p_{j0} + \sum_{k=1}^{\infty} p_{jk} f^*_{k0}$, con lo que la solución de (2) será:

$y_j = f^*_{j0}$, $y_0 = 1$

como $f^*_{j0} < 1$, esta solución es acotada y no constante.

CS Supongamos que (2) tiene una solución $(y_j)_{j=1,2\ldots}$ acotada, no constante. Si en esa solución $y_0 = 0$, esta misma solución lo es de (1), con lo que la cadena es transitoria porque algún $y_j \neq 0$ para que no sea constante la solución de (2).

Si $y_0 \neq 0$, podemos escribir (2) de la siguiente forma:

$$y_j = p_{j0} y_0 + \sum_{k=1}^{\infty} p_{jk} y_k = \left(1 - \sum_{k=1}^{\infty} p_{jk}\right) y_0 + \sum_{k=1}^{\infty} p_{jk} y_k \Longrightarrow y_j - y_0 = \sum_{k=1}^{\infty} p_{jk}(y_k - y_0),$$

y $(y_k - y_0)$ es una solución acotada no-nula de (1). ◆

Ejemplo: Camino aleatorio con probabilidades de transición variables

Consideremos el camino aleatorio definido por la matriz de probabilidades de transición de un paso:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & \ldots & \end{array} \\ \left(\begin{array}{cccccc} r_0 & p_0 & \ldots & & & \\ q_1 & r_1 & p_1 & & & \\ 0 & q_2 & r_2 & p_2 & & \\ 0 & 0 & q_3 & r_3 & p_3 & \\ 0 & 0 & 0 & q_4 & r_4 & \\ & & & & & \end{array} \right) \end{array}$$

siendo $\forall i \geq 1$: $p_i, q_i \neq 0$ (para que sea irreducible)

$p_i + q_i + r_i = 1 \quad \forall i \geq 1$

$r_0, p_0 \neq 0 \quad r_0 + p_0 = 1$

Para determinar si el sistema es transitorio examinemos la ecuación (1):

$$\left.\begin{aligned} y_1 &= r_1 y_1 + p_1 y_2 \\ y_j &= q_j y_{j-1} + r_j y_j + p_j y_{j+1} \qquad j=2,3,\ldots \end{aligned}\right\}$$

dado que $r_j = 1 - p_j - q_j \quad j \geq 1$, las anteriores ecuaciones pueden escribirse:

$$y_2 - y_1 = \frac{q_1}{p_1} y_1 \qquad \ldots \qquad y_{j+1} - y_j = \frac{q_j}{p_j}(y_j - y_{j-1})$$

Recursivamente se obtiene:

$$y_j - y_{j-1} = \frac{q_1 q_2 \ldots q_{j-1}}{p_1 p_2 \ldots p_{j-1}} y_1 \longleftrightarrow y_j - y_{j-1} = \rho_j y_1 \quad \text{siendo} \quad \rho_j \triangleq \frac{q_1 q_2 \ldots q_{j-1}}{p_1 p_2 \ldots p_{j-1}}$$

De ahí que: $y_j = \left(\sum_{k=1}^{j} \rho_k\right) y_1$ con $\rho_1 \equiv 1$

El sistema es transitorio si y_j está acotada, i.e. si $\sum_{k=1}^{\infty} \rho_k < \infty$ y recurrente si $\sum_{k=1}^{\infty} \rho_k = \infty$.

En particular, si $q_j = q$ y $p_j = p$, $j=0,1,2\ldots$ entonces $\rho_k = \left(\frac{q}{p}\right)^{k-1}$ y el sistema será transitorio si, y sólo si, $q < p$.

Nota. Sólo hemos caracterizado el caso en que el sistema infinito irreducible es transitorio o ergódico, habría también que caracterizar el caso en que siendo el sistema recurrente resulta ser nulo. No lo demostraremos, pero esta diferenciación puede establecerse a partir del siguiente resultado:

"Si una cadena de Markov infinita irreducible y aperiódica es recurrente entonces existe una solución positiva de la ecuación de estacionariedad $\sum_{j=0}^{\infty} x_j p_{jk} = x_k$, $k=0,1,\ldots$ que es única salvo constante multiplicativa. La cadena es ergódica o nula según sea, respectivamente, $\sum x_j < \infty$ ó $\sum x_j = +\infty$."

PROCESOS DE NACIMIENTO-MUERTE

Introducción

La Teoría de Colas es el estudio del fenómeno de las líneas de espera originadas cuando los individuos pertenecientes a una determinada población (clientes) llegan, de forma generalmente aleatoria, a uno o más centros de servicio, teniendo que esperar para recibir servicio cuando los servidores están ocupados. El tiempo invertido por un servidor en ofrecer el servicio demandado por un cliente, es también, en general, una variable aleatoria.

El estudio matemático de las colas arranca del trabajo de Erlang, en la segunda década de este siglo, relativo al problema de la congestión de una central telefónica. Desde entonces son muchos los trabajos de investigación desarrollados en este área, a la luz sobre todo de sus numerosísimas aplicaciones prácticas: tráfico telefónico, aeropuertos, carga y descarga en un puerto, hospitales, etc.

En muchas colas, el proceso estocástico subyacente es un proceso de nacimiento-muerte, y una forma de clasificar tales procesos y de determinar las condiciones de convergencia a un régimen estacionario (propuesto por Karlin y McGregor (1957)) consiste en aproximarlos por una cadena de Markov de tiempo discreto, y aplicar los resultados conocidos para tales cadenas.

1. Cadenas de Markov encajadas

Definición

Un proceso estocástico de parámetro continuo $\{X(t), t \geq 0\}$, con estados $S = \{0, 1, 2, \ldots\}$ se denomina proceso de nacimiento-muerte si desde un estado sólo se puede evolucionar a sus vecinos inmediatos, con probabilidades:

$$P(X(t+\Delta t) = j+1 \mid X(t) = j) = \lambda_j \Delta t + o(\Delta t)$$

$$P(X(t+\Delta t) = j-1 \mid X(t) = j) = \mu_j \Delta t + o(\Delta t)$$

$$P(X(t+\Delta t) = j \mid X(t) = j) = 1 - (\lambda_j + \mu_j)\Delta t + o(\Delta t)$$

En un intervalo de tiempo Δt muy pequeño el tamaño de la población, $X(t) = j$, o aumenta en uno con tasa instantánea de probabilidad λ_j, o disminuye en uno con tasa μ_j, o no cambia, bien porque no se produce alteración alguna de la población o bien porque se producen simultáneamente un alta y una baja. Esto último tiene probabilidad:

$$\{1 - \lambda_j \Delta t + o(\Delta t)\}\{1 - \mu_j \Delta t + o(\Delta t)\} + \{\lambda_j \Delta t + o(\Delta t)\}\{\mu_j \Delta t + o(\Delta t)\} = 1 - (\lambda_j + \mu_j)\Delta t + o(\Delta t)$$

Supongamos que se observa el proceso $X(t)$ en una sucesión de instantes discretos, separados Δt (con Δt pequeño). Estas observaciones corresponden a la v.a $X_n = X(n\Delta t)$ $n = 0, 1, \ldots$ que definen un camino aleatorio con matriz de probabilidades de transición del tipo:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \cdots \end{array} \\ \left\{ \begin{array}{ccccc} \psi_0 & \theta_0 & 0 & 0 & \cdots \\ \phi_1 & \psi_1 & \theta_1 & 0 & \cdots \\ 0 & \phi_2 & \psi_2 & \theta_2 & \cdots \\ 0 & 0 & \phi_2 & \psi_3 & \theta_3 \; \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{array} \right\} \end{array}$$

con $\phi_j = \mu_j \Delta t$

$\theta_j = \lambda_j \Delta t$

$\psi_j = 1 - (\lambda_j + \mu_j)\Delta t$

El que μ_j y λ_j sean positivos, $\forall j$, garantiza el carácter irreducible de la cadena (i.e. todos los estados son comunicantes entre sí: $\forall j,k \in S$, $\exists n \geq 0$ / $p_{jk}(n) > 0$), que será a su vez aperiódica (todos los estados tienen periodo 1) por ser $\psi_j > 0$, para Δt suficientemente pequeño.

Para cadenas de Markov infinitas, irreducibles y aperiódicas, sabemos que la cadena será transitoria si, y sólo si, el sistema de ecuaciones

$$\sum_{k=1}^{\infty} p_{jk}\, y_k = y_j \qquad j = 1, 2, \ldots \qquad (1)$$

tiene solución no nula y acotada (a saber $y_j = g_j \equiv$ probabilidad de no encontrar nunca el estado 0 a partir de j; podemos asumir $|y_j| \leq 1$, si y_j está acotada, dividiendo por la cota). En nuestro caso, y para determinar si el sistema es transitorio examinemos las ecuaciones (1):

$$\left.\begin{aligned} y_1 &= \psi_1 y_1 + \theta_1 y_2 \\ y_j &= \phi_j y_{j-1} + \psi_j y_j + \theta_j y_{j+1} \qquad j = 2, 3, \ldots \end{aligned}\right\}$$

y puesto que $\psi_j = 1 - \theta_j - \phi_j$, las anteriores ecuaciones pueden escribirse:

$$y_2 - y_1 = \frac{\phi_1}{\theta_1} y_1 \quad , \quad y_{j+1} - y_j = \frac{\phi_j}{\theta_1}(y_j - y_{j-1}) \qquad j = 2, 3, \ldots$$

recursivamente, se obtiene:

$$y_j - y_{j-1} = \frac{\phi_1 \phi_2 \cdots \phi_{j-1}}{\theta_1 \theta_2 \cdots \theta_{j-1}} y_1 \longleftrightarrow y_j - y_{j-1} = \rho_j y_1 \quad \text{con} \quad \rho_j = \frac{\phi_1 \phi_2 \cdots \phi_{j-1}}{\theta_1 \theta_2 \cdots \theta_{j-1}} = \frac{\mu_1 \cdots \mu_{j-1}}{\lambda_1 \cdots \lambda_{j-1}}$$

De ahí: $y_j = \left(\sum_{k=1}^{j} \rho_k\right) y_1$ con $\rho_1 \equiv 1$

Ahora, el sistema es transitorio si y_j está acotada, i.e., si $\sum_{k=1}^{\infty} \rho_k < \infty$ y recurrente si $\sum_{k=1}^{\infty} \rho_k = \infty$.

Definimos $S_2 = \sum_{n=2}^{\infty} \prod_{i=1}^{n-1} \frac{\mu_i}{\lambda_i} = \sum_{n=2}^{\infty} \rho_n$, luego $S_2 + 1 = \sum_{k=1}^{\infty} \rho_k$. Entonces, la cadena es recurrente si, y sólo si, $S_2 = \infty$.

2. Condiciones de ergodicidad

Si una cadena de Markov infinita, irreducible y aperiódica es recurrente, entonces existe una solución positiva de la ecuación de estacionariedad:

$$\sum_{j=0}^{\infty} x_j\, p_{jk} = x_k \qquad k = 0, 1, 2, \qquad (2)$$

única, salvo constante multiplicativa. La cadena será ergódica si, y sólo si, la solución es absolutamente convergente ($\sum |x_j| < \infty$). En ese caso (ergodicidad), la solución del sistema será múltiplo de las probabilidades de equilibrio π_k:

$$\pi_0 = \pi_0 \psi_0 + \pi_1 \phi_1 \qquad \pi_k = \pi_{k-1}\theta_{k-1} + \pi_k \psi_k + \pi_{k+1}\phi_{k+1} \qquad k = 1, 2, 3, \ldots$$

y de ahí: $\pi_0(1 - \psi_0) = \pi_1 \phi_1 \qquad \pi_k(1 - \psi_k) = \pi_{k-1}\theta_{k-1} + \pi_{k+1}\phi_{k+1} \qquad k = 1, 2, \ldots$

$$\pi_0 \lambda_0 = \pi_1 \mu_1 \qquad \pi_k(\lambda_k + \mu_k) = \pi_{k-1}\lambda_{k-1} + \pi_{k+1}\mu_{k+1} \qquad (3)$$

a partir de la última ecuación se puede deducir: $\mu_n \pi_n = \lambda_{n-1} \pi_{n-1}$ $\quad n \geq 1$ $\quad (4)$.

Para probarlo definamos $a_n = \mu_n \pi_n - \lambda_{n-1} \pi_{n-1}$ $\quad n \geq 1$. De la segunda ecuación de (3) se obtiene que $a_k = a_{k+1}$ para $k \geq 1$, y por la primera ecuación de (3) se deduce que $a_1 = 0$, y por consiguiente $a_k = 0$, $\forall k \geq 1$.

Sustituyendo de forma reiterada: $\pi_n = \frac{\lambda_{n-1}}{\mu_n} \frac{\lambda_{n-2}}{\mu_{n-1}} \cdots \frac{\lambda_0}{\mu_1} \pi_0$ $\quad n \geq 1$

Si $\{\pi_n, n \geq 0\}$ es una distribución estacionaria: $1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 (S_1 + 1)$

siendo $S_1 = \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}} = \sum_{n=1}^{\infty} (\pi_n / \pi_0)$.

Luego, la cadena será ergódica si, y sólo si, $S_1 < \infty$.

En resumen, dados $S_1 = \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}$ y $S_2 = \sum_{n=2}^{\infty} \prod_{i=1}^{n-1} \frac{\mu_i}{\lambda_i}$, la cadena será ergódica si y sólo si, $S_1 < \infty$ y $S_2 = \infty$.

Problema de tráfico telefónico.

Supongamos que los abonados a una central telefónica hacen llamadas en los instantes $0 < \tau_1 < \tau_2 < \ldots$ Es razonable suponer que los tiempos entre llegadas $T_1 = \tau_1$, $T_2 = \tau_2 - \tau_1$, ..., $T_n = \tau_n - \tau_{n-1}$ son v.a. independientes distribuidas exponencialmente con media $1/\lambda$.

La duración de la conversación correspondiente a la llamada que se produce en el instante τ_n es una variable aleatoria S_n que representa el tiempo de ocupación de una línea. También se asumirá que $S_1, S_2, \ldots, S_n, \ldots$ son v.a. independientes exponenciales con media $1/\mu$.

Cada llamada que llega se conecta, dando lugar a una conversación, si se dispone de una línea libre. Distinguiremos los siguientes casos:

nº de líneas disponibles:

- a) Infinitas
- b) $N < \infty$
  - b_1) la llamada que llega cuando las N líneas están ocupadas, se une a una línea de espera, aguardando su turno.
  - b_2) la nueva llamada no es registrada si todas las líneas están bloqueadas.

Para clasificar las colas, atendiendo a sus elementos más característicos, seguiremos a Kendall, que ha propuesto los siguientes símbolos:

M (por Markov), indica que los tiempos entre llegadas o de servicio siguen una distribución exponencial

E_k (por Erlang), se reserva para representar una distribución erlangiana de orden k ($E_1 \equiv M$)

GI (por general independiente), indica que respecto a los tiempos entre llegadas sólo sabemos que son independientes.

G (por general), se reserva para representar los tiempos de servicio con distribución general.

D (por determinístico), indica que los tiempos de servicio o entre llegadas son constantes

Entonces, las colas más simples se identificarán mediante un tripleta A|B|c, donde c es el número de canales de servicio, mientras que A y B representan, respectivamente la distribución de entradas o distribución de los tiempos entre llegadas T, y la distribución de los tiempos de servicio S. En ocasiones se añade un cuarto símbolo al final, que indica la capacidad de la sala de espera.

Caso a: M/M/∞

Sea $X(t)$ el número de líneas ocupadas en el instante t. Entonces $\{X(t), t\geq 0\}$ será un proceso de nacimiento-muerte con:

- $\lambda_n = \lambda$, $n\geq 0$, i.e., la probabilidad de que llegue una nueva llamada en un tiempo Δt es independiente del nº de conferencias que se están desarrollando y vale $\lambda\Delta t$.
- $\mu_n = n\mu$, $n\geq 0$, i.e., la probabilidad de que concluya una conferencia es $n\mu\Delta t$.

$$P\{t< S_n \leq t+\Delta t \,/\, t< S_n\} = \frac{1-e^{-\mu(t+\Delta t)}-1+e^{-\mu t}}{e^{-\mu t}} = 1-e^{-\mu\Delta t} \simeq \mu\Delta t.$$

- la probabilidad de las otras variaciones es un infinitésimo de orden superior a Δt.

Teniendo en cuenta que: $\lim_{t\to\infty} P\{X(t)=n\} = \pi_n = \dfrac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}\pi_0$; $n\geq 1$

Resulta entonces que $\pi_n = \dfrac{1}{n!}\left(\dfrac{\lambda}{\mu}\right)^n \pi_0 \qquad n\geq 1$

Nótese que $S_2 = \sum_{n=2}^{\infty} (n-1)!\left(\dfrac{\mu}{\lambda}\right)^{n-1}$, de modo que $\dfrac{a_{n+1}}{a_n} = n\left(\dfrac{\mu}{\lambda}\right)$, con lo que la serie es divergente: $S_2=\infty$.

Ahora bien: $1=\sum_{n=0}^{\infty}\pi_n = \pi_0\sum_{n=0}^{\infty}\dfrac{1}{n!}\left(\dfrac{\lambda}{\mu}\right)^n = \pi_0 e^{\frac{\lambda}{\mu}}$ $(\Longleftrightarrow S_1<\infty) \longrightarrow \pi_n = e^{-\rho}\dfrac{\rho^n}{n!} \qquad \rho=\dfrac{\lambda}{\mu}$

es decir, $X(t)$ tiene una distribución de larga duración de Poisson con media ρ, siendo

$$\rho = \frac{\lambda}{\mu} = \frac{\frac{1}{\mu}}{\frac{1}{\lambda}} = \frac{\text{tiempo medio de servicio}}{\text{tiempo medio entre llegadas}}$$

Caso b: M/M/N/∞

Sea $X(t)$ el número de clientes que están siendo servidos o que están esperando para recibir servicio. $\{X(t), t\geq 0\}$ es un proceso de nacimiento-muerte con:

- $\lambda_n = \lambda \quad n\geq 0$

- $\mu_n = \begin{cases} n\mu & n=0,1,\ldots,N-1 \\ N\mu & n\geq N \end{cases}$

Las probabilidades de larga duración, si existen, cumplirán:

$$\pi_n = \begin{cases} \dfrac{1}{n!}\left(\dfrac{\lambda}{\mu}\right)^n \pi_0 & n< N \\[2ex] \dfrac{N^N}{N!}\left(\dfrac{\lambda}{N\mu}\right)^n \pi_0 & n\geq N \end{cases}$$

con π_0 dado por la condición: $1=\sum_{k=0}^{\infty}\pi_k = \pi_0\left[\sum_{k=0}^{N-1}\dfrac{1}{k!}\left(\dfrac{\lambda}{\mu}\right)^k + \sum_{k=N}^{\infty}\dfrac{N^N}{N!}\left(\dfrac{\lambda}{N\mu}\right)^k\right]$

Ahora bien, la serie infinita anterior, y en consecuencia la serie, converge sii $\rho = \frac{\lambda}{N\mu} < 1$

Entonces:

$$\Pi_0 = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{n!}(N\rho)^n + \frac{(\rho N)^N}{N!\,(1-\rho)}} \quad \text{puesto que} \quad \frac{N^N}{N!}\sum_{n=N}^{\infty} \rho^n = \frac{N^N}{N!}\,\frac{\rho^N}{1-\rho}$$

y

$$\Pi_n = \begin{cases} \frac{(N\rho)^n}{n!}\,\Pi_0 & n = 0,1 \ldots N \\ \rho^{\,n-N}\,\Pi_N & n = N+1, N+2 \ldots \end{cases}$$

Interés especial tiene el caso de una sola línea: $N=1 \rightarrow M|M|1/\infty$

Entonces la probabilidad estacionaria del tamaño de la cola será:

$$\left.\begin{aligned} \Pi_0 &= \frac{1}{1+\frac{\rho}{1-\rho}} = 1-\rho \\ \Pi_n &= (1-\rho)\,\rho^n \qquad n \geq 1 \end{aligned}\right\} \equiv \text{distribución geométrica}$$

Caso b_2: $M/M/N$

$X(t)$ = el número de líneas ocupadas en t (como en el caso a). Ahora:

$$\lambda_n = \begin{cases} \lambda & n = 0,1 \ldots N-1 \\ 0 & n \geq N \end{cases} \qquad \mu_n = \begin{cases} n\mu & n = 1,2 \ldots N \\ 0 & n > N \end{cases}$$

En este caso, la cadena encajada tiene por matriz de probabilidades de transición, con Δt:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \\ N-1 \\ N \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & N-2 & N-1 & N \end{array} \\ \left(\begin{array}{cccccc} 1-\lambda h & \lambda h & 0 \;\cdots & 0 & 0 & 0 \\ \mu h & 1-(\mu+\lambda)h & \lambda h \;\cdots & 0 & 0 & 0 \\ 0 & 2\mu h & 1-(2\mu+\lambda)h & \cdots & \cdots & \cdots \\ 0 & 0 & 3\mu h \;\cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \;\cdots & (N-1)\mu h & 1-[(N-1)\mu+\lambda]h & \lambda h \\ 0 & 0 & 0 \;\cdots & 0 & N\mu h & 1-N\mu h \end{array}\right) \end{array}$$

La cadena es finita, irreducible y aperiódica luego será ergódica. Se obtienen las probabilidades de equilibrio:

$$\Pi_n = \frac{e^{-\rho}\,\frac{\rho^n}{n!}}{\sum_{k=0}^{N} e^{-\rho}\,\rho^k/k!} \qquad n = 0,1 \ldots N$$

Se trata de la distribución de Poisson truncada, también llamada de Erlang de la pérdida.

COLA M|G|1

Introducción

Para el análisis de la cola M|G|1 un procedimiento que proporciona muy buenos resultados consiste en encajar una cadena de Markov de parámetro discreto, lo que constituye una forma alternativa a la utilizada en el tema anterior de discretizar el proceso. Consideraremos la longitud de la cola sólo en aquellos instantes en que un cliente acaba de completar su servicio.

Otra característica que interesa conocer de la cola, además del análisis de ergodicidad y de la distribución de equilibrio, es su longitud media y la distribución del tiempo de espera del cliente.

Cadena inducida de Markov

Consideremos un sistema con un solo puesto de servicio, cuyas llegadas siguen una distribución de Poisson de parámetro λ y tal que los tiempos de servicio V siguen una distribución general $B(v)$, siendo $E(V)<\infty$, e independientes.

Sea $Z(t)$ el número de clientes en la cola en el momento t. Se observa el proceso en el momento justo t_n en que un cliente acaba de ser servido. Se forma así un proceso estocástico discreto: $X_0=0$ $X_n=Z(t_n)$. Cumpliéndose

$$X_{n+1}=\begin{cases} X_n-1+N & X_n\geq 1 \\ N & X_n=0 \end{cases} \qquad n\geq 0$$

siendo N el número de clientes que llegan durante el servicio al $(n+1)$-ésimo cliente.

Como N no depende más que de V y no de la longitud de la cola ni del momento en que comience el servicio: $\{X_n\}$ es una cadena de Markov.

La distribución de probabilidad de N será:

$$P(N=n)=\int_0^\infty P(N=n\,/\,V=v)\,dB(v)$$

siendo $P(N=n\,/\,V=v)=e^{-\lambda v}\dfrac{(\lambda v)^n}{n!}$

por ser el número de individuos que llegan a una cola en el tiempo v una v.a. $Po(\lambda v)$

Entonces la matriz P de probabilidades de transición de un paso tiene por elemento genérico:

si $i\geq 1$ $\quad p_{ij}=P(X_{n+1}=j\,/\,X_n=i)=P(N=j-i+1)=\begin{cases} 0 & \text{si } j<i-1 \\ \displaystyle\int_0^\infty e^{-\lambda v}\frac{(\lambda v)^{j-i+1}}{(j-i+1)!}\,dB(v) & \text{si } j\geq i-1\geq 0 \end{cases}$

si $i=0$ se repite la fila $i=1$.

