

• Estado k > 1:  $\lambda p_k + \mu p_k = \lambda p_{k-1} + \mu p_{k+1}$ 

El sistema anterior más la condición de normalización  $\sum_k p_k = 1$  determinan completamente la solución  $\{p_k\}$ . Para resolver el sistema anterior basta observar que:

$$\lambda p_k - \mu p_{k+1} = \lambda p_{k-1} - \mu p_k$$

$$= \lambda p_{k-2} - \mu p_{k-1}$$

$$= \dots$$

$$= \lambda p_0 - \mu p_1 = 0$$
(8.4)

luego

$$p_{k+1} = (\lambda \ over \mu) p_k = (\frac{\lambda}{\mu})^2 p_{k-1} = \dots = (\frac{\lambda}{\mu})^{n+1} p_0$$
 (8.5)

usando  $\sum_{k=0}^{\infty} p_k = 1$  se tiene que

$$p_0 \cdot \sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k p_k = 1 \tag{8.6}$$

la ecuación anterior admite solución si sólo si  $\rho = \frac{\lambda}{\mu} < 1$ , en dicho caso se tiene que

$$p_0 = 1 - \frac{\lambda}{\mu} = 1 - \rho \tag{8.7}$$

y en general

$$p_k = (1 - \rho) \cdot \rho^k \tag{8.8}$$

es decir tiene una distribución geométrica. Anteriormente vimos que la condición  $\lambda < \mu$  era necesaria para la existencia de probabilidades estacionarias en un sistema G/G/1. Ahora bien, del desarrollo anterior vemos que para el caso M/M/1 la condición es además suficiente.

### Medidas de Efectividad

El estudio práctico de un sistema de espera no se limita a conocer cual es la distribución de probabilidades estacionarias, en lo que realmente se está interezado es en determinar las

medidas de efectividad del sistema como por ejemplo número promedio de entidades, tiempo promedio de permanencia en el sistema, fracción del tiempo que el servidor está ocupado, etc. Estas medidas de efectividad reflejan el funcionamiento del sistema y permiten apoyar decisiones sobre el manejo de los sistemas de espera. Por ejemplo si al momento de diseñar una cola se considera un sólo servidor y se calcula que en estas condiciones el 98% del tiempo el servidor estará ocupado entonces parece razonable pensar en aumentar el número de servidores a utilizar, la conclusión sería distintas si se hubiese obtenido una fracción de ocupación del 40%. Sea N la variable aleatoria número de entidades en el sistema y W la

variable aleatoria tiempo de permanecia en el sistema. En forma análoga se pueden definir  $N_Q$  y  $W_Q$  para la cola. El número promedio de entidades en el sistema se calcula como <sup>1</sup>

$$L = E(N) = \sum_{k=0}^{\infty} k p_k = \sum_{k=1}^{\infty} k (1 - \rho) \rho^n = \frac{\rho}{1 - \rho}$$
 (8.9)

Utilizando la formula de Little se tiene que

$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda \cdot (1 - \rho)} = \frac{1}{\mu - \lambda} \tag{8.10}$$

Además, es directo que  $W_S = \frac{1}{\mu}$  y por tanto es posible calcular  $W_Q = W - W_S$ , es decir

$$W_Q = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu \cdot (\mu - \lambda)} = \rho \cdot W \tag{8.11}$$

Finalmente conocido  $W_Q$ ,  $L_Q$  se obtiene usando Little

$$L_Q = \lambda \cdot W_Q = \frac{\lambda^2}{\mu \cdot (\mu - \lambda)} = \rho \cdot L \tag{8.12}$$

Las medidas de efectividad anteriormente calculadas representan los valores medios del número de entidades en el sistema o en la cola  $(L, L_Q)$  y también del tiempo de permanecia en el sistema o en la cola  $(W, W_Q)$ . Si bien esta información es de por sí es muy útil para el estudio de un sistema, puede ser insuficiente por no entregar un medida de la variabilidad. Es necesario conocer también como es la varianza de N y W por ejemplo para tener una visión más completa del funcionamiento del sistema. Varianza de N

La varianza del número de entidades en el sistema var(n) se puede calcular fácilmente usando la relación:

$$var(N) = E(N^2) - [E(N)]^2$$
 (8.13)

