20

Teoría de líneas de espera (también teoría de colas)

Todos nosotros hemos pasado mucho tiempo esperando en una cola. En este capítulo se estudian ciertos modelos matemáticos para las líneas de espera. Se empieza por tratar los términos que se usan con mayor frecuencia en las líneas de espera en la sección 20.1. Luego se examinan algunas distribuciones (las distribuciones exponencial y Erlang) en la sección 20.2; dichas distribuciones son necesarias para explicar los modelos de líneas de espera. En la sección 20.3 se presenta la idea de un proceso de nacimiento-muerte, el cual es básico en muchos modelos de líneas de espera que requieren distribución exponencial. Diversos modelos de sistemas de líneas de espera se tratan en el resto del capítulo; estos modelos se pueden usar para responder preguntas como las siguientes:

1 ¿Cuánto tiempo está ocioso cada servidor?

2 ¿Cuál es el número esperado de clientes presentes en la cola?

3 ¿Cuál es el tiempo previsto que un cliente debe pasar en la cola?

4 ¿Cuál es la distribución de probabilidad de la cantidad de clientes presentes en la cola?

5 ¿Cuál es la distribución de probabilidad del tiempo de espera de un cliente?

6 Si un gerente de un banco desea tener la certeza de que sólo 1% de todos los clientes tendrá que esperar más de cinco minutos a que lo atienda un cajero, ¿cuántos cajeros se deben emplear?

20.1 Terminología para las líneas de espera

Es necesario especificar un proceso de entrada y uno de salida para describir un sistema de líneas de espera. Algunos ejemplos de procesos de entrada y de salida se proporcionan en la tabla 1.

Proceso de entrada o llegada

El proceso de entrada se denomina, por lo regular, **proceso de llegada**. Las llegadas se llaman **clientes**. En todos los modelos que se estudian, se supone que no más de una llegada ocurre en un instante dado. En el caso de un restaurante, es una suposición irreal. Si hay más de una llegada en un instante dado se dice que se permiten **llegadas en masa**.

Se supone, por lo común, que el número de clientes presentes en el sistema no afecta el proceso de llegada. En el contexto de un banco, esto requeriría que si hay 500 o cinco personas en el lugar, el proceso que rige las llegadas permanece sin cambios.

Hay dos situaciones comunes en las cuales el proceso de llegadas podría depender de la cantidad de clientes presentes. La primera se presenta cuando las llegadas se extraen de una pequeña población. Suponga que hay sólo cuatro barcos en un astillero. Si los cuatro barcos están en reparación, entonces ningún barco puede estropearse en un futuro cercano. Por otro lado, si los cuatro barcos están navegando, existe una probabi-

TABLA 1
Ejemplos de sistemas de colas

Situación	Proceso de llegada	Proceso de salida
Banco	Los clientes llegan al banco	Los cajeros atienden a los clientes
Pizzería	Se reciben los pedidos de pizzas	La pizzería envía al mensajero a entregar pizzas
Banco de sangre de un hospital	Llegada de sangre	Los pacientes utilizan la sangre
Astillero	Los barcos que estan navegando sufren descomposturas y se envian al astillero para su reparación	Los barcos son reparados y regresan al mar

lidad muy alta de una descompostura en el futuro cercano. Los modelos en los cuales las llegadas se toman de una pequeña población reciben el nombre de **modelos de origen finito**. Otra situación en la cual el proceso de llegadas depende de la cantidad de clientes presentes ocurre cuando la razón a la cual llegan los clientes a cierta instalación disminuye cuando ésta se llena. Por ejemplo, si usted ve que el estacionamiento del banco está lleno, se retira y va otro día. Si un cliente llega, pero no entra al sistema, se dice que el cliente **ha renunciado**. El fenómeno de la renuncia fue descrito por Yogi Berra cuando dijo: "Ya nadie va a ese restaurante; siempre está lleno."

Si el número de clientes presente no afecta al proceso de llegada, entonces se le describe a éste mediante la especificación de la distribución de probabilidad que rige el tiempo entre llegadas sucesivas.

Proceso de salida o de servicio

Para describir el proceso de salida (con frecuencia se le llama proceso de servicio) de un sistema de líneas de espera, se especifica una distribución de probabilidad –**distribución del tiempo de servicio**– la cual rige el tiempo de servicio a un cliente. Se supone que, en la mayoría de los casos, la distribución del tiempo de servicio es independiente de la cantidad de clientes presentes. De aquí se infiere que el servidor, o canal, no trabaja más rápido cuando hay más clientes presentes.