Sea $k_r=\displaystyle\int_0^\infty e^{-\lambda v}\frac{(\lambda v)^r}{r!}\,dB(v)$, $r\geq 0$ (1), así pues, $p_{ij}=k_{j-i+1}$, si $i\geq 1$. Entonces:

$$P=\begin{pmatrix} k_0 & k_1 & k_2 & \cdots \\ k_0 & k_1 & k_2 & \cdots \\ 0 & k_0 & k_1 & \cdots \\ 0 & 0 & k_0 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

Trivialmente se trata de una cadena irreducible y aperiódica. Veamos en qué condición es ergódica existiendo una distribución de equilibrio $\pi = (\pi_0\ \pi_1 \ldots \pi_n \ldots)$: $\pi P = \pi$ (2)

Desarrollando (2):

$$\pi_i = \pi_0 k_i + \sum_{r=1}^{i+1} \pi_r k_{i-r+1} \qquad i = 0, 1, 2 \ldots \qquad (3)$$

Sean las funciones generatrices de probabilidad: $\Pi(s) = \sum_{i=0}^{\infty} \pi_i s^i$ $\quad K(s) = \sum_{i=0}^{\infty} k_i s^i \quad s \le 1$

Multiplicando (3) miembro a miembro por s^i y sumando, obtenemos:

$$\Pi(s) = \pi_0 K(s) + \sum_{i=0}^{\infty} \left(\sum_{r=0}^{i+1} \pi_r k_{i-r+1} \right) s^i - \sum_{i=0}^{\infty} \pi_0 k_{i+1} s^i =$$

$$= \pi_0 K(s) + \frac{1}{s} \Pi(s) K(s) - \frac{1}{s} \pi_0 k_0 - \sum_{i=0}^{\infty} \pi_0 k_{i+1} s^i = \pi_0 K(s) + \frac{1}{s} \Pi(s) K(s) - \frac{1}{s} \pi_0 K(s)$$

de donde $s\Pi(s) = s\pi_0 K(s) + \Pi(s) K(s) - \pi_0 K(s) \longrightarrow \Pi(s) = \dfrac{\pi_0 K(s)(s-1)}{s - K(s)}$ (4)

Sea $\mathcal{C}(s) = \dfrac{1-K(s)}{1-s} = \{1 - k_0 - k_1 s - k_2 s^2 - k_3 s^3 \ldots\}\{1 + s + s^2 + s^3 + \ldots\} =$

$$\left.\begin{array}{l} = 1 - k_0 - k_1 s - k_2 s^2 - k_3 s^3 \ldots + \\ + \quad s - k_0 s - k_1 s^2 - k_2 s^3 \ldots + \\ + \quad s^2 - k_0 s^2 - k_1 s^3 \ldots + \\ + \quad s^3 - k_0 s^3 \ldots + \end{array}\right\} = (1-k_0) + (1-k_0-k_1)s + (1-k_0-k_1-k_2)s^2 + \ldots$$

$$= \sum_{j=0}^{\infty} s^j \left(k_{j+1} + k_{j+2} + \ldots \right)$$

Se tiene: $\mathcal{C}(1) = \sum_{j=0}^{\infty} \sum_{n=j+1}^{\infty} k_n = \sum_{n=0}^{\infty} n k_n = \rho$ = media de la distribución de las llegadas durante un servicio

De (1) se deduce, por otra parte:

$$\rho = \int_0^{\infty} e^{-\lambda v} \left(\sum_{n=0}^{\infty} n \frac{(\lambda v)^n}{n!} \right) dB(v) = \int_0^{\infty} \lambda v \, dB(v) = \lambda E(v) \qquad (5)$$

Asimismo, de (4) y de la definición de $\mathcal{C}(s)$: $\Pi(s) = \dfrac{\pi_0 K(s)}{1 - \mathcal{C}(s)}$ (6)

y tomando en (6) $s = 1$: $\begin{cases} \Pi(1) = 1 \\ K(1) = 1 \end{cases} \longrightarrow 1 = \dfrac{\pi_0}{1-\rho} \longrightarrow \pi_0 = 1 - \rho$

¡Desde luego ρ no puede ser ≥ 1, puesto que π_0 ha de ser positivo!

La condición de ergodicidad — $\rho = \dfrac{E(V)}{1/\lambda} = \dfrac{\text{t. m. de servicio}}{\text{t. medio entre llegadas}} < 1$

En definitiva obtendremos todas las π_k a partir de la función generatriz:

$$\Pi(s) = (1-\rho) \frac{K(s)(s-1)}{s - K(s)}$$

2. Servicio en fases: cola $M/E_k/1$

La distribución de Erlang E_k es la distribución de la suma de k v.a. exponenciales, independientes y equidistribuidas, por lo tanto nos permite describir modelos de colas en las que el servicio consta de una serie de fases idénticamente distribuidas e independientes, que deben ser completadas antes de dar servicio a un nuevo cliente. Como ejemplo de tales situaciones podemos considerar los análisis de laboratorio, la revisión de automóviles en un taller realizada por un solo operario, si cabe imaginar que las fases se distribuyen exponencialmente con idéntica media. Sabemos que E_k corresponde a k $\exp(k\mu)$ es una gamma $\Gamma(k, \mu k)$ por lo tanto:

$$b(v) = \frac{(\mu k)^k}{(k-1)!} v^{k-1} e^{-\mu k v}$$

y de ahí:

$$k_r = \int_0^\infty e^{-\lambda v} \frac{(\lambda v)^r}{r!} b(v)\, dv = \int_0^\infty e^{-\lambda v} \frac{(\lambda v)^r}{r!} \frac{(\mu k)^k}{(k-1)!} v^{k-1} e^{-\mu k v} dv = \int_0^\infty e^{-(\lambda+\mu k)v} \frac{(\lambda v)^r}{r!} \frac{(\mu k)^k}{(k-1)!} v^{k-1} dv$$

$$K(s) = \sum_{r=0}^\infty k_r s^r = \int_0^\infty e^{-(\lambda+\mu k)v} \frac{(\mu k)^k}{(k-1)!} v^{k-1} \left(\sum_{r=0}^\infty \frac{[\lambda v s]^r}{r!} \right) dv = \frac{(\mu k)^k}{(k-1)!} \int_0^\infty e^{-[\lambda(1-s)+\mu k]v} v^{k-1} dv =$$

$$= \frac{(\mu k)^k}{(k-1)!} \frac{\Gamma(k)}{[\lambda(1-s)+\mu k]^k} = \frac{1}{\left[\frac{\lambda}{\mu k}(1-s)+1\right]^k}$$

Para la cola $M/E_k/1$ existe distribución de equilibrio si

$$\rho = \frac{\text{tiempo medio de servicio}}{\text{tiempo medio entre llegadas}} = \frac{\frac{k}{\mu k}}{\frac{1}{\lambda}} = \frac{\lambda}{\mu} < 1$$

La función generadora de tal probabilidad límite es:

$$\Pi(s) = (1-\rho) \frac{K(s)(s-1)}{s-K(s)} = (1-\rho) \frac{s-1}{\left[\frac{\lambda}{\mu k}(1-s)+1\right]^k s - 1}$$

3 Longitud media de la cola en régimen estacionario: Fórmula de Pollaczek-Khintchine

Supongamos que el sistema ha alcanzado el régimen estacionario. Sea q el número de personas en la cola después de la partida de un cliente. Se trata de determinar $E(q)$.

Si q' es el número de clientes esperando después de la partida siguiente:

$$q' = q - 1 + \delta + N \qquad (7)$$

con N = número de llegadas durante la duración de un servicio

$$\delta = \begin{cases} 1 & \text{si } q = 0 \\ 0 & \text{si } q > 0 \end{cases}$$, i.e., δ es una función de q: $\delta \equiv \delta(q)$

En régimen estacionario q y q' tienen la misma distribución con lo que $E(q) = E(q')$ y tomando esperanzas en (7): $E(q') = E(q) - 1 + E(\delta) + E(N) \longrightarrow E(\delta) = 1 - E(N) = 1 - \rho = 1 - \lambda E(v)$ (8)

Es evidente que $\delta^2 = \delta$ y tomando cuadrados en (7):

$$q'^2 = q^2 + \delta + (N-1)^2 + 2q\delta + 2\delta(N-1) + 2q(N-1)$$

y como $q\delta = 0$: $q'^2 = q^2 + \delta + N(N-1) + (1-N) + 2\delta(N-1) + 2q(N-1)$

Ahora bien N es independiente de q y en consecuencia de δ, lo que nos lleva a:

$$E(q'^2) = E(q^2) + 1-\rho + E[N(N-1)] + 1-\rho + 2(1-\rho)(\rho-1) + 2E(q)(\rho-1)$$

como $E(q'^2) = E(q^2)$: $\quad 0 = 1-\rho + E[N(N-1)] + 1-\rho + 2(1-\rho)(\rho-1) + 2E(q)(\rho-1)$

$0 = 2 - 2\rho + 4\rho - 2 - 2\rho^2 + E[N(N-1)] + 2E(q)(\rho-1)$

$2E(q)(1-\rho) = 2\rho(1-\rho) + E(N(N-1)) \longrightarrow E(q) = \rho + \dfrac{E[N(N-1)]}{2(1-\rho)}$ (9)

Recordemos que k_n es la probabilidad de que lleguen n individuos durante un servicio

$$K(s) = \sum_{n=0}^{\infty} k_n s^n = \sum_{n=0}^{\infty} \left\{ \int_0^{\infty} e^{-\lambda v} \frac{(\lambda v)^n}{n!} dB(v) \right\} s^n = \int_0^{\infty} e^{-\lambda v(1-s)} dB(v) \qquad (10)$$

luego $E[N(N-1)] = \sum_{n=2}^{\infty} n(n-1) k_n = K''(1)$

Derivando (10) con respecto a s: $K''(s) = \lambda^2 \int_0^{\infty} v^2 e^{-\lambda v(1-s)} dB(v)$

entonces $K''(1) = \lambda^2 \int_0^{\infty} v^2 dB(v) = \lambda^2 [\operatorname{Var}(V) + E^2(V)] = \operatorname{Var}(\lambda V) + \rho^2 \longrightarrow E[N(N-1)] = \operatorname{Var}(\lambda V) + \rho^2$

Sustituyendo en (9): $E(q) = \rho + \dfrac{\operatorname{Var}(\lambda V) + \rho^2}{2(1-\rho)}$ (11)

que es la fórmula de Pollaczek-Kintchine.

4. Valor medio del tiempo de espera

Supongamos que un cliente espera un tiempo W hasta que empieza a ser servido, y que su servicio dura V, así como que al partir deja q clientes en la cola, que han llegado durante el tiempo $W+V$, según un proceso de Poisson de media λ (distribución de llegadas).

En régimen estacionario: $E(q) = \lambda E(W+V) = \lambda\{E(W) + E(V)\}$ (12) que se conoce como fórmula de Little. Entonces

$$\lambda E(W) + \rho = \rho + \frac{\operatorname{Var}(\lambda V) + \rho^2}{2(1-\rho)} \longrightarrow E(W) = \frac{\operatorname{Var}(\lambda V) + \rho^2}{2\lambda(1-\rho)} \qquad (13)$$

Veamos unos ejemplos: cola $M|M|1$

La cola $M|M|1$ es la $M|G|1$, en este caso la varianza de V es $\frac{1}{\mu^2}$, con lo que

$$E(q) = \rho + \frac{\frac{\lambda^2}{\mu^2} + \rho^2}{2(1-\rho)} = \rho + \frac{\rho^2}{1-\rho} = \frac{\rho}{1-\rho}$$

que corresponde a la distribución geométrica $\Pi_n = (1-\rho)\rho^n \quad n = 0, 1, 2 \ldots$. En efecto:

$$E(q) = (1-\rho)\sum_{n=0}^{\infty} n\rho^n = (1-\rho)\rho \sum_{n=1}^{\infty} n\rho^{n-1} = (1-\rho)\rho \frac{d}{d\rho}\left\{\sum_{n=0}^{\infty}\rho^n\right\} = (1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) = \frac{(1-\rho)\rho}{(1-\rho)^2} = \frac{\rho}{1-\rho}$$

y el valor medio del tiempo de espera:

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

cola $M|D|1$

Cada servicio tiene una duración constante μ. N = número de personas que llegan a la cola durante un servicio: $N \stackrel{d}{=} P_o(\lambda\mu)$

luego $k_r = P(N=r) = e^{-\lambda\mu} \dfrac{(\lambda\mu)^r}{r!} \quad r \geq 0$

Así, pues, la función generatriz de probabilidad de N será:

$$K(s) = \sum_{i=0}^{\infty} k_i s^i = e^{-\lambda\mu(1-s)}$$

y por lo tanto, $\Pi(s) = (1-\rho)\dfrac{e^{-\lambda\mu(1-s)}(s-1)}{s - e^{-\lambda\mu(1-s)}} = (1-\rho)\dfrac{s-1}{s\,e^{\lambda\mu(1-s)} - 1}$

La longitud media de la cola es:

$$E(q) = \rho + \frac{\rho^2}{2(1-\rho)} = \frac{2\rho - 2\rho^2 + \rho^2}{2(1-\rho)} = \rho\,\frac{2-\rho}{2-2\rho} = \rho\,\frac{1-\rho/2}{1-\rho}$$

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

COLA GI/M/1

Introducción

Para el análisis de la cola GI/M/1 un procedimiento que proporciona buenos resultados es encajar una cadena de Markov, pero a diferencia del tema anterior, para discretizar el proceso, consideramos la longitud de la cola sólo en aquellos instantes en que un nuevo cliente llega a la cola.

1. Cadena inducida de Markov

Consideremos un sistema con un solo puesto de servicio, cuyos tiempos entre llegadas son independientes y siguen una distribución general $H(u)$, y tal que los tiempos de servicio son exponenciales de parámetro μ.

Sea $Z(t)$ el número de personas en la cola, incluyendo al que recibe servicio, en el instante t. Se observa el proceso en el momento justo en que un nuevo cliente llega a la cola, se forma así un proceso estocástico discreto $X_0 = 0$, $X_n = Z(t_n)$, se cumple:

$$X_{n+1} = X_n + 1 - N \qquad X_n \geq 0 \qquad X_n + 1 \geq N \qquad (1)$$

siendo N el número de personas servidas en el periodo de tiempo que media entre las llegadas sucesivas, que sólo depende de la longitud de dicho intervalo y de X_n.

Puesto que en este modelo se asume que los intervalos de tiempo entre llegadas son independientes, resulta una cadena de Markov con transición en los instantes en que se produce una llegada. $\{X_n\}$ es una cadena de Markov.

La distribución de probabilidad de N, por el teorema de la probabilidad total, será:

$$P(N=n) = \int_0^\infty P(N=n \,/\, T=t)\, dH(t)$$

siendo $P(N=n \,/\, T=t) = e^{-\mu t}\dfrac{(\mu t)^n}{n!} \qquad n \geq 0$

que es N el número de individuos servidos para un intervalo entre llegadas de longitud t, cuando el tiempo de servicio es exponencial μ. Sabemos, por el cálculo de probabilidades, que en este caso $N \sim P_o(\mu t)$

Introduciremos la siguiente notación simplificadora:

$$a_k = P(N=k) = \int_0^\infty e^{-\mu t}\frac{(\mu t)^k}{k!}\, dH(t) \qquad k \geq 0 \qquad (2)$$

Entonces, la matriz de transición de probabilidades de un paso tiene por elemento genérico:

$$p_{ij} = P(X_{n+1} = j \,/\, X_n = i) = \begin{cases} P(N = i+1-j) = a_{i+1-j} & i+1 \geq j \geq 1 \\ 0 & j > i+1 \end{cases}$$

Además, $p_{i0} = 1 - \sum_{j=1}^{\infty} p_{ij} = 1 - \sum_{j=1}^{i+1} p_{ij} = 1 - a_0 - a_1 - \dots - a_i = r_i$, p_{i0} es la probabilidad de que por lo menos hayan sido servidos $i+1$ clientes.

Entonces:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \end{array} \\ \left(\begin{array}{ccccc} r_0 & a_0 & 0 & 0 & \cdots \\ r_1 & a_1 & a_0 & 0 & \cdots \\ r_2 & a_2 & a_1 & a_0 & \cdots \\ r_3 & a_3 & a_2 & a_1 & a_0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{array} \right) \end{array}$$

Esta cadena es irreducible y aperiódica. Veamos directamente si existen las probabilidades de equilibrio: $\Pi P = \Pi$

$$\left.\begin{array}{l} \Pi P = \Pi \\ \Pi = (\Pi_0\, \Pi_1 \, \ldots \, \Pi_n \ldots) \end{array}\right\} \quad (3)$$

condiciones que desarrolladas se expresan: $\sum_{j=0}^{\infty} \Pi_j \, p_{jk} = \Pi_k \quad k=0,1,2\ldots \quad (4)$, de donde

$$\Pi_k = \sum_{j=0}^{\infty} \Pi_{j+k-1}\, a_j \qquad k \geq 1$$

$$\Pi_0 = \sum_{j=0}^{\infty} \Pi_j r_j = \sum_{j=0}^{\infty} \Pi_j \left(1 - \sum_{l=0}^{j} a_l\right)$$

Ensayamos la solución $\Pi_j = \alpha s^j$, con $0<s<1$. Se tiene entonces la solución $\forall k \geq 1$

$$\alpha s^k = \alpha \sum_{j=k-1}^{\infty} s^j a_{j-k+1} \longrightarrow \sum_{j=k-1}^{\infty} s^{j-k+1} a_{j-k+1} = s$$

y mediante un cambio de índices resulta:

$$\sum_{l=0}^{\infty} a_l s^l = s \qquad (5)$$

Si s es una solución de (5), i.e., es solución de (4) $\forall k \geq 1$, resulta con $k=0$:

$$\alpha \sum_{j=0}^{\infty} s^j r_j = \alpha \sum_{j=0}^{\infty} s^j \left(\sum_{k=j+1}^{\infty} a_k\right) = \alpha \left\{ \begin{array}{l} a_1 + a_2 + a_3 + a_4 + a_5 + \ldots + \\ + s(a_2 + a_3 + a_4 + a_5 + \ldots) + \\ + \quad s^2(a_3 + a_4 + a_5 \ldots) + \\ + \quad\quad s^3(a_4 + a_5 \ldots) + \\ \ldots \end{array} \right\} = \alpha \sum_{k=1}^{\infty} a_k \left(\sum_{j=0}^{k-1} s^j\right)$$

sumando las series geométricas que resultan:

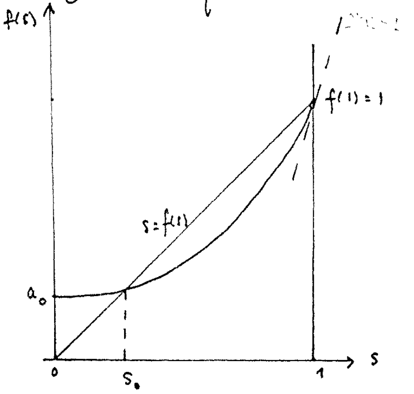
$$\alpha \sum_{j=0}^{\infty} s^j r_j = \alpha \sum_{k=1}^{\infty} a_k \frac{1-s^k}{1-s} = \frac{\alpha}{1-s}\left(1 - a_0 - \sum_{k=1}^{\infty} a_k s^k\right) = \frac{\alpha}{1-s}\left[1 - a_0 - (s - a_0)\right] = \alpha \equiv \alpha s^0$$

Por lo tanto si s_0 es solución de (5), con la condición $0 < s_0 < 1$, $\alpha s_0^n \;\; \forall n \geq 0$ es una solución de (4).

Consideremos la función $f(s) = \sum_{k=0}^{\infty} a_k s^k$, en $[0,1]$, tenemos que $f(0) = a_0$ y $f(1) =$ puesto que $f(1) = \sum_{k=0}^{\infty} a_k$, además:

$$f'(s) = \sum_{k=1}^{\infty} k a_k s^{k-1} \geq 0 \qquad y \qquad f''(s) = \sum_{k=2}^{\infty} k(k-1) a_k s^{k-2} \geq 0$$

luego, la función considerada en $[0,1]$ es creciente y convexa.



Para que tenga solución $f(s)=s$ en $(0,1)$ se debe ocurrir evidentemente que $f'(1) > 1$ (se trata de una condición suficiente de optimalidad, dado que se ha encontrado una solución particular), siendo

$f'(1) = \sum_{k=0}^{\infty} k a_k \equiv$ nº medio de servicios entre dos llegadas consecutivas

Como $\sum_{n=0}^{\infty} \pi_n = 1 = \alpha \sum_{n=0}^{\infty} s_0^n = \frac{\alpha}{1-s_0} \longrightarrow \alpha = (1-s_0)$

y en consecuencia:

$$\pi_n = (1-s_0)\, s_0^n \qquad \forall n \geq 0 \qquad (6)$$

Veamos, por otra parte, que

$$f'(1) = \sum_{k=0}^{\infty} k a_k = \frac{\text{tiempo medio entre llegadas}}{\text{tiempo medio del servicio}} = \frac{1}{\rho}$$

con lo que la condición de ergodicidad es la de siempre $\rho < 1$. En efecto:

$$\sum_{k=0}^{\infty} k a_k = \sum_{k=1}^{\infty} k a_k = \int_0^{\infty} e^{-\mu t} \left\{ \sum_{k=1}^{\infty} \frac{(\mu t)^k k}{k!} \right\} dH(t) = \int_0^{\infty} e^{-\mu t} \mu t \left\{ \sum_{k=1}^{\infty} \frac{(\mu t)^{k-1}}{(k-1)!} \right\} dH(t) =$$

$$= \int_0^{\infty} \mu t \, dH(t) = \mu \int_0^{\infty} t \, dH(t) = \frac{\int_0^{\infty} t \, dH(t)}{\frac{1}{\mu}} = \frac{\text{tiempo medio entre llegadas}}{\text{tiempo medio del servicio}} = \frac{1}{\rho}$$

Distribución del tiempo de espera

Sea W el tiempo de espera de un cliente antes de su servicio. La probabilidad de no esperar es la probabilidad de que a su llegada no haya nadie (π_0, suponiendo que la cola haya alcanzado su régimen de equilibrio). Suponiendo que por el contrario se encuentran a su llegada $n \geq 1$ clientes se tendrá:

$$P(W \leq t \,/\, n \text{ clientes delante}) = F^{(n)}(t) = P(T_1 + T_2 + \dots + T_n \leq t) \quad (*) \qquad t \geq 0$$

siendo $F^{(n)}(\cdot)$ la función de distribución de una v.a $\Gamma(n,\mu)$, que es la suma de las n v.a exponenciales de parámetro μ, que corresponden a los tiempos de servicio T_i de los n clientes que hay en la cola. Por lo tanto,

$$P(W \leq t) = \sum_{n=1}^{\infty} P(W \leq t \,/\, n \text{ delante}) \cdot \pi_n + P(W \leq t \,/\, n=0)\, \pi_0 = \qquad \pi_n = (1-s_0)s_0^n$$

$$= (1-s_0) \int_0^t e^{-\mu\tau} \left\{ \sum_{n=1}^{\infty} \frac{\mu^n \tau^{n-1} s_0^n}{(n-1)!} \right\} d\tau + (1-s_0) \qquad (7)$$

Resolvemos aparte la integral:

$$\int_0^t e^{-\mu\tau} s_0 \mu \left\{ \sum_{n=1}^{\infty} \frac{(\mu\tau s_0)^{n-1}}{(n-1)!} \right\} d\tau = \int_0^t s_0 \mu\, e^{-\mu\tau(1-s_0)} d\tau = -\frac{s_0}{1-s_0} e^{-\mu\tau(1-s_0)} \Big|_0^t = \frac{-s_0}{1-s_0} \left(e^{-\mu t (1-s_0)} - 1 \right)$$

y reemplazando en (7):

$$P(W \le t) = (1-s_0) + s_0\left[1 - e^{-\mu(1-s_0)t}\right] \qquad (8)$$

es decir, se obtiene una mixtura de una distribución exponencial de parámetro $\mu(1-s_0)$ y de una distribución degenerada cuyo único valor posible es 0 con prob. $(1-s_0)$, probabilidad de que el cliente que llega no tenga que esperar.