Dejamos propuesto al lector chequear que:

$$E(N^2) = \frac{\rho^2 + \rho}{(1 - \rho)^2} \tag{8.14}$$

y que por tanto  $var(N) = \frac{\rho}{(1-\rho)^2}$ . Se puede ver tanto de E(N) como de var(N) que el comportamiento del sistema se hace más inestable cuando  $\rho \to 1$ . En estos casos el número

<sup>&</sup>lt;sup>1</sup>El resultado se obtiene facilmente utilizando la relación  $\sum_n n \rho^n = \rho \cdot \frac{d}{d\rho} (\sum_n \rho^n)$ .

promedio de entidades en el sistema es muy grande al igual que la varianza lo que implica que el número de entidades observadas en el sistema en un instante cualquiera tiene una probabilidad alta de ser muy diferente del valor medio. <u>Varianza de W</u>

Para poder calcular var(W) es necesario conocer cuál es la distribución de probabilidades de esta variable. Para ello observemos los siguiente, supongamos que una entidad E llega al sistema y encuentra n entidades en su interior, entonces el tiempo que deberá permanecer antes de salir  $W_n$  puede escribirse como:

$$W_n = v_1 + v_2 + \dots + v_{n+1} (8.15)$$

en donde  $v_1'$  representa el tiempo residual de atención de la entidad que está siendo atendida cuando llega  $E, v_2,...,v_n$  representan los tiempos de atención de las n-1 entidades en la cola al momento de llegar E y  $v_{n+1}$  es el tiempo de atención de E. Como el proceso de atención es exponencial entonces  $v_2,...,v_{n+1}$  son claramente variables aleatorias independientes exponencialmente distribuidas, más aún la falta de memoria de la distribución exponencial permite asegurar que  $v_1$  también está exponencialmente distribuida. Por lo tanto,  $W_n$  es la suma de n+1 variables aleatorias exponecial independientes de tasa  $\mu$  por lo que  $W_n$  tiene una distribución gamma de parámetros  $\mu$  y n+1 y su función de densidad viene dada por:

$$f_{W_n}(w) = \frac{\mu^{n+1} \cdot w^n \cdot e^{-\mu \cdot w}}{\Gamma(n+1)}$$
 (8.16)

Sea por otro lado  $f_W(w)$  la función de densidad del tiempo de permanencia en el sistema. Entonces la probabilidad que una entidad permanezca en el sistema un tiempo comprendido entre [w, w + dw] viene dada por:

$$Prob(w \le W \le w + dw) = f_W(w) \cdot dw \tag{8.17}$$

Además la probabilidad anterior puede reescribirse condicionándola al número de entidades que nuestra entidad de prueba E detecta al llegar como:

$$Prob(w \le W \le w + dw) = \sum_{n=0}^{\infty} Prob(w \le W_n \le w + dw) \cdot p_n$$

$$= \sum_{n=0}^{\infty} \frac{\mu^{n+1} \cdot w^n \cdot e^{-\mu \cdot w}}{\Gamma(n+1)} \cdot (1-\rho)\rho^n \cdot dw$$

$$= \mu(1-\rho)e^{-\mu \cdot w} \sum_{n=0}^{\infty} \frac{(\mu w \rho)^n}{\Gamma(n+1)} dw$$

$$= \mu(1-\rho)e^{-\mu(1-\rho)w} dw$$
(8.18)

Luego igualando 8.17 con 8.18 se tiene que

$$f_W(w) = \mu(1-\rho)e^{-\mu(1-\rho)w}$$
(8.19)

Por lo tanto W sigue una distribución exponencial de tasa  $\mu(1-\rho)$ . Luego la varianza de W viene dada por  $var(W) = \frac{1}{(\mu(1-\rho))^2}$  y nuevamente los problemas de variabilidad para W se alcanzan cuando  $\rho \to 1$ .

# 8.4.2 Sistema M/M/c

Consideremos un sistema de espera cuyo proceso de llegada es poissoniano de tasa  $\lambda$  y que dispone de c servidores en paralelo, teniendo cada uno un tiempo de atención exponencialmente distribuido con tasa  $\mu$ .