En este capítulo se tratan dos tipos de servidores: **servidores en paralelo** y **servidores en serie**. Los servidores están en paralelo si todos ofrecen el mismo tipo de servicio y un cliente sólo requiere pasar por un servidor para completar el servicio. Por ejemplo, los cajeros de un banco están en paralelo; cualquier cliente sólo necesita ser atendido por un cajero, y cualquier cajero puede proporcionar el servicio deseado. Los servidores están en serie cuando un cliente debe pasar por varios servidores antes de terminar el servicio. Una línea de ensamble es un ejemplo de un sistema de espera en serie.

Disciplina de las líneas de espera

Para describir por completo un sistema de líneas de espera, se debe describir también la disciplina de las líneas de espera y el modo en el cual los clientes forman las líneas de espera.

La **disciplina de las líneas de espera** explica el método usado para determinar el orden con el cual se atiende a los clientes. La disciplina más común es la **disciplina FCFS** (*first come, first served*, es decir, al primero que llega se le atiende primero), en la cual se atiende a los clientes según el orden en que llegan. En la **disciplina LCFS** (*last come, first served*, el último en llegar es el primero en salir), las llegadas más recientes son los primeros en entrar al servicio. Si se considera la salida de un elevador como un servicio, entonces un elevador lleno ilustra una disciplina LCFS. El orden en el cual los clientes llegan no afecta, algunas veces, el orden en el cual son atendidos. Éste sería el caso cuando el siguiente cliente en pasar al servidor es elegido en forma aleatoria de entre los clientes que están esperando atención. A esta situación se le denomina **disciplina SIRO** (*service in random order*, servicio en orden aleatorio). Cuando a una persona que llama a una aerolínea se le hace esperar, la suerte determina con frecuencia quién será la siguiente persona en ser atendida por un operador.

Se consideran, por último, las **disciplinas de prioridad en las colas**. Una disciplina de prioridad clasifica cada llegada en una categoría. Cada categoría recibe luego un nivel de prioridad, y dentro de cada nivel de prioridad, los clientes entran al servicio de acuerdo a FCFS. Las disciplinas de prioridad se usan a menudo en salas de urgencia con el objeto de determinar el orden en el cual los pacientes reciben atención, así como en las instalaciones de copiado y para compartir computadoras, donde la prioridad se da a trabajos con tiempos de proceso más cortos.

Método usado para que las llegadas se sumen a las colas

Otro factor de gran efecto en el comportamiento de los sistemas de las líneas de espera es el método que los clientes utilizan para determinar a qué cola sumarse. Por ejemplo, en algunos bancos, los clientes deben formar una sola cola, pero en otros bancos los clientes eligen la cola en la que desean formarse. Cuando hay varias colas, los clientes se forman en la más corta. En muchas situaciones, por ejemplo en el supermercado, es difícil definir cuál es la cola más corta. Si hay varias colas, es importante saber si se permite a los clientes cambiarse de cola o adelantarse en la cola. En la mayor parte de sistemas de líneas de espera con varias colas está permitido adelantarse en la cola, pero no se recomienda hacerlo en una caseta de cobro.

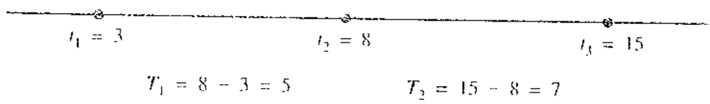
20.2 Modelado de procesos de llegada y servicio