Ejemplo:

Cola M/M/1. En este caso:

$$a_k = \int_0^\infty e^{-\mu t}\frac{(\mu t)^k}{k!}\lambda e^{-\lambda t}\,dt = \int_0^\infty \lambda e^{-(\mu+\lambda)t}\frac{(\mu t)^k}{k!}\,dt$$

$$f(s) = \sum_{k=0}^\infty a_k s^k = \int_0^\infty \lambda e^{-(\mu+\lambda)t}\left\{\sum_{k=0}^\infty \frac{(\mu t s)^k}{k!}\right\}dt = \int_0^\infty \lambda e^{-[\mu(1-s)+\lambda]t}\,dt =$$

$$= -\frac{\lambda}{\mu(1-s)+\lambda}e^{-[\mu(1-s)+\lambda]t}\Big|_0^\infty = \frac{\lambda}{\mu(1-s)+\lambda} = \frac{1}{\frac{(1-s)}{\rho}+1}$$

entonces:

$$f(s) = s \longrightarrow \frac{1-s}{\rho}+1 = \frac{1}{s} \longrightarrow \frac{1-s}{\rho} = \frac{1}{s}-1 = \frac{1-s}{s} \longrightarrow s_0 = \rho$$

Cola D/M/1

En este caso los intervalos de tiempo entre las llegadas son constantes, de longitud λ. Y por lo tanto, $a_k \equiv$ probabilidad de que se realicen k servicios en un intervalo entre dos llegadas sucesivas será: $a_k = \frac{(\mu\lambda)^k}{k!}e^{-\mu\lambda}$, i.e., $N \doteq Po(\lambda\mu)$

$$f(s) = \sum_{k=0}^\infty a_k s^k = \sum_{k=0}^\infty e^{-\mu\lambda}\frac{(\mu\lambda s)^k}{k!} = e^{-\mu\lambda}e^{\mu\lambda s} = e^{-\mu\lambda(1-s)}$$

$$f'(s) = \mu\lambda\, e^{-\mu\lambda(1-s)} \longrightarrow f'(1) = \mu\lambda > 1 \qquad \mu\lambda = \frac{\lambda}{\frac{1}{\mu}} = \frac{1}{\rho} > 1$$

Resolvamos:

$$f(s) = s \quad : \quad e^{-\mu\lambda(1-s_0)} = s_0$$

entonces:

$$P(W \le t) = (1-s_0) + s_0\left(1 - s_0^{t/\lambda}\right) = 1 - s_0^{t/\lambda + 1}$$

Cola E_k/M/1

Sea el tiempo entre llegadas consecutivas una v.a. Erlang $\Gamma(k, \lambda k)$. Entonces

$$a_n = \int_0^\infty e^{-\mu t}\frac{(\mu t)^n}{n!}\frac{(\lambda k)^k t^{k-1}e^{-\lambda k t}}{(k-1)!}\,dt = \frac{(\lambda k)^k}{(k-1)!}\int_0^\infty e^{-(\mu+\lambda k)t}\frac{(\mu t)^n}{n!}t^{k-1}\,dt$$

luego

$$f(s) = \sum_{n=0}^\infty a_n s^n = \frac{(\lambda k)^k}{(k-1)!}\int_0^\infty e^{-(\mu+\lambda k)t}e^{\mu t s}t^{k-1}\,dt = \frac{(\lambda k)^k}{(k-1)!}\int_0^\infty e^{-[\mu(1-s)+\lambda k]t}t^{k-1}\,dt =$$

$$= \frac{(\lambda k)^k}{\cancel{\Gamma(k)}} \frac{\cancel{\Gamma(k)}}{[\mu(1-s)+\lambda k]^k} = \frac{1}{\left[\frac{\mu}{\lambda k}(1-s)+1\right]^k} = \left[\frac{1-s}{k\rho}+1\right]^{-k} \qquad \rho = \frac{1/\mu}{1/\lambda}$$

ahora bien:

$$f'(s) = -k\left[\frac{1-s}{k\rho}+1\right]^{-k-1}\left(-\frac{1}{k\rho}\right) \longrightarrow f'(1) = (-k)\left(-\frac{1}{k\rho}\right) = \frac{1}{\rho} > 1 \Leftrightarrow \rho < 1$$

$$f(s) = s \Leftrightarrow \frac{1-s}{k\rho}+1 = s^{-\frac{1}{k}}$$

sea $r = s^{1/k} \longrightarrow \dfrac{1-r^k}{k\rho}+1 = \dfrac{1}{r} \longrightarrow (1+k\rho-r^k)\,r = k\rho$

$$r^{k+1} - r(1+k\rho)+k\rho = 0$$

Comprobamos que para $k=1$ obtenemos el resultado de la cola M/M/1:

$$r^2 - r(1+\rho)+\rho = 0 \longrightarrow r = \frac{(1+\rho) \pm \sqrt{(1+\rho)^2-4\rho}}{2} = \frac{1+\rho \pm (1-\rho)}{2} \begin{cases} \nearrow \rho \\ \searrow \cancel{1} \end{cases}$$

Capítulo 6

Colas M/M/1

TEMA VI: COLAS M|M|1.

Introducción

La Teoría de colas es el estudio del fenómeno de las líneas de espera, originadas cuando los individuos pertenecientes a una determinada población (clientes o usuarios) llegan, de forma generalmente aleatoria, a uno o más centros de servicio, teniendo que esperar para recibir servicio cuando los servidores están ocupados. El tiempo invertido por un servidor en ofrecer el servicio demandado por el cliente es también, en general, una variable aleatoria.

El de cliente o usuario es un término genérico que puede hacer referencia a clientes pidiendo servicio en un mostrador, barcos entrando en un puerto, flujo de mensajes en una oficina, máquinas averiadas esperando ser reparadas, etc.

Los procesos de colas se clasifican de acuerdo con:

1) distribución de las llegadas, distribución del modelo de llegada de los clientes en el tiempo, o más específicamente: distribución de los tiempos entre llegadas;

2) distribución del servicio, distribución del tiempo de servicio a un cliente; y

3) disciplina de la cola, que indica el número de servidores o canales de servicio y la organización de la línea de espera y servicio. En la mayoría de los modelos la disciplina es "el primero que llega es el primero al que se sirve", y el servicio empieza en cuanto el cliente alcanza la cabeza de la cola. Nosotros sólo consideraremos esta disciplina. FIFO

Para clasificar las colas atendiendo a sus elementos más característicos seguiremos a Kendall (1951), que ha propuesto los siguientes símbolos:

M (por Markov): indica que los tiempos entre llegadas o de servicio siguen una distribución exponencial.

E_k (por Erlang): se reserva para representar una distribución erlangiana (gamma) de orden k ($E_1 \equiv M$)

GI (general independiente): indica que respecto a los tiempos entre llegadas sólo sabemos que son independientes.

G (general): se reserva para representar los tiempos de servicio con distribución general

D (determinista): indica que los tiempos de servicio o entre llegadas son constantes

Las colas más simples se representan mediante un triplete A|B|C, donde c es el número de canales de servicio (servidores), mientras que A y B representan, respectivamente, la distribución de las llegadas y la distribución del servicio. En ocasiones se añade un cuarto símbolo, al final, que indica la capacidad (o del sistema) de la sala de espera.

Veamos, en primer lugar, que bajo ciertas hipótesis sobre los procesos de llegada y de servicio, una gran cantidad de sistemas de colas pueden ser modelados como un proceso de nacimiento-muerte:

(1) si el nº de clientes en el sistema es n en el instante t, durante el intervalo $(t, t+\Delta t)$ llegará un nuevo cliente con prob. $\lambda_n \Delta t + o(\Delta t)$, no llegará ninguno con prob. $1-\lambda_n \Delta t + o(\Delta t)$ y llegará más de uno con prob. $o(\Delta t)$. Suponemos, asimismo, que el nº clientes llegados en el intervalo $(t, t+\Delta t)$ es independiente del tiempo transcurrido desde la última llegada.

(2) si en el instante t hay n clientes en el sistema, hay una prob. $\mu_n \Delta t + o(\Delta t)$ de que durante el intervalo $(t, t+\Delta t)$ se complete un servicio, de que no se complete un servicio la prob. es $1-\mu_n \Delta t + o(\Delta t)$ y de que se complete más de un servicio es $o(\Delta t)$. Se supondrá, asimismo, que el tiempo para completar un servicio es independiente del instante en que empezó. (pérdida de memoria)

(3) Además, suponemos que la dependencia entre llegadas y servicios se restringe a las hipótesis establecidas en (1) y en (2), en las que ambos dependen del tamaño de la población. Se asume $\mu_0 = 0$ y $\lambda > 0$, lógicamente.

En estas condiciones, si llamamos $Q(t)$ al número de clientes en el sistema en el instante t y definimos:

$$p_n(t) = P[\,Q(t)=n \mid Q(0)=i\,]$$

siendo las condiciones iniciales $p_n(0)=\begin{cases}1 & n=i\\ 0 & n\neq i\end{cases}$, tendremos la siguiente ecuación para describir el sistema:

$$\left.\begin{aligned} p_0'(t) &= -\lambda_0\, p_0(t) + \mu_1\, p_1(t)\\ p_n'(t) &= -(\lambda_n+\mu_n)\, p_n(t) + \lambda_{n-1}\, p_{n-1}(t) + \mu_{n+1}\, p_{n+1}(t) \qquad n\geq 1\end{aligned}\right\}$$

que corresponden a un proceso de nacimiento-muerte generalizado: ecuaciones de Kolmogorov (forward).

En cada uno de los casos que estudiemos, es decir, para cada modelo de cola, especificaremos los valores de λ_n y μ_n y si se satisfacen las condiciones que aquí se han indicado.

1. Problema de tráfico telefónico

El estudio matemático de las colas arranca del trabajo de Erlang en la segunda década de este siglo, relativo al problema de la congestión de una central telefónica.

Supongamos que los abonados a una central telefónica hacen llamadas en los instantes $0<\tau_1<\tau_2<\dots<\tau_n<\dots$. Es razonable suponer que las llamadas se producen aleatoriamente, lo que se modelaría según un proceso de Poisson $\lambda>0$, lo (véase tema 5) que implica que los tiempos entre llegadas: $T_1=\tau_1$, $T_2=\tau_2-\tau_1$, ... $T_n=\tau_n-\tau_{n-1}$ son v.a. independientes, distribuidas exponencialmente con media $1/\lambda$.

Supondremos también que la duración de la llamada producida en el instante τ_n es una v.a. S_n, que representa el tiempo de ocupación de una línea, y que tiene distribución exponencial con media $1/\mu$; siendo las variables $S_1\, S_2 \dots$ independientes y equidistribuidas.

Cada llamada que llega se conecta, si se dispone de una línea libre, dando lugar a una conversación. Distinguiremos los siguientes casos según el número de líneas disponibles:

a) infinitas, ninguna restricción de capacidad: M|M|∞

b) finitas N: (b1) cuando hay N líneas ocupadas las llamadas que llegan se incorporan a una línea de espera: M|M|N|∞

M|M|N

(b2) si todas las líneas de espera están ocupadas, la nueva llamada no es registrada, se pierde: M|M|N (también M|M|N|0)

Caso a : $M|M|\infty$

Sea $X(t)$ el número de líneas ocupadas en el instante t, $t\geq 0$. En las condiciones generales que se han enumerado para el problema del tráfico telefónico, cuando no hay restricción de capacidad tendremos que $\{X(t),\ t\geq 0\}$ es un proceso de nacimiento-muerte con las características siguientes: $S=\{0,1,2,\dots\}$

(a1) la probabilidad de que llegue una nueva llamada en un intervalo de amplitud Δt es independiente del nº de conferencias que se están desarrollando:

$$P(\tau_{n+1}\leq \tau_n+\Delta t)=P(T_{n+1}\leq \Delta t)=1-e^{-\lambda\Delta t}\simeq \lambda\Delta t \quad \text{cuando } \Delta t\to 0 \qquad \forall n\geq 0$$

siendo $\lambda_n=\lambda \quad n\geq 0$

(a2) la probabilidad de que concluya una llamada en un intervalo de amplitud Δt será:

$$P(t<S_j\leq t+\Delta t\ /\ t<S_j)=\frac{P(t<S_j\leq t+\Delta t)}{P(t<S_j)}=\frac{1-e^{-\mu(t+\Delta t)}-1+e^{-\mu t}}{e^{-\mu t}}=1-e^{-\mu\Delta t}\underset{\Delta t\to 0}{\simeq}\mu\Delta t$$

luego $\mu_n=n\mu \quad n\geq 0$ (modelo binomial $(n,\mu\Delta t)$) $\quad P(t<S_j\leq t+\Delta t)=P(S_j\leq t+\Delta t)-P(S_j\leq t)$

(a3) la probabilidad de las otras variaciones es un infinitésimo de orden superior a Δt.

Se sabe que la condición de ergodicidad para un proceso de nacimiento muerte era:

$$S_1=\sum_{n=1}^{\infty}\left(\prod_{i=0}^{n-1}\frac{\lambda_i}{\mu_{i+1}}\right)<+\infty \qquad \text{y} \qquad S_2=\sum_{k=2}^{\infty}\left(\prod_{j=1}^{k-1}\frac{\mu_j}{\lambda_j}\right)=+\infty$$

En nuestro caso:

$$S_1=\sum_{n=1}^{\infty}\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n=e^{\lambda/\mu}-1<+\infty$$

$$S_2=\sum_{n=2}^{\infty}\left(\prod_{i=1}^{n-1}\mu_i/\lambda_i\right)=\sum_{n=2}^{\infty}(n-1)!\left(\frac{\mu}{\lambda}\right)^{n-1}=+\infty \quad \left(\text{puesto que } \frac{a_{n+1}}{a_n}=n\left(\frac{\mu}{\lambda}\right)(+)\right)$$

En consecuencia:

$$\lim_{t\to\infty}P[X(t)=n]=\pi_n=\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\cdots\mu_n}\pi_0 \qquad n\geq 1$$

es decir: $\pi_n=\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\pi_0 \qquad n\geq 1$

Ahora bien, $1=\sum_{n=0}^{\infty}\pi_n=\pi_0\left(\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right)=\pi_0\, e^{\lambda/\mu} \longrightarrow \boxed{\pi_n=e^{-\rho}\frac{\rho^n}{n!},\ n\geq 0,\quad \rho=\frac{\lambda}{\mu}}$

(*) Criterio de la razón:

(a) si $\exists\, n_0 \in \mathbb{N}$: $\forall n \geq n_0$ $\left|\frac{a_{n+1}}{a_n}\right| \geq 1 \rightarrow$ serie $\sum a_n$ diverge

(b) si $\limsup_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| < 1 \rightarrow$ serie $\sum a_n$ converge

Es decir, $X(t)$ tiene una distribución de equilibrio que es de Poisson de parámetro $\rho = \lambda/\mu$, lo que puede ser expresado como:

$$\underset{\lambda}{\rho} = \frac{1/\mu}{1/\lambda} = \frac{\text{tiempo medio de servicio}}{\text{tiempo medio entre llegadas}} = \text{número medio de la cola en régimen de equilibrio.}$$

intensidad del tráfico

Nota:

· CASO b₁ : M/M/N/∞

Sea $X(t)$ el número de clientes que están siendo servidos o que están esperando para recibir servicio, en t, $t \geq 0$, se trata de un p. nacimiento-muerte $S=\{0,1,2,\ldots\}$ en el que los parámetros para las llegadas siguen siendo los mismos que en el caso a: $\lambda_n = \lambda$, $n \geq 0$, ahora bien ha cambiado la situación para los tiempos de servicio:

$$\mu_n = \begin{cases} n\mu & n = 0,1,2,\ldots,N-1 \\ N\mu & n \geq N \end{cases}$$

Nótese que cuando $n \geq N$, hay N clientes recibiendo servicio y $n-N$ que están esperando ser atendidos, lo que explica la formulación anterior.

En este caso el estudio de las condiciones de ergodicidad conduce a:

$$S_1 = \underbrace{\sum_{n=1}^{N-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n}_{\leq e^{\lambda/\mu}} + \frac{N^N}{N!}\sum_{n=N}^{\infty}\left(\frac{\lambda}{N\mu}\right)^n < +\infty \quad \text{si} \quad \frac{\lambda}{N\mu} < 1$$

$$S_2 = \sum_{n=2}^{N}(n-1)!\left(\frac{\mu}{\lambda}\right)^{n-1} + \frac{N!}{N^N}\sum_{n=N+1}^{\infty}\left(\frac{N\mu}{\lambda}\right)^{n-1} = +\infty \quad \left(\frac{a_{n+1}}{a_n} = \frac{N\mu}{\lambda}\right) \quad \text{si} \quad \frac{N\mu}{\lambda} \geq 1$$

Si existe la distribución de equilibrio, $\frac{\lambda}{N\mu} < 1$, las probabilidades estacionarias serán:

$$\pi_n = \begin{cases} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \pi_0 & n < N \\ \frac{N^N}{N!}\left(\frac{\lambda}{N\mu}\right)^n \pi_0 & n \geq N \end{cases} \qquad \pi_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n}\pi_0 \quad n \geq 1$$

siendo $1 = \sum_{k=0}^{\infty}\pi_k = \pi_0\left[\sum_{k=0}^{N-1}\frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \sum_{k=N}^{\infty}\frac{N^N}{N!}\left(\frac{\lambda}{N\mu}\right)^k\right]$, y la serie converge si y sólo si, $\frac{\lambda}{N\mu} < 1$ (como ya sabíamos).

(*) a efectos de rapidez y simplificación, parece lógico invertir el orden de exposición: b_2, b_1 y esto [illegible]

Luego, si $\rho = \dfrac{\lambda}{N\mu} < 1$ tendremos: $\dfrac{N^N}{N!} \sum_{k=N}^{\infty} \left(\dfrac{\lambda}{N\mu}\right)^k = \dfrac{N^N}{N!} \dfrac{\rho^N}{1-\rho}$ y

$$\pi_0 = \left\{ \sum_{n=0}^{N-1} \frac{1}{n!} (N\rho)^n + \frac{(N\rho)^N}{N!\,(1-\rho)} \right\}^{-1} \qquad \pi_0 = \frac{1}{1+S_1}$$

siendo:

$$\pi_n = \begin{cases} \dfrac{(N\rho)^n}{n!} \pi_0 & n = 1, 2, \ldots N \\ \rho^{n-N} \pi_N & n = N+1, \ldots \end{cases}$$

CASO b_2: M/M/N

Sea $X(t)$ el nº de líneas ocupadas en t, $t \geq 0$, en este caso tenemos una restricción importante dada por [no hay sala de espera y] sólo hay N canales de servicio(*), en consecuencia $S = \{0, 1, \ldots N\}$. Ahora bien, el proceso puede ser modelado como un proceso de nacimiento-muerte, análogo al caso (a), en el que los parámetros son:

$$\lambda_n = \begin{cases} \lambda & n = 0, 1, \ldots N-1 \\ 0 & n \geq N \end{cases}$$

$$\mu_n = \begin{cases} n\mu & n = 1, 2 \ldots N \\ 0 & n > N \end{cases}$$

En este caso, para $\Delta t = h$, la cadena encajada tiene por matriz de probabilidades de transición:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \\ N-1 \\ N \end{array} \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & 2 & 3 & & N-2 & N-1 & N \end{array} \\ \left(\begin{array}{cccccccc} 1-\lambda h & \lambda h & 0 & - - - & & 0 & 0 & 0 \\ \mu h & 1-(\mu+\lambda)h & \lambda h & - \cdots & & 0 & 0 & 0 \\ 0 & 2\mu h & 1-(2\mu+\lambda)h & \lambda h & & 0 & 0 & 0 \\ 0 & 0 & 3\mu\lambda & 1-(3\mu+\lambda)h & & 0 & 0 & 0 \\ & & & & & & & \\ 0 & 0 & 0 & & & (N-1)\mu h & 1-[(N-1)\mu+\lambda]h & \lambda h \\ 0 & 0 & 0 & & & 0 & N\mu h & 1-N\mu h \end{array} \right) \end{array}$$

(*) Nótese que al medir líneas ocupadas y ser el nº de estados finito nos ha quedado $S=\{0,1,\ldots,N\}$, ahora bien desde el punto de vista del nº de clientes en el sistema cabría la posibilidad de plantearse el truncamiento de la cola atendiendo a la capacidad de la sala de espera, y podríamos tener modelo $M|M|N|K$, $K\geq 0$. En realidad este modelo es un truncamiento del tipo (b_1), más que otra versión de (b_2), y tendríamos que $S=\{0,1,\ldots,N+K\}$ y habría que determinar p_{ij} con precaución. (G-H p. 104)

La cadena es finita, irreducible y aperiódica, luego será ergódica y existirá, por tanto, distribución de equilibrio. Al resolver las ecuaciones $\pi = \pi P$ se obtiene:

$$\pi_1 = \frac{\lambda}{\mu}\pi_0 \ , \quad \pi_2 = \frac{\lambda}{2\mu}\pi_1 = \frac{\lambda^2}{2\mu^2}\pi_0 \ , \quad \pi_3 = \frac{\lambda}{3\mu}\pi_2 = \frac{1}{3!}\left(\frac{\lambda}{\mu}\right)^3\pi_0 \quad \ldots \quad \pi_N = \frac{\lambda}{N\mu}\pi_{N-1} = \cdots = \frac{1}{N!}\left(\frac{\lambda}{\mu}\right)^N \pi_0$$

luego, para $n = 1, \ldots N$: $\pi_n = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \pi_0$, siendo $\sum_{n=0}^{N} \pi_n = 1 = \left(\sum_{n=0}^{N} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right)\pi_0$

en total $\pi_0 = \dfrac{1}{\sum_{n=0}^{N} \rho^n/n!}$ $\quad \rho = \frac{\lambda}{\mu}$ $\quad$ y

$$\pi_n = \frac{1/n! \ \rho^n}{\sum_{k=0}^{N} \rho^k/k!} = \frac{1/n! \ e^{-\rho}\rho^n}{\sum_{k=0}^{N} e^{-\rho}\rho^k/k!} \qquad n = 1, 2 \ldots N$$

La distribución de equilibrio es la distribución de Poisson truncada(*), llamada también Erlangiana de pérdida.

2. La cola M|M|1|∞

Consideremos una línea de espera en la que las llegadas ocurren una a una según un proceso de Poisson de parámetro λ. Estos clientes son atendidos por un servidor único y sus tiempos de servicio son v.a. independientes e idénticamente distribuidas $\exp(\mu)$, lo que significa que cuando haya bastantes clientes esperando ser servidos, el nº de servicios acabados (completados) será un proceso de Poisson de parámetro μ.(**) Supondremos, además, que el servicio no se interrumpe mientras haya clientes esperando ser atendidos.