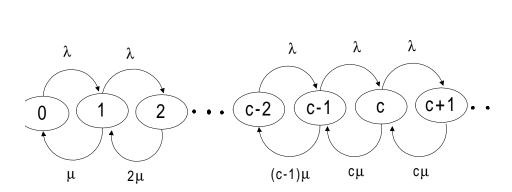
## Probabilidades Estacionarias

El sistema M/M/c corresponde a un sistema de nacimiento y muerte para el cual las tasa de transición entre estados no son siempre iguales. En efecto si existen n entidades en el sistema y n < c entonces sólo n servidores estarán ocupados y el tiempo entre dos salidas consecutivas del sistema está exponencialmente distribuidas con tasa  $n \cdot \mu$ . Si se tiene en cambio que  $n \ge c$  entonces todos los servidores estarán ocupados y el tiempo entre salidas consecutivas del sistema está exponencialmente distribuido con tasa  $c \cdot \mu$ . Por lo tanto, el sistema M/M/c es un proceso de nacimiento y muerte con tasa de nacimiento constante  $\lambda_n = \lambda$  y con tasa de muerte  $\mu_n$  dependiente del estado n del sistema con:

$$\mu_n = n \cdot \mu$$
  $n = 0, 1, 2, ..., c - 1$ 

$$\mu_n = c \cdot \mu$$
  $n = c, c + 1, \dots$ 

El grafo representante se muestra en la Figura 8.2:



Suponiendo que el estado estacionario existe entonces la probabilidad estacionaria de que el sistema se encuentre en el estado n  $(p_n)$  viene dada por:

$$p_n = \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^n \mu_i} p_0 \tag{8.20}$$

es decir,

$$p_{n} = \frac{\lambda^{n}}{n! \cdot \mu^{n}} p_{0} \quad n = 0, 1, 2, ...., c - 1$$

$$p_{n} = \frac{\lambda^{n}}{c! \cdot c^{n-c} \cdot \mu^{n}} p_{0} \quad n = c, c + 1, ....$$
(8.21)

Utilizando el resultado anterior y el hecho que  $\sum_{n=0}^{\infty} p_n = 1$  se obtiene

$$p_0 = \left[ \sum_{n=0}^{c-1} \frac{\lambda^n}{n! \cdot \mu} + \frac{\lambda^c}{c! \cdot \mu^c \cdot (1 - \frac{\lambda}{c\mu})} \right]^{-1}$$
(8.22)

El resultado anterior asume que se satisface la condición  $\lambda < c\mu$  que es condición necesaria y suficiente para la existencia del estado estacionario para un sistema M/M/c. Una pregunta

interesante en este modelo es la que tiene relación con la probabilidad que una entidad que llega al sistema tenga que esperar para ser atendidad o equivalentemente que encuentre a los c servidores ocupados.

$$C \equiv Prob(N \ge c) = \sum_{n=c}^{\infty} p_n = \frac{p_c}{1-\rho}$$

donde  $\rho = \frac{\lambda}{c\mu}$ . La expresión anterior se conoce como la fórmula C de Erlang y aparece tabulada para valores diferentes de c y  $\frac{\lambda}{\mu}$ .

#### Medidas de Efectividad

#### Número Esperado de Servidores Ocupados

Una medida de efectividad para estudiar el dimensionamiento en términos del número de servidores a utilizar es el número promedio de servidores que están ocupados en un momento dado. Mientras más cercano es este valor al número de servidores totales mayor será la tasa de ocupación de estos. A partir de las relaciones de Little se tiene que  $L=\lambda W$  y  $L_Q=\lambda W_Q$  y por tanto  $L-L_Q=\lambda(W-W_Q)$ . El término  $L-L_Q$  representa el número promedio de entidades que están siendo atendidas en un momento dado del tiempo y es exactamente igual al número de servidores ocupados en ese momento. Por otro lado,  $W-W_Q$  representa el tiempo de atención de una entidad el cual viene dado por  $\frac{1}{\mu}$ . Por lo tanto, el número promedio de servidores ocupados es igual a  $\frac{\lambda}{\mu}=c\rho$ . Número Esperado de entidades en el sistema