Modelado de procesos de llegadas

Como ya se mencionó, suponemos que la mayor parte de las llegadas ocurre en un instante dado. Definamos t_i como el tiempo en el cual llega el i-ésimo cliente. Observe la figura 1 para ejemplificar lo anterior. Para $i \geq 1$, definimos $T_i = t_{i+1} - t_i$ como el i-ésimo tiempo entre llegadas. Por lo tanto, en la figura, $T_1 = 8 - 3 = 5$ y $T_2 = 15 - 8 = 7$. Al modelar el proceso de llegadas, suponemos que las T_i son variables continuas, aleatorias e independientes descritas por la variable aleatoria **A**. La suposición de independencia significa, por ejemplo, que el valor de T_2 no afecta al valor de T_3, T_4 o cualquier otra T_i superior. La suposición de que cada T_i es continua es, por lo regular, una buena aproximación de la realidad. Después de todo, un tiempo entre llegadas no necesita ser exactamente de uno o dos minutos; podría ser con toda facilidad 1.55892 minutos. La suposición de que cada tiempo entre llegadas está regido por la misma variable aleatoria, implica que la distribución de llegadas es independiente del momento del día o del día de la semana. Esta es la suposición de los tiempos estacionarios entre llegadas. Debido a fenómenos como las horas pico, esta última suposición es a menudo irreal, pero podríamos aproximarnos con frecuencia a la realidad descomponiendo la duración del día en segmentos. Por ejemplo, si estuviéramos modelando el flujo del tránsito, podríamos descomponer el día en tres partes: un segmento de horas pico por la mañana, un segmento a mediodía y otro por la noche. Durante cada uno de estos segmentos los tiempos entre llegadas podrían ser estacionarios.

Suponemos que **A** tiene una función de densidad $a(t)$. Según la sección 12.5, para Δt pequeñas, $P(t \leq \mathbf{A} \leq t + \Delta t)$ es de alrededor $\Delta t a(t)$. Naturalmente un tiempo entre llegadas negativo es imposible. Todo esto nos permite escribir:

FIGURA 1
Definición de tiempos entre llegadas



$$P(\mathbf{A} \leq c) = \int_0^c a(t)dt \qquad \text{y} \qquad P(\mathbf{A} > c) = \int_c^\infty a(t)dt \tag{1}$$

Definimos $\frac{1}{\lambda}$ como el tiempo medio o promedio entre llegadas. Sin perder la generalidad, suponemos que el tiempo se mide en unidades de horas. Entonces, las unidades de $\frac{1}{\lambda}$ son horas por llegada. Según la sección 12.5, podríamos estimar $\frac{1}{\lambda}$ a partir de $a(t)$ mediante la ecuación siguiente:

$$\frac{1}{\lambda} = \int_0^\infty ta(t)dt \tag{2}$$

Definimos λ como la **tasa de llegadas**, la cual tiene unidades de llegadas por hora.

En la mayor parte de las aplicaciones de líneas de espera, un aspecto importante es cómo escoger **A** de tal manera que refleje la realidad y siga siendo manejable desde el punto de vista del cálculo. La elección más común para **A** es la **distribución exponencial**. Una distribución exponencial con parámetro λ tiene una densidad $a(t) = \lambda e^{-\lambda t}$. En la figura 2 se ilustra la función de densidad para una distribución exponencial. Se observa que $a(t)$ disminuye con rapidez para t pequeñas. Esto quiere decir que son improbables los tiempos entre llegadas muy grandes. Se puede demostrar, mediante la ecuación (2) y la integración por partes, que el tiempo medio o promedio entre llegadas (llámelo $E(\mathbf{A})$) está dado por

$$E(\mathbf{A}) = \frac{1}{\lambda} \tag{3}$$

Debido al hecho de que var $\mathbf{A} = E(\mathbf{A}^2) - E(\mathbf{A})^2$, es posible demostrar que

$$\text{var } \mathbf{A} = \frac{1}{\lambda^2} \tag{4}$$

Propiedad de carencia de memoria de la distribución exponencial

La razón de que la distribución exponencial se use a menudo para modelar los tiempos entre llegadas se encuentra en el lema siguiente.

LEMA 1

Si **A** tiene una distribución exponencial, entonces para todos los valores no negativos de t y h,

$$P(\mathbf{A} > t + h | \mathbf{A} \geq t) = P(\mathbf{A} > h) \tag{5}$$

Demostración Observe primero que por la ecuación (1) se tiene

$$P(\mathbf{A} > h) = \int_h^\infty \lambda e^{-\lambda t} = [-e^{-\lambda t}]_h^\infty = e^{-\lambda h} \tag{6}$$

Entonces,

$$P(\mathbf{A} > t + h | \mathbf{A} \geq t) = \frac{P(\mathbf{A} > t + h \cap \mathbf{A} \geq t)}{P(\mathbf{A} \geq t)}$$

Según (6)

$$P(\mathbf{A} > t + h \cap \mathbf{A} \geq t) = e^{-\lambda(t+h)} \qquad \text{y} \qquad P(\mathbf{A} \geq t) = e^{-\lambda t}$$