Se trata de un proceso de nacimiento-muerte cuyos parámetros serán:

- $\lambda_n = \lambda \qquad n \geq 0$
- $\mu_n = \mu \qquad n > 0 \ , \ \mu_0 = 0$

Es un caso particular del (b₁), y ya se ha visto que la condición de ergodicidad es $\rho = \frac{\lambda}{\mu} < 1$. Siendo la distribución de equilibrio del tamaño de la cola:

$$\left.\begin{array}{l} \pi_0 = \dfrac{1}{1 + \frac{\rho}{1-\rho}} = 1 - \rho \\ \pi_1 = \rho(1-\rho) \\ \pi_n = (1-\rho)\rho^n \qquad n \geq 1 \end{array}\right\}$$

(*) es, en realidad, un truncamiento de la cola $M/M/1/\infty$

(**) la hipótesis "que haya bastantes clientes esperando", bastante ambigua por otra parte, aseguraría la distribución del nº de servicios completados, pero persiste el problema de determinar cuántos clientes serían "bastantes". Nótese que en (3) se establece un resultado, aparentemente contradictorio con éste, y que sólo es válido cuando el sistema alcanza el equilibrio.

que corresponde a una distribución geométrica de parámetro $(1-\rho) = \frac{\mu-\lambda}{\mu}$. El tamaño medio del sistema en régimen de equilibrio será:

$$L = E(Q) = \frac{\rho}{1-\rho} = \frac{\lambda/\mu}{\frac{\mu-\lambda}{\mu}} = \lambda(\mu-\lambda)^{-1} \qquad \text{siendo } \rho = \frac{\lambda}{\mu} < 1$$

Nótese que si el sistema lleva mucho tiempo funcionando y $\lambda<\mu$, la probabilidad de ser atendido inmediatamente, al llegar a la cola, es:

$$\pi_0 = (1-\rho) = \frac{\mu-\lambda}{\mu} \quad = \text{prob. de éxito de la distr. geométrica}$$

Otra información útil, en problemas prácticos, es el conocimiento de la distribución de probabilidad del tiempo que un cliente ha de permanecer en la cola, T (si se supone que $\lambda<\mu$ y el sistema está en equilibrio), y del tiempo de espera antes de ser atendido, W. Veamos cuales son estas distribuciones:

(i) Sea T = tiempo de permanencia en la cola. Si al llegar a la cola tenemos n clientes delante (incluido el que está siendo atendido), tardaremos en abandonar la cola un tiempo que se distribuye $\Gamma(n+1,\mu)$, luego

$$P(T\le t / \text{ hay } n \text{ delante}) = \int_0^t \frac{\mu^{n+1}\tau^n e^{-\mu\tau}}{n!}\, d\tau \qquad t>0$$

en consecuencia: $t>0$

$$P(T\le t) = \sum_{n=0}^{\infty} P(T\le t/ \text{ hay } n \text{ delante}) \left(\frac{\lambda}{\mu}\right)^n \left(1-\frac{\lambda}{\mu}\right) =$$

$$= \sum_{n=0}^{\infty} \int_0^t \frac{\mu^{n+1}\tau^n e^{-\mu\tau}}{n!} \left(\frac{\lambda}{\mu}\right)^n \left(1-\frac{\lambda}{\mu}\right) d\tau =$$

$$= \int_0^t \mu e^{-\mu\tau}\left(1-\frac{\lambda}{\mu}\right) \sum_{n=0}^{\infty} \frac{(\lambda\tau)^n}{n!}\, d\tau = \int_0^t \mu e^{-\mu\tau}\left(1-\frac{\lambda}{\mu}\right) e^{\lambda\tau}\, d\tau =$$

$$= \int_0^t \left(1-\frac{\lambda}{\mu}\right)\mu \exp\{-\mu\tau\left(1-\frac{\lambda}{\mu}\right)\}\, d\tau = 1 - e^{-\mu t(1-\lambda/\mu)}$$

es decir

$$T \overset{D}{=} \exp\{\mu(1-\tfrac{\lambda}{\mu})\} = \exp\{\mu-\lambda\} \qquad \longrightarrow \qquad E(T) = \frac{1}{\mu-\lambda} = \text{tiempo medio de permanencia en la cola}$$

$\mu-\lambda>0$

(iii) Sea W el tiempo de espera en la cola, antes de ser atendido. Hemos de diferenciar 2 situaciones: que al llegar no haya nadie en la cola (lo que ocurre con probabilidad π_0) o que se encuentre con n clientes, $n\geq 1$, que llegaron antes que él y esperan todavía para ser atendidos. Es decir:

$$P(W=0) = \pi_0 = (1-\rho) = 1-\frac{\lambda}{\mu}$$

$$P(W\leq t / \text{ hay } n \text{ delante}) = P(T_1+T_2+\ldots+T_n\leq t) = \int_0^t \mu^n \frac{\tau^{n-1}e^{-\mu\tau}}{(n-1)!}\,d\tau \qquad t>0$$

En consecuencia, $t\geq 0$:

$$P(W\leq t) = \sum_{n=1}^{\infty} P(W\leq t/\text{ hay } n \text{ delante})\cdot\pi_n + \pi_0 =$$

$$= \left(1-\frac{\lambda}{\mu}\right) + \left(1-\frac{\lambda}{\mu}\right)\int_0^t e^{-\mu\tau}\left(\sum_{n=1}^{\infty}\mu^n\frac{\tau^{n-1}}{(n-1)!}\rho^n\right)d\tau =$$

$$= (1-\rho)+(1-\rho)\lambda\int_0^t e^{-\mu\tau}\left(\sum_{n=1}^{\infty}\frac{(\lambda\tau)^{n-1}}{(n-1)!}\right)d\tau = (1-\rho)+(1-\rho)\lambda\int_0^t e^{-(\mu-\lambda)\tau}\,d\tau$$

$$= (1-\rho)+\rho\int_0^t(\mu-\lambda)e^{-(\mu-\lambda)\tau}\,d\tau = (1-\rho)+\rho\left[1-e^{-(\mu-\lambda)t}\right] = \quad (1)$$

$$= 1-\rho e^{-(\mu-\lambda)t} = 1-\rho\, e^{-\mu(1-\rho)t}$$

Lo que se obtiene (1) es una mixtura de una distribución exponencial de parámetro $\mu-\lambda=\mu(1-\rho)$ y de una distribución degenerada cuyo único valor posible es el cero con probabilidad $(1-\rho)$.

Veamos cuanto es el tiempo medio de espera:

$$E(W) = 0\cdot\pi_0 + \int_0^\infty t\,f_W(t)\,dt = \rho\int_0^{t}\mu(1-\rho)t\,e^{-\mu(1-\rho)t}\,dt \underset{p.p.}{=} -\rho t\,e^{-\mu(1-\rho)t}\Big|_0^{+\infty} +$$

$$+\rho\int_0^{+\infty}e^{-\mu(1-\rho)t}\,dt = \frac{\rho}{\mu(1-\rho)} = \frac{\rho}{\mu-\lambda} \longrightarrow E(W) = \rho/_{\mu-\lambda}$$

Nótese que se verifica la relación: $E(T) = E(W)+\frac{1}{\mu}$, i.e. el tiempo medio de permanencia en la cola es la suma del tiempo medio de espera y del tiempo medio de servicio.

3. Cola con recursos finitos

A diferencia de los modelos estudiados hasta ahora, supondremos que la población, i.e. el número de posibles usuarios, es finita de tamaño K, y que las probabilidades de que se produzca una llegada son funciones que dependen del pasado del sistema (lo que se reflejará en los parámetros del mismo). Una aplicación típica de este modelo es la que corresponde a la cola que forman las máquinas averiadas a la espera de ser reparadas.

Supondremos que hay c canales de servicio, $c < K$, y que los tiempos de servicio son v.a. independientes equidistribuidas $\exp(\mu)$, siendo el proceso de llegadas el siguiente: si una unidad no está en el sistema en t, la probabilidad de que haya entrado en $(t, t+h)$ es $\lambda h + o(h)$, para $h>0$, $h \to 0$, lo que significa que el tiempo que transcurre hasta que la unidad entra en el sistema es exponencial λ.

Sea $X(t)$ el número de unidades que están en el sistema en t, $t \geq 0$. Se trata de un proceso de nacimiento-muerte, con espacio de estados $S = \{0, 1, \dots, K\}$ y parámetros:

$$\lambda_n = \begin{cases} (K-n)\lambda & 0 \leq n < K \\ 0 & n \geq K \end{cases}$$

$$\mu_n = \begin{cases} n\mu & 0 \leq n < c \\ c\mu & n \geq c \end{cases}$$

Y diremos que se tiene una cola $M/M/c/p$ con recursos finitos. Dado que se trata de un proceso con un nº finito de estados, la cadena de Markov encajada será recurrente positiva, y en consecuencia, las probabilidades estacionarias verificarán:

$$\pi_n = \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu_1 \cdots \mu_n} \pi_0 \qquad n = 1, \dots, K$$

y multiplicando adecuadamente:

$$\pi_n = \begin{cases} \binom{k}{n} \left(\frac{\lambda}{\mu}\right)^n \pi_0 & 0 \le n < c \\ \binom{k}{n} \left(\frac{\lambda}{\mu}\right)^n \frac{n!}{c!\, c^{n-c}} \pi_0 & c \le n \le k \end{cases}$$

y dado que $\sum_{n=0}^{k} \pi_n = 1 \longrightarrow \pi_0 = \left[\sum_{n=0}^{c-1} \binom{k}{n} \left(\frac{\lambda}{\mu}\right)^n + \sum_{n=c}^{k} \binom{k}{n} \left(\frac{\lambda}{\mu}\right)^n \frac{n!}{c!\, c^{n-c}} \right]^{-1}$

Por otra parte, el número medio de máquinas (unidades) en el sistema sería:

$$L = \sum_{n=0}^{k} n \pi_n$$

y en este caso, aunque no se obtiene una expresión sencilla para este parámetro ha de señalarse que para el cálculo de π_0 han debido ser calculados todos los elementos $(k+1)$ que aparecen en la expresión anterior.

Puede obtenerse, además, una expresión que permite relacionar directamente L con el número medio de unidades que esperan ser reparadas, i.e. con L_q:

$$L_q = \sum_{n=c}^{k} (n-c)\pi_n = \sum_{n=c}^{k} n\pi_n - c \sum_{n=c}^{k} \pi_n = L - \sum_{n=0}^{c-1} n\pi_n - c\left[1 - \sum_{n=0}^{c-1} \pi_n\right] = L - c + \sum_{n=0}^{c-1} (c-n)\pi_n$$

$$L_q = L - c + \pi_0 \sum_{n=0}^{c-1} \binom{k}{n} (c-n) \left(\frac{\lambda}{\mu}\right)^n$$

Si queremos aplicar la fórmula de Little para determinar los valores de $E(W)$ y $E(T)$, hemos de calcular previamente la tasa media de llegadas al sistema. Nótese que si hay n unidades en el sistema, sólo pueden ingresar en el mismo $k-n$ unidades, cada una de ellas con tasa instantánea λ, luego la tasa media de llegadas al sistema sería $\lambda(k-n)$ y sumando para cada n tendríamos:

$$\lambda' = \sum_{n=0}^{k} (k-n)\lambda \pi_n = k\lambda \sum_{n=0}^{k} \pi_n - \lambda \sum_{n=0}^{k} n\pi_n = \lambda(k-L)$$

La ecuación $\lambda' = \lambda(k-L)$ es bastante intuitiva; dado que hay en término medio L unidades en el sistema, fuera de él habrá en media $k-L$

unidades, que tienen todas ellas tasa instantánea de llegada λ, luego la tasa media de llegada sería $\lambda(K-L)$.

En consecuencia, al aplicar la fórmula de Little tendríamos:

$$E(W) = L_q / \lambda' = \frac{L_q}{\lambda(K-L)}$$

$$E(T) = L / \lambda' = \frac{L}{\lambda(K-L)}$$

4. Salida de una cola $M/M/m/\infty$: Colas en serie

Consideremos un sistema formado por una serie de estaciones de servicio, en el que hay que ir progresando de una estación a la siguiente hasta recorrerlas todas en forma ordenada (colas en serie). Como ejemplo de tales sistemas podemos considerar una línea de proceso de un producto manufacturado, un chequeo médico (análisis de laboratorio, electrocardiograma, rayos X, ...), etc. (*)

Una cuestión importante relacionada con estos sistemas es el estudio de la distribución de la salida de una cola que determina el régimen de ingreso en la cola siguiente. Se ha demostrado (* Burke, 1968) la independencia de los intervalos de tiempo entre salidas para el sistema $M/M/m/\infty$, en régimen estacionario. Para este sistema las salidas siguen un proceso de Poisson con el mismo parámetro que la distribución de entrada.

Si consideramos un modelo de colas en serie, con sala de espera entre estaciones de servicio sin restricción de capacidad, donde las llegadas se producen siguiendo un proceso de Poisson de parámetro λ, y donde el tiempo de servicio de cada servidor en la estación i-ésima ($i=1$—n) es exponencial con media $1/\mu_i$. Puede verse fácilmente que, dado que no hay restricciones sobre la espera entre estaciones, cada estación puede ser analizada separadamente como una cola sencilla.

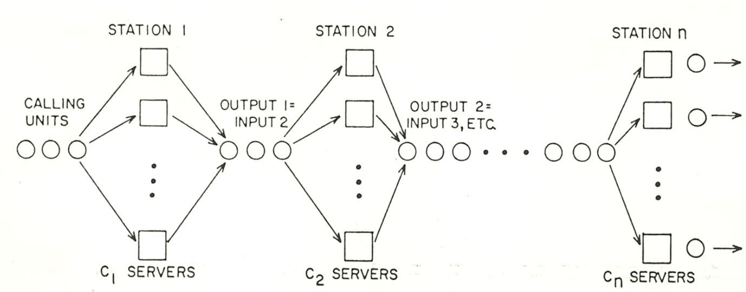


Fig. 4.3 Series queue, infinite waiting room

La primera estación es una cola $M/M/c_1/\infty$ y necesitamos encontrar la distribución de salida para determinar el régimen de llegada a la siguiente estación. Demostraremos que el tiempo entre salidas de una cola $M/M/m/\infty$ en régimen estacionario es idéntico al tiempo entre llegadas necesarias (en su respectiva distribución de probabilidad), i.e. exponencial con

(* Burke P.J. (1968): Ann. Math. Statist. 39, pp 1144-1152.)

(*) puede verse en Gross-Harris (p. 204) otro tipo de colas en serie, aquellas que al no permitir colas entre las estaciones de servicio se bloquean (series queues with blocking) media $1/\lambda$, y, por consiguiente todas las relaciones anteriores valen sólo M/M/c/∞ (λ, μ).

Consideremos una cola M/M/m/∞ en estado de equilibrio (régimen estacionario). Sea $N(t)$ el número de usuarios en el sistema después de transcurrir el tiempo t ($t\geq 0$) desde la última salida, dado que estamos en el estado de equilibrio.

$$t\geq 0: \quad P(N(t)=n) = \pi_n = \begin{cases} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n \pi_0 & n<m \\ \frac{m^m}{m!}\left(\frac{\lambda}{m\mu}\right)^n \pi_0 & n\geq m \end{cases}$$

Sea T_s la variable aleatoria que representa el tiempo entre salidas sucesivas, y

$$F_n(t) = P\{N(t)=n \text{ y } T_s>t\} \quad \xrightarrow{\text{prob. total}} \quad P(T_s>t) = \sum_{n=0}^{\infty} P(T_s>t,\ N(t)=n) = \sum_{n=0}^{\infty} F_n(t)$$

es decir, $F_n(t)$ es la probabilidad conjunta de que haya n usuarios en el sistema en el instante t, desde la última salida, y que t sea menor que el tiempo entre dos salidas sucesivas, esto es, no ha ocurrido ninguna otra salida. Definimos, $t\geq 0$

$$C(t) = P(T_s\leq t) = 1 - P(T_s>t) = 1 - \sum_{n=0}^{\infty} F_n(t)$$

así pues, para conocer la distribución de probabilidad que corresponde a T_s hemos de encontrar primero $F_n(t)$.

Podemos formular las siguientes ecuaciones en diferencias que conciernen a $F_n(t)$:

para M/M/m/∞: $\lambda_n=\lambda$, $\mu_n = \begin{cases} n\mu & 0\leq n\leq m \\ m\mu & n\geq m \end{cases}$

$$\begin{cases} F_n(t+\Delta t) = (1-\lambda\Delta t)(1-m\mu\Delta t)F_n(t) + \lambda\Delta t(1-m\mu\Delta t)F_{n-1}(t) + o(\Delta t) & m\leq n \\ F_n(t+\Delta t) = (1-\lambda\Delta t)(1-n\mu\Delta t)F_n(t) + \lambda\Delta t(1-(n-1)\mu\Delta t)F_{n-1}(t) + o(\Delta t) & 1\leq n\leq m \\ F_0(t+\Delta t) = (1-\lambda\Delta t)F_0(t) + o(\Delta t) \end{cases}$$

procediendo análogamente a como hicimos en otros casos, cuando $\Delta t\to 0$:

$$\left.\begin{array}{ll} \frac{dF_n(t)}{dt} = -(\lambda+m\mu)F_n(t) + \lambda F_{n-1}(t) & m\leq n \\ \frac{dF_n(t)}{dt} = -(\lambda+n\mu)F_n(t) + \lambda F_{n-1}(t) & 1\leq n\leq m \\ \frac{dF_0(t)}{dt} = -\lambda F_0(t) & \end{array}\right\}$$

con la condición de contorno $F_n(0) = P\{N(0)=n,\ T_s>0\} = P\{N(t)=n\} = \pi_n$.

Resolvemos, por inducción, el sistema de ecuaciones anterior:

$$\frac{dF_0(t)}{dt} = -\lambda F_0(t) \longrightarrow \left.\begin{array}{l} F_0(t) = C_0 e^{-\lambda t} \\ F_0(0) = \pi_0 \end{array}\right\} C_0 = \pi_0 \longrightarrow F_0(t) = \pi_0 e^{-\lambda t}$$

Para $n=1$:

$$\frac{dF_1(t)}{dt} + (\lambda+\mu) F_1(t) = \lambda F_0(t) = \lambda \pi_0 e^{-\lambda t} \qquad \cdot\, e^{(\lambda+\mu)t}$$

$$\frac{d}{dt}\left[F_1(t) e^{(\lambda+\mu)t} \right] = \lambda \pi_0 e^{\mu t} = \frac{d}{dt}\left[\frac{\lambda \pi_0}{\mu} e^{\mu t} + C_1 \right] \longrightarrow F_1(t) = \frac{\lambda \pi_0}{\mu} e^{-\lambda t} + C_1 e^{-(\lambda+\mu)t}$$

Ahora bien, sabemos que

$$\pi_n = \frac{(m\rho)^n}{n!}\pi_0 \quad \text{con} \quad \rho = \frac{\lambda}{m\mu} < 1 \quad \text{para} \quad 1 \le n \le m$$

es decir $\pi_n = \dfrac{\lambda^n}{\mu^n n!}\pi_0 \longrightarrow \pi_1 = \dfrac{\lambda}{\mu}\pi_0$, con lo que:

$$\left.\begin{array}{l} F_1(t) = \pi_1 e^{-\lambda t} + C_1 e^{-(\lambda+\mu)t} \\ F_1(0) = \pi_1 + C_1 = \pi_1 \longrightarrow C_1 = 0 \end{array}\right\} F_1(t) = \pi_1 e^{-\lambda t}$$

Supongamos que es cierto para $n-1 \le m-1$: $F_{n-1}(t) = \pi_{n-1} e^{-\lambda t}$, entonces:

$$F_n'(t) + (\lambda + n\mu) F_n(t) = \lambda \pi_{n-1} e^{-\lambda t} \qquad \cdot\, e^{(\lambda+n\mu)t}$$

luego $$\frac{d}{dt}\left[F_n(t) e^{(\lambda+n\mu)t} \right] = \lambda \pi_{n-1} e^{n\mu t} = \frac{d}{dt}\left[\frac{\lambda \pi_{n-1}}{n\mu} e^{n\mu t} + C_n \right]$$

con lo que

$$\left.\begin{array}{l} F_n(t) = \dfrac{\lambda}{n\mu}\pi_{n-1} e^{-\lambda t} + C_n e^{-(\lambda+n\mu)t} \\ \pi_n = \dfrac{\lambda}{n\mu}\pi_{n-1} \qquad F_n(0) = \pi_n + C_n = \pi_n \longrightarrow C_n = 0 \end{array}\right\} F_n(t) = \pi_n e^{-\lambda t} \qquad 1 \le n \le m$$

Teniendo en cuenta que, para $n > m$, la probabilidad de equilibrio son

$$\pi_n = \rho^{n-m}\pi_m = \rho\, \pi_{n-1} = \frac{\lambda}{m\mu}\pi_{n-1}$$

y dado que $F_n'(t) + (\lambda + m\mu) F_n(t) = \lambda F_{n-1}(t)$ $\quad n \ge m$, es inmediato que

$F_n(t) = \pi_n e^{-\lambda t}$, también para $n > m$.

Luego : $P(T_s > t) = \sum_{n=0}^{\infty} F_n(t) = \sum_{n=0}^{\infty} \pi_n e^{-\lambda t} = e^{-\lambda t} \sum_{n=0}^{\infty} \pi_n = e^{-\lambda t}$, y , en consecuencia :

$$C(t) = P(T_s \leq t) = 1 - e^{-\lambda t} \qquad t \geq 0 \longrightarrow T_s \stackrel{D}{=} \exp(\lambda)$$

Hemos demostrado que los tiempos entre salidas sucesivas siguen una distribución exponencial con media $1/\lambda$, idéntica a la distribución que siguen los tiempos entre llegadas sucesivas a la cola.

Es también cierto que las variables $N(t)$ y T_s son independientes[(*)] y que los tiempos entre salidas sucesivas $(T_1, T_2, \ldots)$ de una misma cola son mutuamente independientes.

Ejemplo : (Gross-Harris p. 201)

Una cadena de supermercados quiere cambiar su funcionamiento y como experiencia ha remodelado una de sus tiendas como sigue. Cuando un cliente completa su compra entra en una sala y si todas las cajas están ocupadas recibe un número, y se sienta. Cuando una caja queda libre se llama al número siguiente y ese cliente empieza a ser atendido. Supondremos que no hay restricción de capacidad en la sala de espera.

Se ha estimado, en periodos de demanda máxima, que los clientes llegan al supermercado de acuerdo con un p.P. de media 40 / hora, tardando un tiempo medio de 3/4 hora en llenar su carro, tiempo que sigue una distribución exponencial. Además el tiempo de servicio de las cajas es exponencial de media 4 minutos.

Queremos saber :

(i) número mínimo de cajas operando en periodos de máxima demanda

(ii) si decidimos añadir 1 caja al número mínimo anterior ¿cuál es el tiempo medio de espera en la sala? ¿cuál es el nº medio de personas en la sala? y ¿cuál es el nº medio de personas en todo el supermercado?

(i) Podemos modelar esta situación como una cola en serie con dos estaciones. La primera es la parte de la tienda en donde se efectúan las compras en régimen de autoservicio, luego se tratará de una cola $M/M/\infty$ con $\lambda = 40$ y $\mu = 4/3$. La segunda estación es la sala de espera : $M/M/c$, y en régimen de equilibrio sabemos que las llegadas a esa cola son Poisson de media 40 por hora. Ahora bien, para que se alcance el equilibrio $\rho < 1$, siendo $\rho = \frac{\lambda}{\mu c}$ y $\frac{1}{\mu} = \frac{4}{60} = \frac{1}{15} \rightarrow \mu = 15$, entonces $c > \frac{\lambda}{\mu} = \frac{40}{15} \simeq 2.67 \longrightarrow c_{min} = 3$

(*) Dist. conjunta = producto de distribuciones marginales

(ii) Si decidimos tener 4 cajas operando, la segunda situación será una cola M/M/4/∞ y hemos de calcular para esa cola el tiempo medio de espera: E(W) y el tamaño medio de la cola: E(Q).

Calculemos, en primer lugar, la distribución de equilibrio: $c=4$

$$\pi_0 = \left\{ \sum_{n=0}^{3} \frac{1}{n!}(4\rho)^n + \frac{(4\rho)^4}{4!\,(1-\rho)} \right\}^{-1} \qquad \text{con } \rho = \frac{\lambda}{4\mu} = \frac{40}{60} = 2/3 \rightarrow 4\rho = 8/3 = r = \frac{\lambda}{\mu} \qquad r/c = \rho$$

luego $$\pi_0 = \left\{ \sum_{n=0}^{3} \frac{1}{n!}(8/3)^n + \frac{1}{4!}(8/3)^4 \cdot 3 \right\}^{-1} = 0.06$$

siendo
$$\left.\begin{array}{ll} \pi_n = \dfrac{1}{n!} r^n \pi_0 & 1 \le n \le c \\[2ex] \pi_n = \dfrac{1}{c!\, c^{n-c}} r^n \pi_0 & n > c \end{array}\right\}$$

Si definimos $L = E(Q)$, el número medio de individuos que están en el sistema en régimen de equilibrio (incluyendo los que están siendo atendidos), entonces
$$L = E(Q) = \sum_{n=0}^{\infty} n\,\pi_n$$

Si llamamos L_q al número medio de personas en la cola $\equiv$ sala de espera, en régimen de equilibrio (sin contar los que están siendo atendidos), tendríamos que

$$L_q = 0(\pi_0+\pi_1+\cdots+\pi_c) + 1\,\pi_{c+1} + 2\,\pi_{c+2} + \cdots = \sum_{n=c}^{\infty} (n-c)\,\pi_n$$

Se ha visto que E(W) era el tiempo medio de espera en la cola, hasta que empieza a ser atendido, y dado que las llegadas se producen siguiendo un proceso de Poisson de media λ, tenemos que

$$L_q = \lambda E(W)$$

puesto que L_q es el número medio de personas que han llegado detrás de un individuo mientras éste esperaba para ser atendido. Análogamente, dado que el tiempo de permanencia en la cola es $E(W) + \frac{1}{\mu}$, se tendría que: $L = \lambda E(W) + \lambda/\mu$.(*) Con lo que será suficiente con determinar L_q para la cola M/M/4. Lo haremos en general, para una cola M/M/c/∞ y luego lo calcularemos para $c=4$, $r=8/3$ y $\pi_0 = 0.06$.