El número esperado de entidades en el sistema se puede calcular como la suma del número esperado de entidades siendo atendidas más el número esperado de entidades en la cola. Del punto anterior el número esperado de entidades siendo atendidas es  $c\rho$ . Por otro lado, el número esperado de entidades en la cola viene dado por:

$$L_Q = \sum_{n=c}^{\infty} (n-c) \cdot p_c$$

$$= \sum_{n=c}^{\infty} (n-c) \frac{\lambda^n}{\mu^n \cdot c! \cdot c^{n-c}} p_0$$

$$= \frac{\rho}{1-\rho} Prob(N \ge c)$$
(8.23)

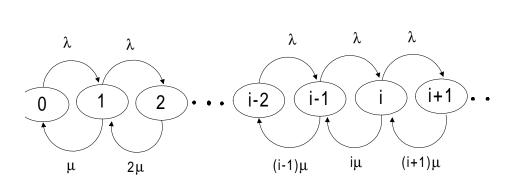
Por lo tanto, el número esperado de entidades en el sistema es igual a  $L = c\rho + \frac{\rho C}{1-\rho}$  con  $C = Prob(N \ge c)$ . A partir de la fórmula de Little se pueden deducir facilmente los valores de W y  $W_Q$ .

$$W_Q = \frac{L_Q}{\lambda} = \frac{C}{c \cdot \mu \cdot (1 - \rho)}$$

$$W = \frac{L}{\lambda} = \frac{1}{\mu} + W_Q$$
(8.24)

# 8.4.3 Sistema $M/M/\infty$

Consideremos un sistema con llegadas poissonianas, tiempos de atención exponenciales y que tenga un número ilimitado de servidores. Este tipo de modelos se ajusta bien a situaciones de autoservicio, es decir, en donde cada entidad que llega al sistema se proporciona el servicio. Este sistema puede modelarse como un proceso de nacimiento y muerte con el grafo representante que se muestra en la Figura 8.3:



Las probabilidades estacionarias vienen dadas en este caso por:

$$p_n = \frac{\lambda^n}{\mu^n \cdot n!} p_0 \tag{8.25}$$

Luego imponiendo que  $\sum_{n=0}^{\infty} p_n = 1$  se tiene que:

$$p_n = \frac{\lambda^n}{\mu^n \cdot n!} e^{\frac{-\lambda}{\mu}} \qquad \forall n = 0, 1, 2, ....$$
 (8.26)

Es decir, el número de personas en el sistema tiene una distribución Poisson de media  $L = \frac{\lambda}{\mu}$ . El tiempo promedio de permanencia de una entidad en este sistema es claramente  $\frac{1}{\mu}$ , ya al exsistir capacidad ilimitada en la atención no se forma cola y el tiempo en el sistema es igual al tiempo de atención cuya media es  $\frac{1}{\mu}$ .

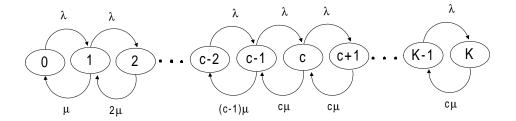
# 8.4.4 Sistema M/M/c/K

Consideremos un sistema de espera cuyo proceso de llegada es Poisson de tasa  $\lambda$  y que dispone de c servidores en paralelo, teniendo cada uno un tiempo de atención exponencialmente distribuido con tasa  $\mu$ . El sistema tiene capacidad finita, no pudiendo haber más de K entidades simultáneamente en el sistema (i.e. sólo hay K-c lugares de espera). Podemos asumir  $c \leq K$  pues si c > K el sistema se comporta igual que un M/M/K/K.

#### Probabilidades Estacionarias

El sistema M/M/c/K corresponde a un sistema de nacimiento y muerte en el cual las tasas de transición entre estados son iguales a las del sistema M/M/c para los estados  $0, 1, \ldots, K$  y la tasa de transición a estados con índice mayor que K es nula. Si existen n entidades en el sistema y n < c entonces sólo n servidores estarán ocupados y el tiempo entre dos salidas consecutivas del sistema (muertes) está exponencialmente distribuidas con tasa  $\mu_n = n \cdot \mu$ . Si se tiene en cambio que  $n \ge c$  entonces todos los servidores estarán ocupados y el tiempo entre salidas consecutivas del sistema está exponencialmente distribuido con tasa  $\mu_n = c \cdot \mu$ . Por otro lado, si hay n < K entidades en el sistema la tasa de nacimiento es  $\lambda_n = \lambda$ , mientras que si hay n = K entidades la tasa de nacimiento es  $\lambda_n = 0$ . El grafo representante se muestra en la Figura 8.4:

Figura 8.4: Grafo representante de una cola M/M/C/K



Como vemos, la evolución del número de entidades en el sistema puede ser representado mediante una cadena de Markov en tiempo continuo irreducible y finita, de modo que con seguridad existe una ley de probabilidades estacionarias, no importa cuál sea la relación entre  $\lambda, \mu, cyK$  (asumiendo  $\lambda > 0, \mu > 0, 0 < c \le K < \infty$ ).