Por lo tanto,

$$P(\mathbf{A} > t + h | \mathbf{A} \geq t) = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h} = P(\mathbf{A} > h)$$

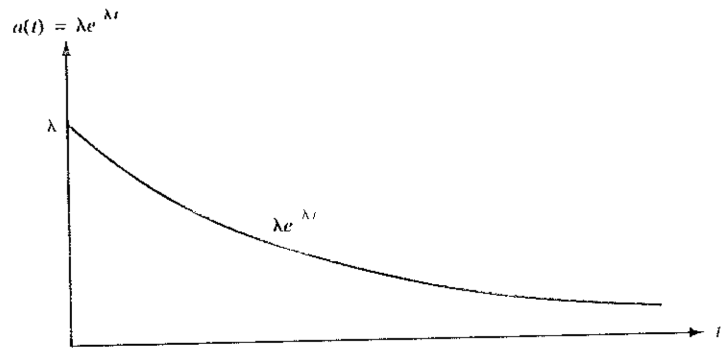


FIGURA 2
Función de densidad de la distribución exponencial

Puede demostrarse que ninguna otra función de densidad puede cumplir con (5) (ver Feller, 1957). Por razones que son naturales, una densidad que satisface (5) tiene la **propiedad de carencia de memoria**. Suponga que nos advierten que no ha habido llegada alguna en las últimas t horas (equivale a que nos digan que $\mathbf{A} \geq t$) y que nos preguntan cuál es la probabilidad de que no haya llegadas durante las siguientes h horas (es decir, $\mathbf{A} > t + h$). Entonces se infiere de (5) que esta probabilidad *no depende del valor de* t, y para todos los valores de t, esta probabilidad es igual a $P(\mathbf{A} > h)$. En resumen, si sabemos que han transcurrido por lo menos t unidades de tiempo desde que ocurrió la última llegada, entonces la distribución del tiempo restante hasta la siguiente llegada (h) no depende de t. Por ejemplo, si $h = 4$, entonces (5) genera, para $t = 5$, $t = 3$, $t = 2$ y $t = 0$,

$$\begin{aligned} P(\mathbf{A} > 9|\mathbf{A} \geq 5) &= P(\mathbf{A} > 7|\mathbf{A} \geq 3) = P(\mathbf{A} > 6|\mathbf{A} \geq 2) \\ &= P(\mathbf{A} > 4|\mathbf{A} \geq 0) = e^{-4\lambda} \end{aligned}$$

La propiedad de carencia de memoria de la distribución exponencial es importante porque implica que si queremos conocer la distribución de probabilidad del tiempo hasta la llegada siguiente, entonces *no importa cuánto tiempo ha transcurrido desde la última llegada*. En términos concretos, suponga que los tiempos entre llegadas están distribuidos en forma exponencial con $\lambda = 6$. Entonces, se infiere de la propiedad de carencia de memoria que no importa cuánto tiempo transcurrió desde la última llegada, la distribución de probabilidad que rige el tiempo hasta la llegada siguiente tiene la función de densidad de $6e^{-6t}$. Esto significa que para poder predecir patrones de llegadas futuras no necesitamos rastrear cuánto ha pasado desde la última llegada. Esta observación puede simplificar notoriamente el análisis del sistema de líneas de espera.

Para saber qué tiempo pasó desde la última llegada sí afecta la distribución del tiempo hasta la llegada siguiente en la mayor parte de las situaciones, suponga que $\mathbf{A}$ es discreta con $P(\mathbf{A} = 5) = P(\mathbf{A} = 100) = \frac{1}{2}$. Si nos avisan que no ha habido llegadas durante las últimas seis unidades de tiempo, *sabemos con certeza* que hay $100 - 6 = 94$ unidades de tiempo hasta la siguiente llegada. En cambio, si sabemos que no ha habido llegadas durante la última unidad de tiempo, entonces hay alguna probabilidad de que el tiempo hasta la siguiente llegada sea $5 - 1 = 4$ unidades de tiempo y alguna probabilidad que será de $100 - 1 = 99$ unidades de tiempo. Por lo tanto, en esta situación, la distribución de tiempo siguiente entre llegadas no se puede predecir con facilidad cuando se sabe qué tiempo transcurrió desde la última llegada.