(*) es, simplemente, una propiedad ya conocida del pPo.

$$L_q = \sum_{n=c}^{\infty}(n-c)\pi_n = \sum_{n=c}^{\infty}\frac{n}{c!\,c^{n-c}}r^n\pi_0 - \sum_{n=c}^{\infty}\frac{c}{c!\,c^{n-c}}r^n\pi_0$$

ahora bien:

$$\sum_{n=c}^{\infty}\frac{c}{c!\,c^{n-c}}r^n\pi_0 = \frac{\pi_0 c}{c!}r^c\sum_{n=c}^{\infty}\left(\frac{r}{c}\right)^{n-c} \overset{r/c=\rho<1}{=} \frac{\pi_0 c}{c!}r^c\frac{1}{1-r/c} = \frac{\pi_0 c\, r^c}{c!(1-\rho)}$$

$$\frac{\pi_0}{c!}\sum_{n=c}^{\infty}\frac{n r^n}{c^{n-c}} = \frac{\pi_0}{c!}\frac{r^{c+1}}{c}\left\{\underbrace{\sum_{n=c}^{\infty}(n-c)\left(\frac{r}{c}\right)^{n-c-1}}_{\sum_{m=0}^{\infty} m\left(\frac{r}{c}\right)^{m-1}\ (*)} + \underbrace{\sum_{n=c}^{\infty}c\left(\frac{r}{c}\right)^{n-c-1}}_{c^2/r\sum_{m=0}^{\infty}\left(\frac{r}{c}\right)^{m}}\right\} = \frac{\pi_0 r^{c+1}}{c!\,c}\left\{\frac{1}{(1-r/c)^2} + \frac{c^2/r}{1-r/c}\right\}$$

luego

$$L_q = \pi_0\frac{r^{c+1}}{c!\,c}\left\{\frac{1}{(1-\rho)^2} + \frac{c^2/r}{1-\rho} - \frac{c^2/r}{1-\rho}\right\} = \pi_0\frac{r^{c+1}}{c!\,c(1-\rho)^2} = \pi_0\frac{(\lambda/\mu)^{c+1}}{(c-1)!\,(c-\lambda/\mu)^2} = \frac{\pi_0(\lambda/\mu)^c\lambda\mu}{(c-1)!\,(c\mu-\lambda)^2}$$

y dado que $E(W) = L_q/\lambda$:

$$E(W) = \pi_0\frac{(\lambda/\mu)^c\mu}{(c-1)!\,(c\mu-\lambda)^2} \qquad \text{y} \qquad L = \lambda/\mu + L_q \qquad \text{con } L_q = 0.76$$

En nuestro caso:

$$E(W) = 0.06\,\frac{(8/3)^4\,15}{3!\,(60-40)^2} \doteq 0.019\text{ hr} = 1.14 \text{ minutos}$$

$$L = \lambda\left(\frac{1}{\mu} + E(W)\right) = 40\left(\frac{4}{60} + 0.019\right) \doteq 3.44$$ (Nótese que $L < c$, luego el sistema no se congestiona; tengo siempre más servicios del que es necesario)

Sabemos el número de personas en la segunda estación del sistema, pero para la primera estación: cola M/M/1/∞, el tamaño medio de la cola era $\rho = \frac{\lambda}{\mu} = \frac{40}{4/3} = 30$, en consecuencia el nº medio de personas en todo el supermercado será de 33.44, en la hora de demanda máxima, si hay operando 4 cajas. Puede comprobarse que en el caso de 3 cajas el sistema se congestiona. Nótese que $\rho = \frac{\lambda}{3\mu} = 8/9$ y $r = 8/3$, en este caso $\pi_0 = 1/369 \simeq 0.024$ con lo que:

$L_q \simeq 1.82$ $\quad E(W) = 0.045$ hr $= 2.73$ min $\quad$ y $\quad L = 4.48 > 3$ (¡congestión!)

$$(*) \quad \sum_{n=0}^{\infty} n\, x^{n-1} = \frac{d}{dx}\left(\sum_{n=0}^{\infty} x^{n}\right) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2} \qquad \text{para} \quad x<1$$

Capítulo 7

Colas M/G/1

TEMA 7: COLA M/G/1

Introducción

Para el análisis de la cola M/G/1, i.e. aquella para la que el tiempo de servicio sigue una distribución de probabilidad cualquiera, seguiremos un camino diferente al que utilizamos para el análisis de la cola M/M/1 (que consistía en estudiar el proceso subyacente de nacimiento-muerte). El procedimiento utilizado consiste en encajar una cadena de Markov de parámetro discreto, considerando la longitud de la cola sólo en aquellos instantes en que un cliente acaba de completar su servicio.

Además del estudio de las condiciones de ergodicidad y de la distribución de equilibrio del sistema, se estudiará la longitud media de la cola y la distribución del tiempo de espera del cliente.

1. Análisis de ergodicidad

Consideremos un sistema con un solo punto de servicio, en donde las llegadas siguen una distribución de Poisson de parámetro λ (lo que significa que los tiempos entre llegadas son exponenciales: M), y tal que los tiempos de servicio son v.a. independientes, equidistribuidas V, siendo V una v.a. con función de distribución $B(v)$, verificándose que $B(0)=0$, i.e. no existe probabilidad positiva asociada a tiempos negativos, y $E(V) < +\infty$.

Sea $Z(t)$ el proceso estocástico que mide el número de clientes en la cola en el instante t, $t\geq 0$. Se observa el proceso en el momento justo t_n en que el cliente n-ésimo acaba su servicio, obteniéndose un proceso estocástico discreto: $X_0 = Z(0)$, $X_n = Z(t_n)$ y, para $n>0$:

$$X_{n+1} = \begin{cases} X_n - 1 + N & X_n \geq 1 \\ N & X_n = 0 \end{cases}$$

siendo N el número de clientes que llegan a la cola en el intervalo (t_n, t_{n+1}), durante el tiempo de servicio al $(n+1)$-ésimo cliente. Dado que N sólo depende de λ y de la distribución de V, i.e. no depende de la longitud de la cola ni del momento en que empieza el servicio, tendremos que $\{X_n,\ n\geq 0\}$ y $S=\{0,1,2\ldots\}$ es una cadena de Markov.

La distribución de probabilidad de N será:

$$P(N=n)=\int_0^\infty P(N=n\ /\ V=v)\,.\,dB(v)\equiv k_n \qquad n=0,1,2\ldots$$

siendo $P(N=n/\ V=v)=e^{-\lambda v}\dfrac{(\lambda v)^n}{n!}$

dado que el número de individuos que llegan a una cola en un tiempo v es una v.a. $Po(\lambda v)$.

Veamos, pues, cuál es la matriz P de probabilidades de transición de un paso para la cadena X_n:

(1) si $i\geq 1 \wedge j\geq 0$

$$p_{ij}=P(X_{n+1}=j\ |\ X_n=i)=P(N=j-i+1)=\begin{cases}0 & \text{si } j<i-1\\[2mm] \displaystyle\int_0^\infty e^{-\lambda v}\frac{(\lambda v)^{j-i+1}}{(j-i+1)!}\,dB(v) & \text{si } j\geq i-1\end{cases}$$

(2) si $i=0$, se repite la fila $i=1$

Puesto que hemos llamado $k_n=\displaystyle\int_0^\infty e^{-\lambda v}\frac{(\lambda v)^n}{n!}\,dBv$, $n\geq 0$, tenemos que $p_{ij}=k_{j-i+1}$, si $i\geq 1$, de modo que:

$$P=\begin{array}{c|ccccccc} & 0 & 1 & 2 & 3 & 4 & & \\ \hline 0 & k_0 & k_1 & k_2 & k_3 & k_4 & & \\ 1 & k_0 & k_1 & k_2 & k_3 & k_4 & \ldots & \\ 2 & 0 & k_0 & k_1 & k_2 & k_3 & \ldots & \\ 3 & 0 & 0 & k_0 & k_1 & k_2 & \ldots & \\ & \ldots & \ldots & \ldots & \ldots & \ldots & & \end{array}$$

Se trata de una cadena irreducible, aperiódica. Veamos en qué condiciones es recurrente positiva, calcularemos $\pi=\pi P$:

$$\pi_i = \pi_0 k_i + \sum_{r=1}^{i+1} \pi_r k_{i-r+1} \qquad i = 0, 1, 2 \ldots \qquad (1)$$

Consideremos la función generatriz de probabilidades para π_i y k_i:

$$\Pi(s) = \sum_{i=0}^{\infty} \pi_i s^i \qquad \text{y} \qquad K(s) = \sum_{i=0}^{\infty} k_i s^i \qquad s \leq 1$$

Multiplicando (1), miembro a miembro, por s^i y sumando, obtenemos:

$$\Pi(s) = \pi_0 K(s) + \sum_{i=0}^{\infty} \left(\sum_{r=0}^{i+1} \pi_r k_{i-r+1} \right) s^i - \sum_{i=0}^{\infty} \pi_0 k_{i+1} s^i$$

ahora bien:

$$\sum_{i=0}^{\infty} \left(\sum_{r=0}^{i+1} \pi_r k_{i-r+1} \right) s^i = \underbrace{(\pi_0 k_1 + \pi_1 k_0)}_{i=0} + \underbrace{(\pi_0 k_2 + \pi_1 k_1 + \pi_2 k_0)}_{i=1} s +$$

$$+ \underbrace{(\pi_0 k_3 + \pi_1 k_2 + \pi_2 k_1 + \pi_3 k_0)}_{i=2} s^2 + \ldots$$

$$= \frac{1}{s} \pi_0 (k_1 s + k_2 s^2 + k_3 s^3 + \ldots) + \frac{1}{s} \pi_1 s \{k_0 + k_1 s + k_2 s^2 + \ldots\} + \frac{1}{s} \pi_2 s^2 \{k_0 + k_1 s + k_2 s^2 + \ldots\} +$$

$$+ \ldots = \frac{1}{s} \Pi(s) K(s) - \frac{1}{s} \pi_0 k_0$$

luego:

$$\Pi(s) = \pi_0 K(s) + \frac{1}{s} \Pi(s) K(s) - \frac{1}{s} \pi_0 k_0 - \sum_{i=0}^{\infty} \pi_0 k_{i+1} s^i = \pi_0 K(s) + \frac{1}{s} \Pi(s) K(s) - \frac{1}{s} \pi_0 K(s)$$

de donde $s\Pi(s) = s\pi_0 K(s) + \Pi(s) K(s) - \pi_0 K(s) \longrightarrow \Pi(s) = \dfrac{\pi_0 K(s)(s-1)}{s - K(s)} \quad (2) \quad s \leq 1 \quad (*)$

Nótese que $K'(1) = \lim_{s \to 1^-} \dfrac{K(1) - K(s)}{1-s} = \lim_{s \to 1^-} \dfrac{1 - K(s)}{1-s} = E(N) = \rho$

Llamamos $\mathcal{C}(s) = \dfrac{1 - K(s)}{1-s} = \{1 - k_0 - k_1 s - k_2 s^2 - \ldots\}\{1 + s + s^2 + s^3 + \ldots\} =$

$$\left.\begin{array}{l} = 1 - k_0 - k_1 s - k_2 s^2 - k_3 s^3 \ldots \\ \quad s - k_0 s - k_1 s^2 - k_2 s^3 - \ldots \\ \quad s^2 - k_0 s^2 - k_1 s^3 - \ldots \\ \quad s^3 - k_0 s^3 - \ldots \end{array}\right\} \sum$$

$$= (1 - k_0) + (1 - k_0 - k_1) s + (1 - k_0 - k_1 - k_2) s^2 + (1 - k_0 - k_1 - k_2 - k_3) s^3 + \ldots = \sum_{j=0}^{\infty} s^j (k_{j+1} + k_{j+2} \ldots) =$$

(*) Sabemos que $K(1) = \Pi(1) = 1$, luego

$$\lim_{s \to 1^-} \Pi(s) = 1 = \pi_0 \lim_{s \to 1^-} \frac{sK(s) - K(s)}{s - K(s)} = \pi_0 \lim_{s \to 1^-} \frac{K(s) + sK'(s) - K'(s)}{1 - K'(s)} = \pi_0 \frac{1 + K'(1) - K'(1)}{1 - K'(1)} = \pi_0 \frac{1}{1 - K'(1)}$$

pero $K'(1) = E(N) = \sum_{n=0}^{\infty} n \, \alpha_n = \lambda E(V) = \frac{E(V)}{1/\lambda} = \rho$

luego $1 = \pi_0 \frac{1}{1-\rho} \longrightarrow \pi_0 = 1 - \rho \quad \wedge \quad \rho < 1$

entonces

$$\Pi(s) = (1-\rho) \frac{K(s)\,(s-1)}{s - K(s)} \qquad s \leq 1$$

$$\mathcal{C}(s) = \sum_{j=0}^{\infty} \left(\sum_{n=j+1}^{\infty} k_n \right) s^j$$

teniéndose que: $\mathcal{C}(1) = \sum_{j=0}^{\infty} \sum_{n=j+1}^{\infty} k_n = \sum_{n=0}^{\infty} n k_n = \rho$, media de la distribución de las llegadas durante un servicio. Sustituyendo k_n:

$$\rho = \int_0^{\infty} e^{-\lambda v} \left(\sum_{n=0}^{\infty} n \frac{(\lambda v)^n}{n!} \right) dB(v) = \int_0^{\infty} \lambda v \, dB(v) = \lambda E(V)$$

Asimismo, de (2) y de la definición de $\mathcal{C}(s)$: $\Pi(s) = \dfrac{\pi_0 K(s)}{1-\mathcal{C}(s)}$, y para $s=1$:

$$\Pi(1) = 1 , \quad K(1) = 1 \longrightarrow 1 = \frac{\pi_0}{1-\rho} \longrightarrow \pi_0 = 1-\rho$$

y como π_0 ha de ser positivo $\longrightarrow \rho < 1$.

La condición de ergodicidad — $\underset{\text{intensidad del tráfico}}{\rho} = \dfrac{E(V)}{1/\lambda} = \dfrac{\text{tiempo medio de servicio}}{\text{tiempo medio entre llegadas}} < 1$

Para obtener la distribución de equilibrio, podemos derivar en $s=0$ la función generatriz de probabilidades $\Pi(s)$, que representamos:

$$\Pi(s) = (1-\rho) \frac{K(s)\,(s-1)}{s-K(s)} \qquad s \leq 1$$

2. Servicio en fases : cola $M/E_k/1$

Nótese que la distribución de Erlang E_k es la distribución de la suma de k v.a. independientes, equidistribuidas exponencialmente μ ($\Gamma(k,\mu)$). Esta distribución permitirá, pues, describir modelos de colas en los que el servicio consta de una serie de fases idénticamente distribuidas e independientes, que deben ser completadas antes de dar servicio a un nuevo cliente. Como ejemplo de tales situaciones podemos considerar los análisis de laboratorio, la revisión de automóviles en un taller, etc. realizadas por un solo operario, y siempre que pueda admitirse que las fases se distribuyen exponencialmente con la misma media.

Supongamos que la distribución de los tiempos de servicio es una Erlang de orden k, correspondiente a una gamma $\Gamma(k, \mu k)$, cuya función de densidad es:

$$b(v) = \frac{(\mu k)^k}{(k-1)!} v^{k-1} e^{-\mu k v} \qquad v > 0$$

lo que significa que tenemos una suma de k v.a. iid $\exp(\mu k)$, con tiempo medio de servicio por fase de $\frac{1}{\mu k}$, en consecuencia el tiempo medio de servicio, total, será de $\frac{1}{\mu} = E(v)$.

Vamos a calcular los valores de k_n:

$$k_n = \int_0^\infty e^{-\lambda v} \frac{(\lambda v)^n}{n!} b(v)\, dv = \int_0^\infty e^{-\lambda v} \frac{(\lambda v)^n}{n!} \frac{(\mu k)^k}{(k-1)!} v^{k-1} e^{-\mu k v}\, dv =$$

$$= \int_0^\infty e^{-(\lambda + \mu k) v} \frac{(\lambda v)^n}{n!} \frac{(\mu k)^k}{(k-1)!} v^{k-1}\, dv \qquad n \geq 0$$

en consecuencia, para $s \leq 1$:

$$k(s) = \sum_{n=0}^\infty k_n s^n = \int_0^\infty e^{-(\lambda+\mu k) v} \frac{(\mu k)^k}{(k-1)!} v^{k-1} \left(\sum_{n=0}^\infty \frac{[\lambda v s]^n}{n!} \right) dv =$$

$$= \frac{(\mu k)^k}{(k-1)!} \int_0^\infty \underbrace{e^{-[\lambda(1-s)+\mu k] v} v^{k-1}\, dv}_{e^{-x} x^{k-1} dx} = \frac{(\mu k)^k}{(k-1)!} \frac{\Gamma(k)}{[\lambda(1-s)+\mu k]^k} = \frac{1}{\left[\frac{\lambda}{\mu k}(1-s)+1\right]^k}$$

Existirá distribución de equilibrio si:

$$\rho = \frac{E(v)}{1/\lambda} = \frac{1/\mu}{1/\lambda} = \frac{\lambda}{\mu} < 1$$

siendo la función generatriz de las probabilidades límite:

$$\Pi(s) = (1-\rho) \frac{k(s)(s-1)}{s - k(s)} = (1-\rho) \frac{s-1}{\left[\frac{\lambda}{\mu k}((1-s)+1)\right]^k s - 1}$$

Nota. El desarrollo en serie de MacLaurin es inmediato. Puede intentarse para $M/E_3/1$ con $\lambda = 6$ y $\mu = 8$, y calcular la probabilidad de que no haya más de 2 clientes en el sistema. (ejercicio 2)

3. Longitud media de la cola en régimen estacionario: fórmula de Pollaczek-Khintchine

Supongamos que el sistema ha alcanzado el régimen de equilibrio, $\rho < 1$. Sea Q el número de clientes en la cola, después de la partida de un cliente que ha sido servido. Vamos a determinar $E(Q) = L$.

Llamaremos Q' al número de clientes en la cola después de la partida siguiente, en consecuencia:

$$Q' = Q - 1 + N + \delta$$

siendo N las llegadas ocurridas durante la duración de un servicio y

$$\delta = \begin{cases} 1 & \text{si } Q = 0 \\ 0 & \text{si } Q > 0 \end{cases}, \text{ i.e. } \delta \triangleq \delta(Q) \longrightarrow Q' = \begin{cases} N & Q = 0 \\ Q - 1 + N & Q \geq 1 \end{cases}$$

En régimen estacionario, Q y Q' han de tener la misma distribución, lo que implica que $E(Q) = E(Q')$, luego

$$E(Q') = E(Q) - 1 + E(\delta) + E(N) \longrightarrow E[\delta] = 1 - E(N) = 1 - \rho = 1 - \lambda E(V) = \pi_0$$

Es inmediato que $\delta = \delta^2$, lo que supone que:

$$(Q')^2 = Q^2 + \delta + (N-1)^2 + 2Q\delta + 2\delta(N-1) + 2Q(N-1)$$

y dado que $Q\delta = 0$, resulta que:

$$(Q')^2 = Q^2 + \delta + N(N-1) + (1-N) + 2\delta(N-1) + 2Q(N-1)$$

Ahora bien, N es independiente de Q, y en consecuencia es independiente de δ, lo que conduce a:

$$E[(Q')^2] = E[Q^2] + 1 - \rho + E[N(N-1)] + 1 - \rho + 2(1-\rho)(\rho - 1) + 2E(Q)(\rho - 1)$$

y $E[(Q')^2] = E[Q^2] \longrightarrow 0 = 1 - \rho + E[N(N-1)] + 1 - \rho + 2(2\rho - \rho^2 - 1) + 2E(Q)(\rho - 1)$

$$0 = \underbrace{2\rho - 2\rho^2}_{2\rho(1-\rho)} + E[N(N-1)] + 2E(Q)(\rho - 1)$$

luego $E(Q) = \rho + \dfrac{E(N(N-1))}{2(1-\rho)}$

Recordemos que $k_n = P(N=n)$, siendo

$$K(s) = \sum_{n=0}^{\infty} k_n s^n = \int_0^{\infty} e^{-\lambda v(1-s)}\, dB(v) \longrightarrow K''(s) = \lambda^2 \int_0^{\infty} v^2\, e^{-\lambda v(1-s)}\, dB(v)$$

además $E[N(N-1)] = \sum_{n=2}^{\infty} n(n-1)\, k_n = K''(1) = \lambda^2 \int_0^{\infty} v^2\, dB(v) = \lambda^2 [\mathrm{Var}\, V + E^2(V)] = \mathrm{Var}(\lambda V) + \rho^2$

$\rho = \lambda E(V)$

luego $E[N(N-1)] = \mathrm{Var}[\lambda V] + \rho^2$, y sustituyendo:

$$L = E(Q) = \rho + \frac{\mathrm{Var}[\lambda V] + \rho^2}{2(1-\rho)} \qquad (3)$$

que se conoce como fórmula de Pollaczek-Khintchine, y que permite conocer la longitud media de la cola (cuando un cliente la abandona) en régimen estacionario, sin necesidad de determinar los valores de la distribución de equilibrio. Nótese que $E(Q) = \sum_{n=0}^{\infty} n\, \pi_n$.

4. Fórmula de Little

Vamos a determinar el valor medio del tiempo de espera de un cliente en una cola M|G|1, y lo haremos utilizando la llamada fórmula de Little.

Supongamos que un cliente espera un tiempo W hasta que empieza a ser servido, durando su servicio una v.a. V. Si al partir este cliente deja Q individuos en la cola, sabemos que estos han llegado durante el tiempo W+V, siguiendo un proceso de llegadas que es Poisson de tasa instantánea λ.

En régimen estacionario: $E(Q) = \lambda E[W+V] = \lambda[E(W) + E(V)]$, lo que se interpreta diciendo que el tiempo medio de permanencia en la cola debe ser igual al número medio de clientes que llegan en ese período, L, por la longitud media del tiempo entre llegadas: $1/\lambda\, E[Q] = E(W+V)$, y se conoce como fórmula de Little (es una fórmula válida para una gran variedad de colas).

lo que se está diciendo (y esto es una demostración intuitiva de la fórmula de Little) es que cuando llega un cliente encontrará en el sistema el mismo número medio de individuos L que cuando lo abandona después de ser atendido, dado que habíamos supuesto que el sistema se encontraba en equilibrio.

Este resultado no involucra ninguna hipótesis específica acerca de la distribución de los tiempos entre llegadas sucesivas o de los tiempos de servicio, ni depende del número de canales de servicio o de la disciplina de la cola, y se aceptó como válido durante muchos años. El primero en dar una demostración formal del mismo fue Little (1961) y, posteriormente, Jewel (1967) hizo algunas simplificaciones.

Asimismo, se designa como fórmula de Little una versión particular de este resultado, que es la siguiente:

$$L_q = \lambda E(W)$$

siendo L_q el número medio de clientes en la sala de espera. Dado que:

$L = L_q + L_s$ y $E(T) = E(W) + E(V)$,

donde L_s es el número medio de clientes recibiendo servicio, se sigue también que:

$$L_s = \lambda E(V),$$

siempre que sean ciertas las fórmulas de Little.

Para la cola M/M/1 se calcularon $E(W)$ y L y se comprobó que $L = \lambda\,[E(W) + 1/\mu]$, i.e. se cumple la fórmula de Little. Para colas M/G/1 la demostración del resultado podría ser la siguiente: En régimen de equilibrio, para este tipo de colas, la probabilidad de que haya n clientes cuando uno abandona el sistema es la de que hayan llegado n clientes durante el tiempo de permanencia en la cola. Luego, sea $T = W+V$ y $F(t)$ la función de distribución del tiempo de permanencia en la cola, entonces por la ley de probabilidad total:

$$\pi_n = \int_0^\infty \underbrace{P(\text{lleguen } n \text{ durante un tiempo } T / T=t)}_{p.Po(\lambda t)} \, . \, dF(t)$$

y para $n \geq 0$:

$$\pi_n = \frac{1}{n!}\int_0^\infty (\lambda t)^n e^{-\lambda t} \, dF(t)$$

dado que

$$L = \sum_{n=0}^{\infty} n\pi_n = \sum_{n=1}^{\infty} \frac{n}{n!} \int_0^{\infty} (\lambda t)^n e^{-\lambda t}\, dF(t) = \int_0^{\infty} dF(t)\, e^{-\lambda t} \sum_{n=1}^{\infty} \frac{(\lambda t)^n}{(n-1)!} = \int_0^{\infty} dF(t)\, e^{-\lambda t} \lambda t\, e^{\lambda t} =$$

$$= \lambda \int_0^{\infty} t\, dF(t) = \lambda E(T) \qquad \longleftrightarrow \qquad L = \lambda E(T)$$

Nótese que sólo hemos utilizado el hecho de que las llegadas siguen un proceso de Poisson λ, cualquiera que sea la distribución del tiempo de servicio, con lo que esta demostración también validaría, para $M|G|c$, el siguiente resultado:

$L_q = \lambda E(W)$

i.e. cuando las llegadas son markovianas y cualquiera que sea el número de canales de servicio. Pero ya se ha dicho que se cumple asimismo en condiciones más generales.