La probabilidad estacionaria de tener n entidades en el sistema,  $p_n$  viene dada por:

$$p_n = \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}} p_0 \tag{8.27}$$

es decir,

$$p_n = \begin{cases} \frac{\lambda^n}{n! \cdot \mu^n} p_0 & n < c \\ \frac{\lambda^n}{c! \cdot c^{n-c} \cdot \mu^n} p_0 & c \le n \le K \end{cases}$$

$$(8.28)$$

Utilizando el resultado anterior y el hecho que  $\sum_{n=0}^{\infty} p_n = 1$  se obtiene

$$p_0 = \left[ \sum_{n=0}^{c-1} \frac{\lambda^n}{n! \cdot \mu} + \frac{\lambda^c \cdot c^c}{c! \cdot \mu^c} \cdot \frac{1 - \left(\frac{\lambda}{c\mu}\right)^{K+1}}{1 - \frac{\lambda}{c\mu}} \right]^{-1}$$
(8.29)

### Medidas de Efectividad

Entre las medidas de efectividad para este sistema resulta de especial interés la tasa de pérdida de entidades en el largo plazo. Una entidad que llega cuando el sistema está lleno no puede entrar, y se retira sin haber recibido servicio. Diremos que esa entidad se ha perdido. Dado que el proceso de llegadas es Poisson, la probabilidad que una entidad que llega encuentre el sistema lleno es igual a  $p_K$ , la probabilidad estacionaria que el sistema esté lleno (recordar "PASTA"). Así, en el largo plazo una fracción  $p_K$  de las entidades encuentra el sistema lleno, y se pierden, de modo que la tasa de pérdida de entidades es igual a  $\lambda p_K$  [entidades/u. de tiempo].

Podemos calcular el número medio de entidades en el sistema como

$$L = \sum_{i=0}^{K} i \cdot p_i$$

. Una vez conocido L podemos calcular el tiempo medio que pasa una entidad en el sistema, W, a partir de la Fórmula de Little. Sin embargo se debe tener cuidado al aplicar aquí la fórmula de Little: hay que tomar en cuenta que al sistema sólo entran las entidades que no lo encuentran lleno, de manera que la tasa efectiva de entrada al sistema es  $\lambda_{ef} = \lambda(1-p_K)$  [entidades/u. de tiempo]. De esa forma se tiene

$$W = \frac{L}{\lambda_{ef}} = \frac{L}{\lambda(1 - p_K)}$$

.

# 8.4.5 Sistema $M/M/1/\infty/N$

Consideramos ahora un sistema en que las llegadas provienen de una población finita, de N entidades. Una entidad que está fuera del sistema llegará a él en un tiempo exponencialmente distribuido de media  $1/\lambda$  [u. de tiempo], independiente de las demás. Los tiempos de atención son variables aleatorias i.i.d con distribución exponencial con media  $1/\mu$  [u. de tiempo].

#### Probabilidades Estacionarias

Este sistema es susceptible de ser modelado como un proceso de nacimiento y muerte. Cuando hay i entidades en el sistema hay N-i fuera de él. El tiempo que transcurre hasta la llegada de la próxima entidad es el mínimo de los tiempos de llegada de cada una de las N-i entidades que están fuera del sistema, y como todos ellos son exponenciales de tasa  $\lambda$ , la tasa de nacimientos es  $(N-i)\lambda$ , para  $0 \le i < N$ . Cuando hay N entidades en el sistema la tasa de nacimientos es nula (no hay ninguna entidad fuera que pueda llegar al sistema). La tasa de muerte es constante e igual a  $\mu$  mientras haya un número positivo de entidades en el sistema. El grafo representante se muestra en la Figura 8.5.