Relación entre la distribución Poisson y la distribución exponencial

Si los tiempos entre llegadas son exponenciales, la distribución de probabilidad de la cantidad de llegadas que suceden en cualquier intervalo de duración t está dada por el importante teorema siguiente:

TEOREMA 1

Los tiempos entre llegadas son exponenciales, con el parámetro λ, si y sólo si la cantidad de llegadas en un intervalo de duración t sigue una distribución Poisson con parámetro λt.

Una variable discreta aleatoria N sigue una distribución de Poisson con parámetro λ si, para $n = 0, 1, 2, \ldots,$

$$P(\mathrm{N} = n) = \frac{e^{-\lambda}\lambda^n}{n!} \qquad (n = 0, 1, 2, \ldots) \tag{7}$$

Si N es una variable aleatoria tipo Poisson, se puede demostrar que $E(\mathrm{N}) = \mathrm{var}\ \mathrm{N} = \lambda$. Si definimos N_t como la cantidad de llegadas que ocurre durante cualquier intervalo de duración t, el teorema 1 establece que

$$P(\mathrm{N}_t = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \qquad (n = 0, 1, 2, \ldots)$$

Como N_t es Poisson con parámetro λt, λt, $E(\mathrm{N}_t) = \mathrm{var}\ \mathrm{N}_t = \lambda t$. Un promedio de λt llegadas se presenta durante un intervalo de duración t, de modo que se podría pensar que λ es el promedio de llegadas por unidad de tiempo, o bien, tasa de llegadas.

¿Qué suposiciones se requieren para que los tiempos entre llegadas sean exponenciales? El teorema 2 ofrece parte de la respuesta. Piense en las siguientes dos suposiciones:

1 Las llegadas definidas en intervalos de tiempos que no se traslapan son independientes (por ejemplo, las llegadas que se presentan entre los tiempos 1 y 10 no ofrecen información acerca del número de llegadas entre los tiempos 30 y 50).

2 Para Δt pequeñas (y cualquier valor de t), la probabilidad de que se presente una llegada entre los tiempos $t + \Delta t$ es $\lambda\Delta t + o(\Delta t)$, donde $o(\Delta t)$ se refiere a cualquier cantidad que satisfaga

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

Asimismo, la probabilidad de que no haya ninguna llegada durante el intervalo entre t y $t + \Delta t$ es $1 - \lambda\Delta t + o(\Delta t)$, y la probabilidad de que haya más de una llegada entre t y $t + \Delta t$ es $o(\Delta t)$.

TEOREMA 2

Si las suposiciones 1 y 2 se sostienen, entonces N_t sigue una distribución Poisson con parámetro λt, y los tiempos entre llegadas son exponenciales con parámetro λ; es decir, $a(t) = \lambda e^{-\lambda t}$.

En esencia, el teorema 2 establece que si la tasa de llegadas es estacionaria, si las llegadas en masa no ocurren y si las llegadas anteriores no afectan las llegadas futuras, entonces los tiempos entre llegadas siguen una distribución exponencial con parámetro λ, y la cantidad de llegadas en cualquier tiempo de duración t es Poisson con parámetro λt. Las suposiciones del teorema 2 podrían parecer demasiado restrictivas, pero los tiempos entre llegadas son, con frecuencia, exponenciales incluso si las suposiciones del teorema 2 no se satisfacen (véase Denardo (1982)). La manera en que se usan los datos para probar si la hipótesis de los tiempos entre llegadas exponenciales es razonable, se estudia en la sección 20.12. En muchas aplicaciones, resulta que la suposición de los tiempos entre llegadas exponenciales es una muy buena aproximación de la realidad.

Cálculo de las probabilidades de Poisson y exponenciales mediante Excel

Excel contiene funciones que facilitan el cálculo de las probabilidades que se relacionan con las variables tipo Poisson y las variables exponenciales aleatorias.

La sintaxis de la función POISSON de Excel es la que se indica:

- =POISSON(x,MEAN,TRUE) encuentra la probabilidad de que una variable aleatoria de Poisson con media = Media sea menor que o igual a x.