Nota. Un procedimiento para asegurarse de que una clase determinada de modelos de colas satisfacen la fórmula de Little sería el de determinar L como esperanza de la distribución estacionaria $\{\pi_n\}_{n=0,1,\dots}$ y obtener una expresión explícita de la distribución de probabilidad de los tiempos de espera en la cola W, de modo que al evaluar $E(W)$ puede comprobarse si se satisface la ecuación $L = \lambda[E(W) + E(V)]$.

Basándonos en el argumento utilizado anteriormente para demostrar la validez de la fórmula de Little para la cola $M|G|1$, podemos encontrar la transformada de Laplace-Stieltjes de la distribución de los tiempos de espera. Se ha visto que:

$$\pi_n = \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!}\, dF(t) \longrightarrow \Pi(s) = \sum_{n=0}^{\infty} \pi_n s^n = \sum_{n=0}^{\infty} \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t s)^n}{n!}\, dF(t) = \int_0^{\infty} e^{-\lambda t(1-s)}\, dF(t)$$

pero esta última expresión es la transformada de L-S de la función de distribución $F(\cdot)$ de los tiempos de permanencia T en el sistema, así que:

$\Pi(s) = F^*(\lambda(1-s))$

pero $T = W + V$ es suma de 2 variables aleatorias independientes, con lo que la transformada de su función de distribución es el producto de las transformadas de las funciones de distribución de W y V, i.e.

$$\Pi(s) = F^*(\lambda(1-s)) = W^*(\lambda(1-s))\, B^*(\lambda(1-s)) \longrightarrow W^*(z) = \frac{\Pi((\lambda - z)/\lambda)}{B^*(z)}$$

En nuestro caso, y dado que la fórmula de Pollaczek-Khintchine nos habrá permitido calcular $E(Q)$, tendremos:

$$\lambda E(W) + \lambda E(V) = \lambda E(W) + \rho = \rho + \frac{Var(\lambda V) + \rho^2}{2(1-\rho)} = \rho + \frac{\lambda^2 Var(V) + \rho^2}{2(1-\rho)}$$

luego
$$E(W) = \frac{Var(\lambda V) + \rho^2}{2\lambda(1-\rho)} \qquad (4)$$

que es el tiempo medio de espera para un cliente en la cola, en régimen estacionario. (*)

<u>Ejemplo</u>: <u>cola $M|D|1$</u>

En este caso el tiempo de servicio es constante, y supongamos que todos duran exactamente $1/\mu$. Sea N el número de personas que llegan a la cola durante un servicio: $N \overset{d}{=} Po(\lambda \frac{1}{\mu})$, siendo λ el parámetro de Poisson para las llegadas a la cola.

Entonces $k_n = P(N=n) = e^{-\lambda/\mu} \dfrac{(\lambda/\mu)^n}{n!}$ $\quad n \geq 0$

llamamos $\rho = \frac{\lambda}{\mu}$, y la condición de ergodicidad es $\rho < 1$. Sabemos que $\rho = \lambda E(V)$ i.e. $\rho = \lambda \frac{1}{\mu}$.

La función generatriz de probabilidad de N será: $Po(\rho)$

$$K(s) = \sum_{i=0}^{\infty} k_i s^i = e^{-\rho(1-s)} \qquad s \leq 1$$

y en consecuencia:

$$\Pi(s) = (1-\rho) \frac{e^{-\rho(1-s)}(s-1)}{s - e^{-\rho(1-s)}} = (1-\rho) \frac{(1-s)}{1 - s e^{\rho(1-s)}}$$

vamos a desarrollar esta función en serie, para calcular las probabilidades de equilibrio, en el caso de que $\rho < 1$:

$$\Pi(s) \underset{(1)}{=} (1-\rho)(1-s) \sum_{k=0}^{\infty} e^{k\rho(1-s)} s^k = (1-\rho)(1-s) \sum_{k=0}^{\infty} e^{k\rho} e^{-k\rho s} s^k = (1-\rho)(1-s) \sum_{k=0}^{\infty} e^{k\rho} \left(\sum_{m=0}^{\infty} \frac{(-k\rho s)^m}{m!} \right) s^k =$$

$s<1$

(1) Dado que $s e^{\rho(1-s)} < 1$ (**)

(*) <u>Nota</u> La fórmula [illegible] y (4) [illegible] un hecho [illegible] [illegible] [illegible] un tiempo medio de servicio $E(V)$ y un tiempo medio entre llegadas [illegible] [illegible] $E(V)<1$, [illegible] [illegible] [illegible] la [illegible] media de [illegible] [illegible] tiempo medio [illegible] [illegible] [illegible] [illegible] [illegible] [illegible], [illegible] [illegible] la [illegible] del tiempo de servicio [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] posible [illegible] [illegible] [illegible] tiempo de servicio constante $\rightarrow$ [illegible] $= 0$.

(**) [illegible] $\rho(1-s) \le 0$

[illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible], cuando [illegible] [illegible], [illegible] [illegible] $\rho \to 0$ y [illegible]

$$= (1-\rho)(1-s) \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} e^{k\rho} (-1)^m \frac{(k\rho)^m}{m!} s^{k+m} = (1-\rho)(1-s) \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} e^{k\rho} (-1)^{n-k} \frac{(k\rho)^{n-k}}{(n-k)!} s^n =$$

$$= (1-\rho)(1-s) \sum_{n=0}^{\infty} \sum_{k=0}^{n} e^{k\rho} (-1)^{n-k} \frac{(k\rho)^{n-k}}{(n-k)!} s^n = (1-\rho) \left\{ \sum_{n=0}^{\infty} \sum_{k=0}^{n} e^{k\rho} \frac{(-k\rho)^{n-k}}{(n-k)!} s^n - \sum_{n=0}^{\infty} \sum_{k=0}^{n} e^{k\rho} \frac{(-k\rho)^{n-k}}{(n-k)!} s^{n+1} \right\}$$

$$= (1-\rho) \left\{ \sum_{n=0}^{\infty} \sum_{k=0}^{n} e^{k\rho} (-1)^{n-k} \frac{(k\rho)^{n-k}}{(n-k)!} s^n - \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} e^{k\rho} (-1)^{n-k-1} \frac{(k\rho)^{n-k-1}}{(n-k-1)!} s^n \right\}$$

ahora ya, por identificación pueden determinarse las probabilidades de equilibrio:

$n=0$ $\qquad \pi_0 = 1-\rho$

$n=1$ $\qquad \pi_1 = (1-\rho)\,(e^{\rho}-1)$

$n \geq 2$ $\qquad \pi_n = (1-\rho) \left\{ \sum_{k=1}^{n} e^{k\rho} (-1)^{n-k} \frac{(k\rho)^{n-k}}{(n-k)!} + \sum_{k=1}^{n-1} e^{k\rho} (-1)^{n-k} \frac{(k\rho)^{n-k-1}}{(n-k-1)!} \right\}$

Por otra parte, en régimen estacionario, la longitud media de la cola, cuando un cliente ha pedido servicio, y el tiempo medio de espera son:

$$E(Q) = \rho + \frac{\rho^2}{2(1-\rho)} = \frac{2\rho - 2\rho^2 + \rho^2}{2(1-\rho)} = \rho\,\frac{1-\rho/2}{1-\rho} = \frac{2\rho-\rho^2}{2(1-\rho)} \qquad Var(V) = 0$$

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

Ejemplo numérico

Los clientes llegan a una tienda según un proceso de Poisson con tasa $\lambda = 2$ por minuto. El tiempo que pasan comprando se distribuye según variables aleatorias independientes e idénticamente distribuidas. Al finalizar su compra acuden con equiprobabilidad a cualquiera de las cajas registradoras, comprobándose que se invierte en el servicio un tiempo medio de 3 minutos, con varianza 9. Se han considerado los costes siguientes:

$C_1 = 0.5$ pts por unidad de tiempo debido a la espera del cliente

$C_2 = 2.5$ pts " " " a los gastos de mantenimiento de la caja

Determinar el número óptimo de cajas registradoras que minimice el coste total esperado del sistema. (Blat p. 236)

$M|M|\infty \xrightarrow{s \text{ colas}} M|G|1$

λ Sea $\hat{s}$ el número de cajas registradoras y W_s el tiempo de espera si hay s cajas. Cada caja se comporta según una cola $M|G|1$, siendo $\lambda_s = 2/s$ /min la tasa instantánea de llegada, en este caso y $E(V) = 3$, el tiempo medio de servicio y $Var(V) = 9$.

Sabemos que $\rho = \lambda_s E(V) = \frac{2}{s} E(V) = 6/s$, luego: $6/s < 1 \Leftrightarrow 6 < s$

$$E(W_s) = \frac{(\lambda/s)^2 \, Var\, V + \rho^2}{2(\lambda/s)(1-\rho)} = \frac{(2/s)^2 \cdot 9 + 36/s^2}{4/s\,(1 - 6/s)} = \frac{2\; 36/s^2}{4/s\,\left(\frac{s-6}{s}\right)} = \frac{18}{s-6}$$

El coste esperado, si hay s cajas será:

$$E(C_s) = \frac{18}{s-6} C_1 + s C_2 = \frac{9}{s-6} + 2.5 s$$

y queremos minimizarlo: $\frac{d}{ds} E(C_s) = 2.5 + \frac{(-9)}{(s-6)^2} = 0 \Longleftrightarrow 2.5(s-6)^2 - 9 = 0$

sii $s = 6 \pm \sqrt{9/2.5} = 6 \pm \sqrt{3.6} = 6 \pm 1.897$ — $s_1 = 7.897$, $s_2 = 4.103$

pero $\rho = 6/s < 1 \longrightarrow \underline{s > 6}$

Elegiremos $[s_1]$ si $E(C_s)\big|_{s=[s_1]} \le E(C_s)\big|_{s=[s_1+1]}$, en caso contrario elegiremos $[s_1+1]$. En nuestro ejemplo:

$E(C)\big|_{s=7} = 26.5$

$E(C)\big|_{s=8} = 24.5 \longrightarrow \boxed{\hat{s} = 8}$

5. Distribución de los tiempos de espera

Ya se vió, al introducir la fórmula de Little, una expresión que involucraba la transformada de Laplace-Stieltjes de la función de distribución de los tiempos de espera: $W(t)$. Para obtener este resultado se ha hecho uso de una identidad que aun siendo cierta para la cola $M/G/1$, no lo es en general. Esa identidad es la siguiente: $\pi_n = p_n$, siendo $p_n = \lim_{t\to\infty} P(Z(t)=n)$.

Nosotros hemos estudiado considerando la longitud de la cola sólo en aquellos instantes en que un cliente acaba de completar su servicio, y para esta cadena de Markov se ha obtenido la distribución estacionaria π_n. Hemos definido $Z(t)$, $t\geq 0$, como el proceso estocástico que mide el tamaño del sistema en un instante arbitrario de tiempo, se trata de ver que: $\pi_n = p_n$.

Consideremos una realización específica del proceso sobre un intervalo de tiempo T suficientemente grande. Para cada estado $n\in S$, sean $A_n(t)$ el nº de llegadas (saltos unitarios hacia adelante) desde el estado $\overset{n\to n+1}{n}$, en el intervalo $(0,t)$ y $D_n(t)$ el nº de salidas (saltos unitarios hacia atrás) hacia $\overset{n+1\to n}{n}$, en $(0,t)$. Dado que las llegadas y salidas se producen singularmente de una en una para el proceso $Z(t)$, para cada n tenemos que
$|A_n(T) - D_n(T)| \leq 1$

Esto ocurre si la intensidad de tráfico es menor que 1, i.e. $\rho<1$, el sistema está insaturado (se alcanzará el equilibrio); y si llamamos $D(T)$ al nº total de salidas y $A(T)$ al nº total de llegadas, cuando T tiende a infinito se cumplirá que:

$$\lim_{T\to\infty} \frac{D(T)}{A(T)} = 1 = \lim_{T\to\infty} \frac{A(T)}{D(T)}, \quad \text{con probabilidad } 1$$

luego

$$\lim_{T\to\infty} \left| \frac{A_n(T)}{A(T)} - \frac{D_n(T)}{A(T)} \right| \leq \lim_{T\to\infty} \frac{1}{A(T)} = 0$$

pues $A(T)$ tenderá a infinito cuando T sea suficientemente grande, así que:

$$\lim_{T\to\infty} \frac{A_n(T)}{A(T)} = \lim_{T\to\infty} \frac{D_n(T)}{A(T)} = \lim_{T\to\infty} \frac{D_n(T)}{A(T)} \cdot \lim_{T\to\infty} \frac{A(T)}{D(T)} = \lim_{T\to\infty} \frac{D_n(T)}{D(T)}$$

Ahora bien, por definición (*): $\pi_n = \lim_{T\to\infty} \frac{D_n(T)}{D(T)}$, y dado que las llegadas ocurren en los [illegible] / según un proceso de Poisson, independientemente del tamaño del sistema / estado del proceso, el límite $\left(\frac{A_n(T)}{A(T)}\right)$

(*) interpretación frecuencial de la probabilidad: $\dfrac{\text{nº de veces que hemos salido del mostrador}}{\text{nº total de salidas}}$, y que $T \to \infty$

implica que estamos en régimen de equilibrio: Tm (prob. calculada en el punto ≡ instante en que un individuo abandona la cola) cuando $T \to \infty$ es idéntico a la probabilidad límite general p_n. Luego $\pi_n = p_n$.

Sabemos, además, que para $F(\cdot)$ función de distribución de los tiempos de permanencia en el sistema:

$$\pi_n = p_n = \frac{1}{n!}\int_0^\infty (\lambda t)^n e^{-\lambda t}\, dF(t) \qquad n \geq 0$$

así, pues, la función generatriz de probabilidad de $\{p_n\}$: $P(z) = \sum_{n=0}^{\infty} p_n z^n = F^*[\lambda(1-z)]$, como ya se había visto; y $F^*(\cdot)$ es la transformada de Laplace-Stieltjes de F.

También habíamos obtenido que:

$$\Pi(s) = F^*(\lambda(1-s)) = P(s)$$

y en consecuencia, del hecho de que $T = W + V \longrightarrow \Pi(s) = F^*(\lambda(1-s)) = W^*(\lambda(1-s))\, B^*(\lambda(1-s))$

$$\longrightarrow W^*(z) = \frac{\Pi(\lambda - z/\lambda)}{B^*(z)} = \frac{P(\lambda - z/\lambda)}{B^*(z)}$$

siendo W^* la transformada de Laplace-Stieltjes de la función de distribución de los tiempos de espera.

6. Servicio en bloque: cola $M/M^{[K]}/1$

Supongamos que las llegadas a un sistema, con un solo servidor y sin restricción de capacidad en la sala de espera, siguen un proceso de Poisson de parámetro λ, y que los clientes son servidos en bloque: K clientes en cada servicio, a menos que haya menos de K en el sistema listos para ser atendidos, en cuyo caso se atienden todos. La cantidad de tiempo requerida para el servicio de un grupo se distribuye exponencialmente (μ), tanto si el grupo consta de K clientes como si consta de menos. Este modelo es conocido con la notación $M/M^{[K]}/1$, y desde luego, este p. est. no es un problema markoviano de nacimiento-muerte.

Las ecuaciones de Chapman-Kolmogorov para este sistema son: $t \geq 0$ $h > 0$

$$p_n(t+h) = (1-(\lambda+\mu)h)p_n(t) + \lambda h p_{n-1}(t) + \mu h p_{n+K}(t) + o(h) \qquad n \geq 1$$

$$p_0(t+h) = (1-\lambda h)p_0(t) + \mu h \sum_{i=1}^{K} p_i(t) + o(h)$$

siendo $p_n(t) = P[N(t) = n] = P$[hay n individuos en el sistema en t], de donde:

$$\begin{cases} p'_n(t) = -(\lambda+\mu)\, p_n(t) + \lambda\, p_{n-1}(t) + \mu\, p_{n+k}(t) & n \geq 1 \\ p'_0(t) = -\lambda\, p_0(t) + \mu \sum_{i=1}^{k} p_i(t) \end{cases}$$

Si suponemos que se satisfacen las condiciones (*) para que el sistema alcance el equilibrio, en ese punto, y como se vió en Tema 5: $p'_n(t) = 0 \quad \forall n \geq 0$, y en consecuencia:

$$\begin{cases} 0 = \mu\, p_{n+k} - (\lambda+\mu)\, p_n + \lambda\, p_{n-1} & n \geq 1 \qquad (1) \\ 0 = \mu\, p_k + \mu\, p_{k-1} + \dots + \mu\, p_1 - \lambda\, p_0 \end{cases}$$

Para resolver este sistema de ecuaciones en diferencias con coef. constantes podemos recurrir a diferentes métodos. Uno de ellos podría ser el de obtener una expresión de la función generatriz de probabilidad de $\{p_n\}$, para ello multiplicamos las ecuaciones primeras (1) por s^{n+k}, para $n \geq 1$ y sumamos, resultando:

$$\mu\, s^{n+k} p_{n+k} = (\lambda+\mu)\, s^{n+k} p_n - \lambda s^{n+k} p_{n-1} \qquad n \geq 1$$

$$\mu \sum_{n=1}^{\infty} s^{n+k} p_{n+k} = (\lambda+\mu) s^k \sum_{n=1}^{\infty} p_n s^n - \lambda s^{k+1} \sum_{n=1}^{\infty} p_{n-1} s^{n-1} = (\lambda+\mu) s^k [P(s) - p_0] - \lambda s^{k+1} P(s)$$

$$\mu \left[P(s) - \sum_{n=0}^{k} s^n p_n \right] = (\lambda+\mu) s^k [P(s) - p_0] - \lambda s^{k+1} P(s)$$

con la condición inicial: $\lambda p_0 = \mu \sum_{i=1}^{k} p_i$, o lo que es lo mismo:

$$P(s) \left[(\lambda+\mu) s^k - \lambda s^{k+1} - \mu \right] = (\lambda+\mu) s^k p_0 - \mu \sum_{n=0}^{k} s^n p_n = \mu s^k \sum_{n=0}^{k} p_n - \mu \sum_{n=0}^{k} s^n p_n = \mu \sum_{n=0}^{k} p_n (s^k - s^n)$$

$$P(s) = \frac{\mu \sum_{n=0}^{k-1} p_n (s^k - s^n)}{(\lambda+\mu) s^k - \lambda s^{k+1} - \mu} \qquad 0 \leq s < 1$$

Pero este resultado dice poco acerca de cómo son las probabilidades p_n, tamaño del sistema cuando se alcanza el equilibrio. Intentemos, pues, resolver las ecuaciones del sistema en equilibrio mediante el uso de operadores. Consideremos el operador lineal D tal que: $D p_n = p_{n+1} \quad \forall n \geq 0$ (lo que implica que: $D^l p_n = p_{n+l} \quad \forall n \geq 0 \quad \forall l \geq 0$), entonces podemos escribir, para $n \geq 0$:

$$[\mu D^{k+1} - (\lambda+\mu) D + \lambda]\, p_n = 0 \qquad (1)''$$

Sean $(r_1, r_2 \dots r_{k+1})$ las raíces del operador ó de la ecuación característica, i.e.:

$$\mu D^{k+1} - (\lambda+\mu) D + \lambda = 0$$

(*) ¿cuales son esas condiciones? Para la existencia de distr. estacionaria es imprescindible que $\exists$ una solución global de la ec característica.

$(D-r_1)(D-r_2)\cdots(D-r_{K+1})=0$, entonces la solución general de la ecuación anterior es: (*)

$$p_n=\sum_{i=1}^{K+1} C_i r_i^n \qquad n\geq 0$$

y dado que $\sum_{n=0}^{\infty} p_n = 1$, tenemos que: $1=\sum_{n=0}^{\infty}\left(\sum_{i=1}^{K+1} C_i r_i^n\right) = C_1\sum_{n=0}^{\infty} r_1^n + C_2\sum_{n=0}^{\infty} r_2^n + \cdots + C_{K+1}\sum_{n=0}^{\infty} r_{K+1}^n$

luego cada r_i debe ser menor que 1, o en caso contrario el correspondiente C_i debe ser cero. Por otra parte, el Teorema de Rouché (**) permite demostrar que la ecuación $\mu D^{K+1}-(\lambda+\mu)D+\lambda=0$ sólo tiene una raíz (r_0) en el intervalo $(0,1)$, con lo que se obtiene:

$$p_n = C r_0^n \qquad n\geq 0 \qquad 0<r_0<1 \qquad \text{y} \qquad \sum_{n=0}^{\infty} p_n = 1 \longrightarrow C = 1-r_0 = p_0$$

así pues, en consecuencia:

$$p_n=(1-r_0)\,r_0^n \qquad n\geq 0 \qquad 0<r_0<1$$

Hemos encontrado que el sistema $M/M^{[K]}/1$ tiene asociada una distribución de probabilidad para el tamaño de la línea de espera en el equilibrio que es idéntica al de la cola $M/M/1$, y en consecuencia:

$$L=E(Q)=\frac{r_0}{1-r_0} \xrightarrow[L=\lambda E(T)]{} E(T)=\frac{r_0}{\lambda(1-r_0)} \qquad \text{y} \qquad E(W)=E(T)-\frac{1}{\mu}=\frac{r_0(\lambda+\mu)-\lambda}{(1-r_0)\lambda\mu}$$

Ejemplo:

Una compañía propietaria de un túnel de lavado de coches decide cambiar su sistema de operación; para ello instalará una máquina que permite lavar 2 coches al mismo tiempo si los hay en la cola, o que lava uno si no hay otro esperando. La compañía espera llegadas Poisson con una media de 20 vehículos por hora y un tiempo de servicio exponencial (función del tamaño del coche) con una media de cinco minutos ¿cuál será el tamaño medio de la cola que se forme?

Se trata de un modelo $M/M^{[K]}/1$ con $\lambda=20/h$, $\mu=1/5\ \text{min}=12/h$, $K=2$. Entonces la ecuación del grado será:

$$\mu r^3-(\lambda+\mu)r+\lambda = 12r^3-32r+20=0 = 3r^3-8r+5$$

una raíz es: $r_1=1$, y dividiendo por $r-1$ se obtiene $3r^2+3r-5=0$, entonces

$$r=\frac{-3\pm\sqrt{69}}{6} \longrightarrow r_0 \cong 0.884$$

$r_2=-1.88$ #

(*) Nótese que (1) es una ecuación en diferencias con coeficientes constantes, y análogamente a lo que ocurre en el caso de ec. diferenciales con coeficientes constantes (en donde las soluciones homogéneas son de la forma e^{nr_i}), las soluciones homogéneas son de la forma r_i^n, siendo r_i una raíz característica, con lo que "$p_n = C_i r_i^n$" es una solución de (1):

$$\mu p_{n+k+1} - (\lambda+\mu) p_{n+1} + \lambda p_n = \mu C_i r_i^{n+k+1} - (\lambda+\mu) C_i r_i^{n+1} + C_i \lambda r_i^n = C_i r_i^n \underbrace{(\mu r_i^{k+1} - (\lambda+\mu) r_i + \lambda)}_{=0} = 0$$

y dado que hay $k+1$ raíces características, la solución general es:

$$p_n = \sum_{i=1}^{k+1} C_i r_i^n \qquad n \geq 0$$

(**) T^ma de Rouché

Si $f(z)$ y $g(z)$ son funciones analíticas en C (contorno cerrado) y si $|g(z)| < |f(z)|$ sobre C, entonces $f(z)$ y $f(z)+g(z)$ tienen el mismo número de ceros dentro de C.

Puede tomarse: $C = [0,1]$, $g = r^{k+1}$ y $f = \frac{\lambda}{\mu} - \left(\frac{\lambda}{\mu}+1\right) r$

con lo que:

$$L = \frac{\rho_0}{1-\rho_0} = \frac{0.884}{0.116} \simeq 7.62 \text{ coches}$$

pues lo que en sistema es $L_q = \lambda E(W) = \lambda E(T) - \lambda/\mu = L - \frac{\lambda}{\mu} = 7.62 - \frac{20}{12} \simeq 5.95$ coches

Deberán, pues, esperar cola de 6 coches, en media.

Comentarios

(1) Un modelo de colas cuya resolución es similar al $M/G/1$ es el que corresponde a un sistema en el que las llegadas se producen en bloques de tamaño aleatorio siguiendo un proceso de Poisson, siendo el tiempo de servicio una colección de v.a. iid, con distribución general: $M^{[X]}/G/1$.

Este problema se resuelve, asimismo, encajando una cadena de Markov, y el criterio de discretización es el de observar el sistema inmediatamente después que un cliente acaba de ser servido (véase Gross y Harris p. 255)

(2) En cuanto al modelo $M/G/c/\infty$, i.e. cuando hay múltiples canales de servicio, $c>1$, y no existe restricción de capacidad de la línea de espera, el problema que presenta es que no posee una cadena de Markov encajada en el sentido usual. Ya que el nº de llegadas entre dos abandonos sucesivos de la cola depende no sólo del tamaño de la cola en la salida inmediatamente anterior, sino de más factores pues hay múltiples servidores.

Ahora bien, para $G=M$ este problema ya ha sido resuelto y para $G=D$ puede resolverse reordenando el proceso original (véase Gross y Harris p. 308).

Capítulo 8

Colas GI/M/1

TEMA 8: COLA GI/M/1

Introducción

Estudiaremos aquellos sistemas en los que los tiempos de servicio se asumen exponenciales y no hay ninguna hipótesis específica respecto al régimen de llegada a la cola excepto que los tiempos entre llegadas sucesivas son independientes y están idénticamente distribuidos, i.e. $GI/M/c$. Estudiaremos el proceso que genera el modelo anterior sólo en aquellos instantes en que un nuevo cliente llega a la cola, y lo haremos, en primer lugar, para $c=1$ y aunque luego se complica un poco extenderemos este resultado para c canales de servicio.

Resulta curioso que podamos obtener resultados para el modelo $GI/M/c$, sin demasiada dificultad, cuando se ha visto en el tema anterior que ello no era posible para el modelo $M/G/c$, siendo sin embargo este último modelo mucho más real que el primero, ya que es más común encontrar llegadas de Poisson que no tiempos de servicio exponencial.

1. Condición de ergodicidad

Consideremos un sistema con un solo puesto de servicio cuyos tiempos entre llegadas sucesivas son v.a. iid con una distribución general común: $H(u)$, siendo los tiempos de servicio exponenciales de parámetro μ. Para el análisis de esta cola encajaremos una cadena de Markov en el proceso asociado.

Sea $Z(t)$ el número de personas en la cola, incluyendo el que recibe servicio, en el instante t. Se observa el proceso en el momento justo antes en que un nuevo cliente llega a la cola, lo que conduce a un proceso estocástico discreto: $X_0=0$, $X_n = Z(t_n)$, $n=1,2, \ldots$, verificándose:

$$X_{n+1} = X_n + 1 - N \qquad X_n \geq 0 \qquad X_n + 1 \geq N$$

siendo N el número de personas atendidas en el período de tiempo que media entre las dos sucesivas llegadas, que sólo depende de la longitud de dicho intervalo (tiempo de servicio exponencial) y del tamaño de la cola X_n en la etapa anterior (no pueden ser atendidos más de los que había en cola). Así, que, resulta una cadena de Markov con transición en los instantes (justo antes) en que se produce una llegada. Tenemos que determinar cuál es la matriz de probabilidad de transición de un paso. Para ello, recordemos que:

$$a_n = P(N=n) = \int_0^\infty P(N=n \,/\, T_e = t)\, dH(t) = \int_0^\infty e^{-\mu t}\frac{(\mu t)^n}{n!}\, dH(t) \qquad n \geq 0$$

siendo T_e el tiempo transcurrido entre dos llegadas sucesivas, y dado que $N|_{T_e=t}$ sigue una distribución de Poisson de parámetro μt. Introducimos la siguiente notación:

$$a_k = P(N=k) \quad k\geq 0 \qquad \longrightarrow \qquad \sum_{k=0}^{\infty} a_k = \int_0^\infty e^{-\mu t}\left(\sum_{n=0}^{\infty}\frac{(\mu t)^n}{n!}\right) dH(t) = \int_0^\infty dH(t) = 1$$

tenemos, que las probabilidades de transición son:

$$p_{ij} = P(X_{n+1}=j \,/\, X_n = i) = \begin{cases} P(N=i+1-j) = a_{i+1-j} & i+1 \geq j \geq 1 \\ \\ 0 & j > i+1 \quad (\text{para } N\geq 0) \end{cases}$$

El caso $j=0$ debe ser tratado separadamente, dado que una parte del tiempo puede corresponder al sistema desocupado y no es suficiente decir que se han servido $i+1-j$ durante el tiempo T_e, pues pueden haberlo sido en un tiempo menor. Se dice, entonces, que p_{i0} es la probabilidad de que por lo menos hayan sido atendidos $i+1$ clientes:

$$p_{i0} = 1 - \sum_{j=1}^{i+1} p_{ij} = 1 - a_0 - a_1 - \cdots - a_i = A_i, \quad i \geq 0.$$

De lo que se sigue:

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ \\ \end{array}\begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \left\{\begin{array}{cccccc} A_0 & a_0 & 0 & 0 & 0 & 0 \\ A_1 & a_1 & a_0 & 0 & 0 & 0 \\ A_2 & a_2 & a_1 & a_0 & 0 & 0 \\ A_3 & a_3 & a_2 & a_1 & a_0 & 0 \\ A_4 & a_4 & a_3 & a_2 & a_1 & a_0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}\right\} \end{array}$$

Se trata pues, de una cadena infinita irreducible y aperiódica $(0<a_k<1 \ , k=0,1,\dots)$. Si existe distribución de equilibrio y denotamos por q_n la probabilidad de que al llegar un individuo encuentre n en el sistema, se cumplirá que:

$$qP = q \qquad \text{siendo} \qquad q = \{q_n\}_{n=0,1,2\dots}$$

· Las condiciones de estacionariedad conducen a:

$$\left.\begin{aligned} q_0 &= \sum_{j=0}^{\infty} q_j A_j = \sum_{j=0}^{\infty} q_j \left(1-\sum_{l=0}^{j} a_l\right) = \sum_{j=0}^{\infty} q_j \left(\sum_{l=j+1}^{\infty} a_l\right) \\ q_i &= \sum_{j=0}^{\infty} q_{i+j-1}\, a_j = \sum_{j=i-1}^{\infty} q_j\, a_{j-i+1} \qquad i \geq 1 \end{aligned}\right\} \quad (1)$$

Nuestro propósito consiste en determinar cuales son las condiciones de ergodicidad de este modelo, y en esas condiciones calcular la distribución estacionaria asociada al sistema. Sabemos que existen diferentes procedimientos para determinar la probabilidad $\{q_n\}$, y es inmediato que no podemos aplicar ningún método iterativo para resolver las anteriores ecuaciones en diferencias, sólo nos quedaría, pues, utilizar funciones generatrices u otros operadores.

Si consideramos un operador D sobre la sucesión $\{q_n\}_{n=0,1,2\dots}$: $Dq_i = q_{i+1}$ para $i \geq 1$, tenemos que (de las ecuaciones en (1)):

$$q_i - [q_{i-1}a_0 + q_i a_1 + q_{i+1} a_2 + \text{——}] = 0$$

luego $$q_{i-1}[D - a_0 - Da_1 - D^2 a_2 - D^3 a_3 \text{——}] = 0$$

con lo que una solución no trivial para q_{i-1}, $i \geq 1$, conduce a:

$$D - \sum_{n=0}^{\infty} D^n a_n = 0 \quad , \text{ i.e. } \quad D = \sum_{n=0}^{\infty} D^n a_n = f(D) \qquad 0 \leq D \leq 1$$

siendo $f(s)$ la función generatriz de probabilidad de $\{a_n\}$. En este caso resulta que $f(\cdot)$ es una función estrictamente convexa $(f''(s) \geq 0)$ y monótona creciente $(f'(s) \geq 0)$ en $[0,1]^{(*)}$, siendo $f(0) = a_0$, $0<a_0<1$ y $f(1) = \sum_{n=0}^{\infty} a_n = 1$. Así, pues, sólo pueden existir cero o una raíz de la ecuación $f(s)=s$ en $(0,1)$. Por otra parte, sabemos que existirá una solución $0<s_0<1$ si, y sólo si, $f'(1)>1$, i.e. si $f'(1) = \sum_{k=0}^{\infty} k a_k$ = número medio de servicios entre 2 llegadas consecutivas es mayor que uno. Además:

$$\sum_{k=0}^{\infty} k a_k = \int_0^{\infty} e^{-\mu t} \left\{ \sum_{k=1}^{\infty} \frac{(\mu t)^k}{k!} k \right\} dH(t) = \int_0^{\infty} e^{-\mu t} \mu t \, dH(t) \left(\sum_{k=1}^{\infty} \frac{(\mu t)^{k-1}}{(k-1)!} \right) = \mu \int_0^{\infty} t \, dH(t) =$$

(*) De nuevo solo nos interesan las raíces de la ecuación $f(D) = D$ en $[0, 1]$, puesto que queremos que se verifique: $\sum q_n = 1 = q_0 + q_1 \sum_{j=0}^{\infty} D^j$, y $D < 1$ es una condición para la convergencia de la serie $\sum_{j=0}^{\infty} D^j$.

$$\sum_{k=1}^{\infty} k a_k = \frac{\int_0^{\infty} t \, dH(t)}{\frac{1}{\mu}} = \frac{\text{tiempo medio entre llegadas}}{\text{tiempo medio de servicio}} = \mu E(T_e) = 1/\rho$$

Con lo que la condición de ergodicidad es la de $\rho < 1$. Supongamos que $\exists\, s : 0 < s < 1$ tal que $f(s) = s$, entonces $q_i = q_1 s^{i-1}$ $i \geq 1$, solución que satisface la relación de (1), veamos qué ocurre con q_0:

$$q_0 = \sum_{j=0}^{\infty} q_j A_j = A_0 q_0 + q_1 \sum_{j=1}^{\infty} s^{j-1} \left(\sum_{k=j+1}^{\infty} a_k \right) = (1-a_0) q_0 + q_1 (a_2 + a_3 + \cdots) + s q_1 (a_3 + a_4 + \cdots) +$$

$$+ s^2 q_1 (a_4 + a_5 + \cdots) + \cdots = q_0 - a_0 q_0 + q_1 \sum_{k=2}^{\infty} a_k \left(\sum_{j=0}^{k-2} s^j \right)$$

con lo que
$$a_0 q_0 = q_1 \sum_{k=2}^{\infty} a_k \left(\frac{1-s^{k-1}}{1-s} \right) = \frac{q_1}{1-s} (1 - a_0 - a_1) - \frac{q_1}{(1-s)s} (s - a_0 - s a_1) = \frac{a_0 q_1 (1-s)}{s(1-s)}$$

luego $a_0 q_0 = a_0 q_1 / s \longrightarrow s q_0 = q_1 \longrightarrow q_j = q_0 s^j \quad j \geq 1$

Pero $\sum_{j=0}^{\infty} q_j = 1 = q_0 \sum_{j=0}^{\infty} s^j = \frac{q_0}{1-s} \longrightarrow q_0 = (1-s)$ y $q_j = (1-s) s^j \quad j \geq 1$

Así, pues, hemos determinado la distribución estacionaria única del sistema, cuando $f'(1) = 1/\rho > 1$, i.e. con $\rho < 1$, y se ha obtenido que es una distribución geométrica de parámetro $(1-s)$, el problema que quedaría pendiente es el de determinar el valor de s, solución de $f(s) = s$, lo que puede hacerse por métodos numéricos (p.e.: método de Newton y Raphson, que converge por convexidad de $f(\cdot)$).

Nota. Puede, asimismo, comprobarse que si $f'(1) < 1$ entonces el sistema es transitorio. Resultando que la condición $f'(1) = 1$ es la que asegura que nos encontramos ante un sistema recurrente nulo. (véase K-T I p. 102)

Se ha obtenido la misma distribución estacionaria que la correspondiente a la cola $M/M/1$, reemplazando ρ por s, con lo que los resultados obtenidos para aquella cola (tiempo medio de espera, longitud de la cola, etc) serían válidos para el modelo $GI/M/1$. Sin embargo, debe señalarse que q_n es la probabilidad (cuando se alcanza el equilibrio) de que haya n individuos en el sistema justo antes de una llegada y no la probabilidad p_n de que haya n individuos en el sistema en equilibrio, i.e. $q_n \neq p_n$. Además, resulta que $p_n = q_n$ si y sólo si $GI = M$, es decir si las llegadas son Poisson.

2. Distribución del tiempo de espera

Supongamos que la cola ha alcanzado su régimen de equilibrio y sea W el tiempo de espera de un cliente en la cola antes de ser atendido. La probabilidad de que no tenga que esperar es la de que a su llegada no haya nadie en el sistema p_0, siendo $p_0 = 1-s$, donde $0<s<1$ tal que $f(s)=s$ y $\rho<1$. Si, por el contrario, encuentra n clientes en el sistema, $n\geq 1$, tendrá que esperar a que les sirvan a todos ellos antes de empezar a ser atendido:

$$n\geq 1 \qquad P(W\leq t \,/\, n \text{ delante}) = P(T_1+T_2+\dots+T_n\leq t) \qquad t\geq 0$$

ahora bien, $\sum_{i=1}^{n} T_i \stackrel{D}{=} \Gamma(n,\mu)$, puesto que es la suma de n v.a. exponenciales indep. de parámetro μ, que corresponden a los tiempos de servicio de los n clientes que hay en el sistema. En consecuencia:

$$t\geq 0 \qquad P(W\leq t) = \underbrace{P(W\leq t/n=0)}_{=1}\cdot p_0 + \sum_{n=1}^{\infty} P(W\leq t/n \text{ delante})\cdot p_n =$$

$$= (1-s) + (1-s)\sum_{n=1}^{\infty} s^n \int_0^t \frac{e^{-\mu\tau}\mu^n \tau^{n-1}}{(n-1)!}\,d\tau \underset{\text{p.p.}}{=} (1-s) + s\{1-e^{-\mu(1-s)t}\}$$

$$= 1-s\,e^{-\mu(1-s)t}$$

Nótese que la función de distribución del tiempo de espera es una mixtura de una distribución exponencial de parámetro $\mu(1-s)$ y de una distribución degenerada cuyo único valor posible es 0 con probabilidad $(1-s)=p_0$ (probabilidad de que el cliente no tenga que esperar para ser atendido).

Podemos, pues, calcular el tiempo medio de espera en la cola:

$$E(W) = 0\cdot p_0 + \int_0^{\infty} t\, f_W(t)\,dt = \int_0^{\infty} s t\,\mu(1-s)\,e^{-\mu(1-s)t}\,dt \underset{\text{p.p.}}{=} \frac{s}{\mu(1-s)}$$

Es curioso constatar que se han obtenido para GI/M/1 los mismos resultados, que en el caso del modelo M/M/1, también para la distribución del tiempo de espera y para la distribución del tiempo de permanencia en la cola (como se puede ver fácilmente). Ello es debido a que en ambos modelos los tiempos de servicio son exponenciales independientes de parámetro μ, y al hecho de que la probabilidad de equilibrio que usamos están evaluadas en el instante en que un cliente llega a la cola; así, aunque en el caso M/M/1 teníamos π_n (probabilidad de n individuos en el sistema) y en el caso GI/M/1 hemos obtenido q_n, como esta última es la probabilidad de encontrar n en el sistema cuando se produce una nueva llegada, los resultados son los mismos. En consecuencia, para T= tiempo de permanencia en la cola tendríamos que:

$$P(T\le t) = 1 - e^{-\mu t(1-s)} \quad t\ge 0 \longrightarrow T \stackrel{D}{=} \exp\{\mu(1-s)\} \quad \text{y} \quad E(T)=\frac{1}{\mu(1-s)}$$

Por otra parte, la longitud media del sistema en el equilibrio, medida en el instante en que llega un cliente será:

$$L=\sum_{n=0}^{\infty} n q_n = \frac{s}{1-s} \qquad \text{puesto que} \quad q_n=(1-s)s^n \quad n\ge 0$$

y en este caso no se cumple la fórmula de Little, puesto que si

$$L=[E(T_e)]^{-1}E(T)=\frac{s}{1-s}, \quad \text{siendo} \quad E(T)=\frac{1}{\mu(1-s)} \longrightarrow [E(T_e)]^{-1}=\mu s \quad (*)$$

cuando $E(T_e)$ es el tiempo medio entre dos llegadas sucesivas y no tiene porqué depender de μ y de s.

Lo que sí es cierto, es que el nº medio de personas en la cola, cuando llega un nuevo cliente es:

$$L_q = 0\, q_0 + \sum_{n=1}^{\infty}(n-1)q_n = \sum_{n=1}^{\infty} n q_n - \sum_{n=1}^{\infty} q_n = L-(1-q_0) = \frac{s}{1-s}-s = \frac{s^2}{1-s}$$

y tampoco ahora se cumple la fórmula de Little.

3. Casos particulares

(i) Cola M/M/1

En este caso los tiempos entre llegadas sucesivas son exponenciales de parámetro λ, con lo que la función de distribución $H(t)=1-e^{-\lambda t}$ $t\ge 0$, con lo que

$$a_k = \int_0^\infty e^{-\mu t}\frac{(\mu t)^k}{k!}\lambda e^{-\lambda t}\,dt = \int_0^\infty \lambda e^{-(\mu+\lambda)t}\frac{(\mu t)^k}{k!}\,dt \qquad k=0,1,2\ldots$$

luego

$$f(s)=\sum_{k=0}^{\infty} a_k s^k = \int_0^\infty \lambda e^{-(\mu+\lambda)t}\left\{\sum_{k=0}^{\infty}\frac{(\mu t s)^k}{k!}\right\}dt = \int_0^\infty \lambda e^{-t[\mu(1-s)+\lambda]}\,dt = \frac{\lambda}{\mu(1-s)+\lambda}$$

(*) de hecho, sabemos que $f'(1) = \sum_{k=1}^{\infty} k\, a_k = m = \mu\, E(T_e) \longrightarrow E(T_e) = \frac{m}{\mu}$

dado que en este caso $\rho = \frac{\lambda}{\mu} \longrightarrow f(s) = \dfrac{1}{1 + \frac{(1-s)}{\rho}}$

Necesitamos resolver la ecuación $f(s)=s$, pero tenemos que esto equivale a:

$$\frac{1}{s} = 1 + \frac{1-s}{\rho} \longleftrightarrow \frac{1-s}{\rho} = \frac{1-s}{s} \longleftrightarrow \rho = s = \frac{\lambda}{\mu} < 1$$

cosa que ya habíamos comentado anteriormente.

(ii) Cola D/M/1

En este caso los intervalos de tiempo entre llegadas son constantes de longitud γ, y, en consecuencia:

$$a_k = \frac{(\mu\gamma)^k}{k!} e^{-\mu\gamma} \qquad k=0,1,2 \ldots$$

i.e. N = nº de clientes atendidos entre 2 llegadas sucesivas ~ Poisson de parámetro $\gamma\mu$

Así, pues, tenemos que

$$f(s) = \sum_{k=0}^{\infty} e^{-\mu\gamma} \frac{(\mu\gamma s)^k}{k!} = e^{-\mu\gamma} e^{\mu\gamma s} = e^{-\mu\gamma(1-s)}$$

y la condición de ergodicidad será: $f'(s) = \mu\gamma e^{-\mu\gamma(1-s)} \longrightarrow f'(1) = \mu\gamma > 1$, pero $\mu\gamma = \frac{\gamma}{\frac{1}{\mu}} = \frac{1}{\rho} > 1$.

Si ~~queremos~~ resolver la ecuación $f(s)=s$, hemos de encontrar solución para la identidad siguiente: $s = e^{-\mu\gamma(1-s)}$ (véase problema nº 1)

(iii) Cola $E_k/M/1$

Supongamos que el tiempo entre llegadas sucesivas sigue una distribución de Erlang de tipo k con media $1/\lambda$, i.e. se trata de una $\Gamma(k, k\lambda)$ cuya función de densidad es:

$$h(t) = \frac{(\lambda k)^k}{(k-1)!} t^{k-1} e^{-\lambda k t} \qquad t>0$$

Esto puede interpretarse como que cada llegada se produce después de pasar necesariamente por k fases, cada una de ellas exponencial con un tiempo medio $1/k\lambda$. Sabemos que, en este caso:

$$a_n = \int_0^\infty e^{-\mu t} \frac{(\mu t)^n}{n!} \frac{(\lambda k)^k t^{k-1} e^{-\lambda k t}}{(k-1)!} dt = \frac{(\lambda k)^k}{(k-1)!} \int_0^\infty e^{-(\mu+\lambda k)t} \frac{(\mu t)^n}{n!} t^{k-1} dt$$

con lo que:

$$f(s) = \sum_{n=0}^{\infty} a_n s^n = \frac{(\lambda k)^k}{(k-1)!} \int_0^\infty e^{-(\mu+\lambda k)t} e^{\mu t s} t^{k-1} dt = \frac{(\lambda k)^k}{\Gamma(k)} \int_0^\infty e^{-t[\mu(1-s)+\lambda k]} t^{k-1} dt =$$

$$f(s) = \frac{(\lambda k)^k}{\Gamma(k)} \; \frac{\Gamma(k)}{[\mu(1-s)+\lambda k]^k} = \frac{1}{\left[\frac{\mu}{\lambda k}(1-s)+1\right]^k} = \frac{1}{\left[\frac{1-s}{k\rho}+1\right]^k} \quad , \quad \text{para } \rho = \frac{1/\mu}{1/\lambda} = \frac{\lambda}{\mu}$$

Ahora bien: $f'(s) = -k\left[\frac{1-s}{k\rho}+1\right]^{-k-1}\left(-\frac{1}{k\rho}\right) = \frac{1}{\rho}\,\frac{1}{\left(\frac{1-s}{k\rho}+1\right)^{k+1}} \longrightarrow f'(1) = \frac{1}{\rho} > 1 \iff \rho < 1$

Para poder tener especificadas las medidas de efectividad de la cola y la distribución de probabilidad asociada a la misma necesitamos resolver, de nuevo: $f(s)=s$, o lo que es lo mismo:

$\frac{1-s}{k\rho}+1 = s^{-1/k}$, si hacemos $r = s^{1/k} \longrightarrow \frac{1-r^k}{k\rho}+1 = \frac{1}{r} \longrightarrow (1+k\rho-r^k)\,r = k\rho$

En consecuencia: $r^{k+1} - r(1+k\rho) + k\rho = 0$, o sustituyendo $\rho = \lambda/\mu$:

$$\mu r^{k+1} - r(\mu+k\lambda) + k\lambda = 0$$

Nota. Puede comprobarse que se trata de la misma ecuación característica que obtuvimos en el Tema 7 al estudiar el servicio en bloque (cola $M/M^{[k]}/1$), para la cual la distribución de equilibrio era también geométrica de parámetro r, con $0<r<1$. Ahora bien, nuestro $s = r^k$.

Ejemplo

Supongamos que las llegadas a una cola con un único servidor son Erlang de tipo 2, con un tiempo medio entre llegadas de 30 minutos. Siendo el tiempo de servicio exponencial de media 25 min. Encontrar la prob. de equilibrio del tamaño del sistema, cuando un cliente se incorpora a la cola y los valores medios de la efectividad del sistema.

Se trata de una cola $E_2/M/1$ con $\lambda = 1/30$ /min $= 2$ /h y $\mu = 1/25$min $= 12/5$ /hr, $k=2$, hemos de resolver, pues, la ecuación $\frac{12}{5}r^3 - \left(\frac{12}{5}+4\right)r + 4 = 0$, o lo que es lo mismo $3r^3 - 8r + 5 = 0$, dado que $r=1$ es una raíz de la ecuación (pues $f(1)=1$), hemos de resolver: $3r^2+3r-5=0 \longrightarrow r = 0.884 \longrightarrow s = r^2 = 0.781$

Así pues: $p_0 = 1-s = 0.219$ y $p_n = 0.219\,(0.781)^n \quad n \geq 1$

$$L = \frac{s}{1-s} = \frac{0.781}{0.219} \simeq 3.56$$

$E(W) = \frac{s}{\mu(1-s)} = 1.48$ hr y $E(T) = \frac{1}{\mu(1-s)} = 1.899$ hr. } no se cumple F. Little

Nota. Para este modelo $E_k/M/1$ sí pueden calcularse con facilidad las probabilidades, en el equilibrio, del tamaño del sistema, i.e. π_n, siguiendo un razonamiento del tipo que introduciremos al estudiar la llegada en bloque. (véase G-H p. 174)

4. Llegada en bloque: cola $M^{[X]}/M/1$

El modelo de cola que corresponde a la llegada en bloque es un modelo markoviano que no es de nacimiento muerte (i.e. permitimos más de un cambio en un intervalo de tiempo infinitesimal). Para abordar este problema utilizaremos la técnica que consiste en obtener las ecuaciones diferenciales de Chapman-Kolmogorov, y a partir de ellas se deducirán las correspondientes ecuaciones en diferencias finitas imponiendo la condición de que la derivada de la función de probabilidad $p_n(t)$ sea nula (sistema en equilibrio).

En este modelo las llegadas siguen un proceso de Poisson, con la particularidad de que el nº de clientes que llega en cada ocasión es una v.a X que toma valores enteros k, con probabilidad c_k. Existe un único servidor cuyo tiempo de servicio es exponencial de media $1/\mu$, y el modelo se denota con la inicial $M^{[X]}/M/1$.

Sea λ_k el promedio de llegadas, según un proceso de Poisson, de grupos de tamaño k y sea $\lambda = \sum_{i=1}^{\infty} \lambda_i$, el promedio conjunto de llegadas de cualquier tamaño, entonces λ_k/λ constituye una estimación (frecuencialista) de $c_k = P(X=k)$ y se escribe $c_k = \lambda_k/\lambda$.

Sea $X(t)$ el número de individuos que se han incorporado al sistema en $(0,t]$, veamos que $X(t)$ es un proceso compuesto de Poisson. Si X_n es el número de individuos que llegan cuando ocurre el n-ésimo suceso de Poisson (= llegada), sabemos que X_n, $n\geq 1$ es una colección de v.a. iid con distribución común la de X, verificándose que:

$$X(t) = \sum_{n=1}^{N(t)} X_n$$

donde $N(t)$ es el proceso de Poisson de llegadas, cuyo parámetro es λ. Así pues, es un modelo markoviano, el $M^{[X]}/M/1$, en el que las llegadas siguen un proceso compuesto de Poisson, para este proceso ya se estudiaron en temas anteriores la distribución de probabilidad y parámetros correspondientes.

Calculemos la ecuación de Chapman-Kolmogorov para $Z(t)$:

$$\left.\begin{array}{ll} n\geq 1 & p_n(t+\Delta t) = (1-\lambda\Delta t)(1-\mu\Delta t)\,p_n(t) + (1-\lambda\Delta t)\,\mu\Delta t\,p_{n+1}(t) + (1-\mu\Delta t)\underbrace{\sum_{k=1}^{n}\lambda_k\,\Delta t\,p_{n-k}(t)}_{(1)} + o(\Delta t) \\ & p_0(t+\Delta t) = (1-\lambda\Delta t)\,p_0(t) + (1-\lambda\Delta t)\,\mu\Delta t\,p_1(t) + o(\Delta t) \end{array}\right\}$$

(1) → la probabilidad de que habiendo $n-k$ en el sistema en el instante t llegue en el intervalo $(t, t+\Delta t)$ un grupo de tamaño k.

Se sigue de forma inmediata:

$$\left.\begin{array}{ll} n\geq 1 & p'_n(t) = -(\lambda+\mu)\,p_n(t) + \mu\,p_{n+1}(t) + \lambda\sum_{k=1}^{n} c_k\,p_{n-k}(t) \\ & p'_0(t) = -\lambda\,p_0(t) + \mu\,p_1(t) \end{array}\right\}$$

Si asumimos que el sistema se encuentra en equilibrio, las ecuaciones estacionarias serán:

$$\left.\begin{array}{ll} n\geq 1 & 0 = -(\lambda+\mu)\,\pi_n + \mu\,\pi_{n+1} + \lambda\sum_{k=1}^{n} c_k\,\pi_{n-k} \\ & 0 = -\lambda\pi_0 + \mu\pi_1 \end{array}\right\} (2)$$

Para resolver este sistema usamos la función generatriz de probabilidad de $\{\pi_n\}$ y $\{c_n\}$

$$\Pi(s) = \sum_{n=0}^{\infty}\pi_n s^n \quad \text{y} \quad C(s) = \sum_{n=1}^{\infty} c_n s^n \qquad 0\leq s<1$$

Multiplicando las ecuaciones del sistema (2) por el s^n, $n\geq 0$, adecuado y sumando se llega a:

$$0 = -\lambda\sum_{n=0}^{\infty}\pi_n s^n - \mu\sum_{n=1}^{\infty}\pi_n s^n + \frac{\mu}{s}\sum_{n=1}^{\infty}\pi_n s^n + \lambda\underbrace{\sum_{n=1}^{\infty}\left(\sum_{k=1}^{n} c_k\,\pi_{n-k}\right)s^n}_{\sum_{k=1}^{\infty} c_k s^k\left(\sum_{n=k}^{\infty}\pi_{n-k}\,s^{n-k}\right)}$$

luego

$$0 = -\lambda\,\Pi(s) - \mu\left(\Pi(s)-\pi_0\right) + \mu/s\left(\Pi(s)-\pi_0\right) + \lambda\,\Pi(s)\,C(s)$$

con lo que:

$$\Pi(s) = \frac{\mu\pi_0\,(1-s)}{\mu(1-s) - \lambda s\,[1-C(s)]} \qquad (3)$$

Y podemos usar la condición: $\Pi(1) = C(1) = 1$, para determinar π_0:

$$1 = \lim_{s\to 1}\Pi(s) \overset{\text{L'Hôpital}}{=} \lim_{s\to 1}\left\{\frac{-\mu\pi_0}{-\mu-\lambda+\lambda C(s)+\lambda s\,C'(s)}\right\} = \frac{-\mu\pi_0}{-\mu+\lambda E(X)}$$

siendo $E(X)$ el tamaño medio de los grupos, y de donde se sigue que:

$$\pi_0 = 1 - \frac{\lambda E(X)}{\mu}$$

si llamamos $\rho = \dfrac{\text{tiempo medio de servicio}}{\text{tiempo medio entre llegadas}} = \dfrac{1/\mu}{\dfrac{1}{\lambda E(X)}} = \dfrac{\lambda E(X)}{\mu}$

resulta que $\pi_0 = 1-\rho$, y que $\rho<1$ es una condición suficiente de ergodicidad.

Entonces, dado $C(s)$, la distribución de equilibrio $\{\pi_n\}$, con $\rho<1$, se obtendría utilizando cualquier método que permita calcular los coeficientes de la serie de expansión de $\Pi(s)$, dada en (3).

Ejemplo

Veamos una ilustración de los resultados anteriores cuando sabemos que X se distribuye geométricamente, es decir: $c_k = (1-\alpha)\alpha^{k-1}$ $k=1,2,\ldots$ $0<\alpha<1$.

Entonces: $C(s) = \sum_{n=1}^{\infty} c_n s^n = (1-\alpha)\sum_{n=1}^{\infty} \alpha^{n-1} s^n = (1-\alpha)s\sum_{n=1}^{\infty}(\alpha s)^{n-1} = \dfrac{(1-\alpha)s}{1-\alpha s}$, para $\alpha s<1$

Si sustituimos en (3), resultaría:

$$\Pi(s) = \frac{\mu\pi_0(1-s)}{\mu(1-s) - \lambda s\left[1 - \frac{s(1-\alpha)}{1-\alpha s}\right]} = \frac{\mu\pi_0(1-s)}{\mu(1-s) - \lambda s\frac{(1-s)}{1-\alpha s}} = \frac{\mu\pi_0}{\mu - \frac{\lambda s}{1-\alpha s}} = \frac{\mu(1-\rho)(1-\alpha s)}{\mu(1-\alpha s)-\lambda s} = \frac{(1-\rho)(1-\alpha s)}{(1-\alpha s) - \lambda s/\mu}$$

Ahora bien:

$$\rho = \frac{\lambda E(X)}{\mu} = \frac{\lambda}{\mu}(1-\alpha)\sum_{n=1}^{\infty} n\alpha^{n-1} = \frac{\lambda(1-\alpha)}{\mu}\left\{\frac{d}{d\alpha}\sum_{n=1}^{\infty}\alpha^n\right\} = \frac{\lambda(1-\alpha)}{\mu}\left\{\frac{d}{d\alpha}\left(\frac{\alpha}{1-\alpha}\right)\right\} = \frac{\lambda(1-\alpha)}{\mu}\frac{1}{(1-\alpha)^2}$$

luego, $\rho = \dfrac{\lambda}{\mu}\dfrac{1}{1-\alpha} = \dfrac{\lambda}{\mu(1-\alpha)}$ y al sustituir:

$$\Pi(s) = \frac{(1-\rho)(1-\alpha s)}{1-\alpha s - \rho s(1-\alpha)} = \frac{(1-\rho)(1-\alpha s)}{1-s[\alpha+\rho-\alpha\rho]} = (1-\rho)\left\{\frac{1}{1-s[\alpha+\rho-\alpha\rho]} - \frac{\alpha s}{1-s[\alpha+\rho-\alpha\rho]}\right\} =$$

$$= (1-\rho)\left\{\sum_{n=0}^{\infty}([\alpha+\rho-\alpha\rho]s)^n - \sum_{n=0}^{\infty}\alpha[\alpha+\rho-\alpha\rho]^n s^{n+1}\right\}$$

con lo que:

$$\pi_n = (1-\rho)\left\{[\alpha+\rho(1-\alpha)]^n - \alpha[\alpha+\rho-\alpha\rho]^{n-1}\right\} = (1-\rho)(\alpha+\rho-\alpha\rho)^{n-1}(1-\alpha)\rho \qquad n=1,2,3\ldots$$

$\pi_0 = 1-\rho$

Cola GI/M/c

Al generalizar el modelo de llegadas general, GI/M/1, para c canales de servicio, nos encontramos con que todos los resultados se mantienen, excepto los cálculos para determinar a_n. Ello es debido a que la tasa media de servicio es μ, $i\mu$ ó $c\mu$, dependiendo del estado en que nos encontremos, es decir, que a_n depende de i y de j. A causa de esta dependencia obtendremos una matriz $P=[p_{ij}]$ algo diferente.

Recuérdese que, para $c=1$, teníamos:

$$p_{ij}=\begin{cases} 0 & \text{si } j>i+1 \\ a_{i+1-j} & \text{si } i+1\geq j\geq 1 \\ 1-\sum_{k=0}^{i} a_k & j=0 \end{cases}$$

Si consideramos que el sistema tiene c servidores, $c>1$, sigue siendo cierto que:

$p_{ij}=0$ si $j>i+1$. (pues $i\geq 0$)

Cuando $j\geq c$, y dado que el tiempo de servicio medio es $1/c\mu$, se tiene que la variable aleatoria $N|_{T_e=t}$ será $P_o(c\mu t)$, con lo que:

$$a_k=\int_0^\infty e^{-c\mu t}\frac{(c\mu t)^k}{k!}\,dH(t) \qquad \text{y} \qquad p_{ij}=a_{i+1-j} \quad \text{para} \quad i+1\geq j\geq c$$

La situación que presenta dificultades es la que corresponde a $j<c$, y habrá que diferenciar dos casos: $i\geq c$ (en donde la tasa media de servicio varía desde $c\mu$ hasta $j\mu$) e $i<c$ (en donde la tasa de servicio media varía desde $i\mu$ hasta $j\mu$)

Consideremos, primero, el caso $j\leq i+1\leq c$. Todos están siendo atendidos y la probabilidad de que alguno complete su servicio en $T_e=t$ es $1-e^{-\mu t}$, independientemente para cada servicio. Para ir de i a j deben haberse completado, hasta el instante t, $i+1-j$ servicios y usando el modelo binomial:

$$p_{ij}=\int_0^\infty \binom{i+1}{i+1-j}(1-e^{-\mu t})^{i+1-j}e^{-\mu tj}\,dH(t) \qquad j\leq i+1\leq c$$

Nos queda el caso en que $i+1>c>j$. El sistema empieza con todos los canales ocupados, pues $i\geq c$, y algunos canales se van desocupando en el intervalo T_e hasta que sólo quedan j servidores ocupados. Definimos V como el tiempo transcurrido desde la llegada del último cliente hasta que empiezan a servirlo ($0<V<T_e$), en cuyo instante los c canales están ocupados; sea $B(v)$ la función de distribución de V.

Para ir de i a j en T_e, tienen que haberse completado $c-j$ servicios en el intervalo $[v, T_e]$, i.e. en un intervalo de amplitud $T_e - V$. Usando de nuevo la binomial y aplicando de nuevo que el tiempo de servicio (exponencial) tiene la propiedad de la pérdida de memoria

$$p_{ij} = \int_0^\infty \int_0^t \binom{c}{c-j} [1-e^{-\mu(t-v)}]^{c-j} e^{-\mu(t-v)j} \, dB(v) \, dH(t)$$

Ahora bien, la v.a. V es simplemente el tiempo transcurrido hasta que $i+1-c$ clientes han sido atendidos, cuando hay c canales de servicio. Luego $V \stackrel{d}{=} \Gamma(i+1-c, \mu c)$ con lo que

$$b(v) = \frac{d}{dv} B(v) = \frac{(c\mu)^{1+i-c} \, v^{i-c}}{(i-c)!} e^{-c\mu v}$$

y sustituyendo quedaría:

$$p_{ij} = \binom{c}{c-j} \frac{(c\mu)^{i+1-c}}{(i-c)!} \int_0^\infty \left(\int_0^t [1-e^{-\mu(t-v)}]^{c-j} e^{-\mu(t-v)j} e^{-c\mu v} v^{i-c} \, dv \right) dH(t), \quad i+1>c>j$$

Nótese que las $c-1$ primeras columnas de la matriz $P=[p_{ij}]$ son las únicas que presentan variación respecto del caso $GI/M/1$. Para $j \geq c$ tienen la misma probabilidad a_{i+1-j}, y sólo ha variado el parámetro: $\mu \to c\mu$.

Las ecuaciones de estacionariedad, para los instantes en que acaba de llegar un cliente, serán: $q_j = \sum_{i=0}^{\infty} p_{ij} q_i$, $j \geq 0$.

Ahora bien, para $j \geq c$, tendremos que:

$$q_j = \sum_{i=0}^{j-2} 0 \, q_i + \sum_{i=j-1}^{\infty} a_{i+1-j} q_i = \sum_{k=0}^{\infty} a_k q_{j+k-1} \qquad j \geq c$$

y por el razonamiento seguido para $c=1$, tendríamos que $q_j = C r_0^j$, siendo r_0 la raíz (única) de la ecuación característica $f(s)=s$, $0<r_0<1$, con $f(s) = \sum_{n=0}^{\infty} a_n s^n$. (*)
La constante C y q_j, $j=0,1,\dots,c-1$, deben ser determinadas a partir de la condición $\sum_{j=0}^{\infty} q_j = 1$ y de las $c-1$ primeras ecuaciones de estacionariedad, usando las p_{ij} que han sido obtenidas. Pero esto no es sencillo, dado que estas $c+1$ ecuaciones con sus $c+1$ incógnitas involucran sumatorios infinitos. Lo que haremos es obtener C en función de $q_0, q_1, \dots, q_c$ y r_0, y desarrollaremos una fórmula de tipo recursivo que permitirá determinar, en cada caso, las q_j, $j=0,1,\dots,c$.

(*) condición de ergodicidad : $\frac{\lambda}{c\mu} < 1$

Tenemos que

$$1 = \sum_{j=0}^{\infty} q_j = \sum_{j=0}^{c-1} q_j + \sum_{j=c}^{\infty} C r_0^j \longrightarrow C = \frac{1-\sum_{j=0}^{c-1} q_j}{\sum_{j=c}^{\infty} r_0^j} = \frac{1-\sum_{j=0}^{c-1} q_j}{\frac{r_0^c}{1-r_0}} = \frac{1-r_0}{r_0^c}\left[1-\sum_{j=0}^{c-1} q_j\right]$$

Por otra parte, tenemos que:

$$q_j = \sum_{i=0}^{\infty} p_{ij} q_i = \sum_{i=0}^{c-1} p_{ij} q_i + \sum_{i=c}^{\infty} a_{i+1-j} C r_0^i$$

y según se vio al calcular las p_{ij}:

$$q_j = \sum_{i=j-1}^{c-1} p_{ij} q_i + C \sum_{i=c}^{\infty} a_{i+1-j} r_0^i \qquad 1 \le j \le c-1$$

luego

$$q_{j-1} = \frac{q_j - \sum_{i=j}^{c-1} p_{ij} q_i - C \sum_{i=c}^{\infty} a_{i+1-j} r_0^i}{p_{j-1,j}} \qquad 1 \le j \le c-1$$

Si hacemos $q'_j = \frac{1}{C} q_j$, $\forall j$, resulta que:

$$q'_{j-1} = \frac{q'_j - \sum_{i=j}^{c-1} p_{ij} q'_i - \sum_{i=c}^{\infty} a_{i+1-j} r_0^i}{p_{j-1,j}} \qquad 1 \le j \le c-1$$

Ahora bien: $q_c = \sum_{i=0}^{\infty} p_{ic} q_i = \sum_{i=c-1}^{c} p_{ic} q_i + C \sum_{i=c+1}^{\infty} a_{i+1-c} r_0^i$, luego

$$q_{c-1} = \frac{(1-p_{cc}) q_c - C \sum_{i=c+1}^{\infty} a_{i+1-c} r_0^i}{p_{c-1,c}} = \frac{(1-a_1) q_c - C \sum_{i=c+1}^{\infty} a_{i+1-c} r_0^i}{a_0}$$

pero $q'_c = q_c/C = r_0^c$, y si dividimos la identidad anterior por C, resulta:

$$q'_{c-1} = \frac{(1-a_1) r_0^c - \sum_{i=c+1}^{\infty} a_{i+1-c} r_0^i}{a_0} = \frac{r_0^c - \sum_{i=c}^{\infty} a_{i+1-c} r_0^i}{a_0}$$

y recursivamente obtendríamos $q'_{c-2}, q'_{c-3} — q'_1, q'_0$. Además:

$$C = \frac{1-r_0}{r_0^c}\left[1 - C\sum_{j=0}^{c-1} q'_j\right] \longrightarrow C\left[1+\sum_{j=0}^{c-1} q'_j\right] = \frac{1-r_0}{r_0^c} \Longrightarrow C = \frac{1-r_0}{r_0^c}\left[1+\sum_{j=0}^{c-1} q'_j\right]^{-1}$$

y nuestro problema estaría resuelto.

Si para la cola GI/M/c pretendemos determinar la distribución del tiempo de espera en la cola, W, tendríamos que la probabilidad de que se le atienda inmediatamente es:

$$P(W=0) = \sum_{i=0}^{c-1} p_i = C \sum_{i=0}^{c-1} q_i = C\left(1/C - \sum_{i=c}^{\infty} q_i\right) = 1 - C\frac{r_0^c}{1-r_0}$$

Por otra parte, si cuando llega un cliente hay en la cola n individuos, $n \geq c$, sabemos que $n-c$ están esperando ser atendidos, con lo que el cliente que acaba de llegar empezará a ser atendido cuando hayan terminado $n+c-1$ servicios, y entonces:

$$t>0: P(W \leq t) = P(W=0) + \sum_{n=c}^{\infty} \underbrace{P(W \leq t \,/\, \text{hay } n \text{ en el sistema})}_{\Gamma(n-c+1,\, c\mu)} \cdot p_n$$

$$= P(W=0) + \sum_{n=c}^{\infty} \int_0^t \frac{(c\mu)^{n-c+1} v^{n-c} e^{-\mu c v}}{(n-c)!} C r_0^n \, dv$$

$$= P(W=0) + C r_0^c \int_0^t \sum_{n=c}^{\infty} \frac{(c\mu v r_0)^{n-c}}{(n-c)!} c\mu \, e^{-c\mu v} \, dv$$

$$= P(W=0) + C r_0^c \int_0^t \mu c \, e^{-\mu c (1-r_0) v} \, dv = C\left[\frac{1}{C} - \frac{r_0^c}{1-r_0}\right] + \frac{C r_0^c}{1-r_0}\left[1 - e^{-\mu c (1-r_0) t}\right]$$

$$P(W \leq t) = 1 - \frac{C r_0^c}{1-r_0} e^{-\mu c (1-r_0) t} \qquad t>0$$

con lo que $f_W(t) = C \mu c r_0^c e^{-\mu c (1-r_0) t}$ $\qquad t>0$

$$E(W) = 0 \cdot P(W=0) + \int_0^{+\infty} t f_W(t)\, dt = -\frac{C r_0^c}{1-r_0} \int_0^{\infty} t\,(-\mu c (1-r_0))\, e^{-\mu c (1-r_0) t} \, dt =$$

$$= \frac{C r_0^c}{1-r_0} \int_0^{+\infty} e^{-\mu c (1-r_0) t} \, dt = \frac{C r_0^c}{1-r_0} \frac{1}{\mu c (1-r_0)} = \frac{C r_0^c}{\mu c (1-r_0)^2} = \frac{1}{\mu c (1-r_0)\left[1 + \sum_{j=0}^{c-1} q_j\right]}$$

Sep 86

Tema 8

(5) Consideremos un sistema atendido por un único servidor en el que las llegadas siguen una distribución erlangiana de tipo 2 con tiempo medio entre llegadas de 30 min. El tiempo medio de servicio es de 25 min y los tiempos de servicio se distribuyen exponencialmente.

a) Determinar la probabilidad de equilibrio en el instante de llegada de nuevo cliente y el tamaño esperado de la cola.

b) ¿cuál sería la distribución del tiempo de espera? ¿y el tiempo medio de espera?

(G-II p. 176)

◇ Se trata de una cola $E_2/M/1$

siendo $E_2 \equiv \Gamma(2, 2\lambda)$ $\qquad \lambda = \frac{1}{30}$ min $= 2$ hr

$\frac{1}{\mu} = 25$ min $\longrightarrow \mu = \frac{1}{25}$ min $= \frac{12}{5}$ hr

La condición de ergodicidad es $\rho < 1$, siendo $\rho = \frac{\lambda}{\mu} = \frac{2}{12/5} = 5/6 < 1$

a) La distribución de equilibrio es:

$$\pi_n = (1-s_0)\, s_0^{\,n} \qquad n \geq 0$$

siendo s_0 la solución en $(0,1)$ de la ecuación $f(s) = s$, siendo $f(s) = \sum_{k=0}^{\infty} a_k s^k$ la función generatriz de probabilidad de los a_k's. En nuestro caso:

$$a_k = (2\lambda)^2 \int_0^{\infty} e^{-(\mu+2\lambda)t} \frac{(\mu t)^k}{k!}\, t\, dt$$

$$f(s) = \frac{1}{\left[1 + \frac{1-s}{2\rho}\right]^2}$$

Se ha visto que $f(s) = s$, $0<s<1 \longleftrightarrow$ tomando $r = s^{1/2}$, $|r|<1$: $r^3 - r(1+2\rho) + 2\rho = 0$

Dado que $\rho = 5/6$: $3r^3 - 8r + 5 = 0$, tenemos que $r = 1$ es una solución de esta ecuación (también se verifica $f(1)=1$), luego:

$$3r^3 - 8r + 5 = (r-1)(3r^2 + 3r - 5) = 0$$

Resolvemos $3r^2+3r-5=0$ $\quad r=\frac{-3\pm\sqrt{9+60}}{6}=\frac{-3\pm 8.24}{6}=$ $\frac{-11.24}{6}>-1$ no vale; 0.873

luego $r_0=0.873 \longrightarrow s_0=(0.873)^2=0.762$. En consecuencia:

$$\pi_n=(1-s_0)\cdot s_0^n = 0.238\,(0.762)^n \quad n\ge 0, \quad \text{con} \quad \pi_0=0.238$$

Nótese que la distribución de equilibrio es geométrica de parámetro $\pi_0=0.238$ siendo

$$E(Q)=\frac{0.762}{0.238}=\underline{\underline{3.2}}$$

b) Sea W el tiempo de espera en la cola, antes de ser servido. Supongamos que la cola está ya en régimen de equilibrio, se ha visto que:

$$P(W\le t)=(1-s_0)+s_0\{1-e^{-\mu(1-s_0)t}\} \qquad t\ge 0$$

i.e. $$P\{W\le t\}=0.238+0.762\{1-e^{-0.56t}\} \qquad t\ge 0$$

Por otra parte sabemos que: $E(W)=\int_0^\infty P(W>t)\,dt=\int_0^\infty [1-P(W\le t)]\,dt$, luego puede calcularse el tiempo medio de espera:

$$E(W)=\int_0^\infty 0.762\{1-1+e^{-0.56t}\}\,dt=\int_0^\infty 0.762\,e^{-0.56t}\,dt=0.762\left(-\frac{1}{0.56}\right)\left[e^{-0.56t}\right]_0^\infty$$

$$\frac{s_0}{\mu(1-s_0)}=\frac{0.762}{0.56}=\underline{\underline{1.36}}\text{ hr.}$$

Además, el tiempo medio de permanencia en la cola será:

$$E(T)=E(W)+\frac{1}{\mu}=1.36+0.41=1.71\text{ hr.}$$