	A	B	C	D	E
3					
4	Poisson	Lambda			
5	P(X=40)	40	0.541918		
6					
7	P(X<=40)	40	0.062947		
8					
9	Exponencial				
10		Lambda			
11	P(X<=10)	0.1	0.632121		
12	Densidad para X = 10	0.1	0.036788		
13					
14					
15					
16					

FIGURA 3

- =POISSON(x,MEDIA,FALSO) =POISSON(x,MEAN,FALSE) en el programa en inglés] obtiene la probabilidad de que una variable aleatoria de Poisson con media = Media sea igual a x.

Por ejemplo, si un promedio de 40 clientes llega en una hora y las llegadas siguen una distribución Poisson, entonces la función =POISSON(40,40,VERDADERO) [(=POISSON(x,MEAN,TRUE) en el programa en inglés] da la probabilidad .542 de que 40 o menos clientes lleguen durante una hora. La función =POISSON(40,40,FALSE) da la probabilidad.063 de que *exactamente* 40 clientes llegan durante una hora.

La sintaxis de la función de Excel EXPONDIST es como se indica

- =DISTR.EXP(x,LAMBDA,VERDADERO) [=EXPONDIST(x,LAMBDA,TRUE) en el programa en inglés] da la probabilidad de que una variable aleatoria exponencial con parámetro λ asuma un valor menor que o igual a x.
- =DISTR.EXP(x,LAMBDA,FALSE) [=EXPONDIST(x,LAMBDA,FALSE) en el programa en inglés] da el valor de la función de densidad para una variable aleatoria exponencial con parámetro λ.

Por ejemplo, suponga que el tiempo promedio entre las llegadas siguen una distribución exponencial con media 10. Entonces $\lambda = 0.1$ y =DISTR.EXP(10,0.1,VERDADERO) = EXPONDIST(10,0.1,TRUE) da la probabilidad .632 de que el tiempo entre llegadas es 10 min o menos.

Poissexp.xls

La función =DISTR.EXP(10,0.1,FALSO) =EXPONDIST(10,0.1,FALSE) da el valor .037 de la función de densidad para $x = 10$ y $\lambda = 0.1$. Véase el archivo Poissexp.xls y la figura 3.

Por medio del ejemplo 1 se ilustra la relación entre las distribuciones exponencial y de Poisson.

EJEMPLO 1 Pedidos de cerveza

La cantidad de vasos de cerveza ordenada por hora en la cantina de Dick sigue una distribución Poisson, con un promedio de 30 cervezas pedidas por hora.

1 Estime la probabilidad de que se pidan exactamente 60 cervezas entre las 10 y 12 de la noche.

2 Encuentre la media y la desviación estándar de la cantidad de cervezas pedidas entre las 9 PM y la 1 AM.

3 Determine la probabilidad de que el tiempo entre dos pedidos consecutivos está entre 1 y 3 minutos.

Solución **1** La cantidad de cervezas pedida entre 10 y 12 de la noche, se apega a una distribución de Poisson con parámetro $2(30) = 60$. Según la ecuación (7), la probabilidad de que se pidan 60 cervezas entre las 10 y las 12 de la noche es

$$\frac{e^{-60}60^{60}}{60!}$$

Otra manera de hallar la respuesta es con la función de Excel =POISSON(60,60,FALSO). Así se obtiene .051.

2 Resulta que $\lambda = 30$ cervezas por hora; $t = 4$ horas. Por lo tanto, la media de cervezas ordenada entre las 9 PM y la 1 AM es $4(30) = 120$ cervezas. La desviación estándar de la cantidad de cervezas ordenadas entre 10 PM y 1 AM es $(120)^{1/2} = 10.95$.

3 Sea X el tiempo (en minutos) entre los pedidos sucesivos de cerveza. El número promedio de pedidos por minuto es exponencial con parámetro o razón $\frac{30}{60} = 0.5$ cerveza por minuto. Por lo tanto, la función de densidad de la probabilidad del tiempo que transcurre entre pedidos de cervezas $0.5e^{-0.5t}$. Entonces

$$P(1 \leq \mathrm{X} \leq 3) = \int_1^3 (0.5e^{-0.5t})dt = e^{-0.5} - e^{-1.5} = .38$$

También se puede usar Excel para determinar la respuesta con la fórmula EXPONDIST(3,0.5TRUE−EXPONDIST(1,0.5TRUE)

=DISTR.EXP(3,0.5, VERDADERO)−DISTR.EXP(1,0.5, VERDADERO)

Con esto se obtiene una probabilidad de .383.

Distribución de Erlang

Si los tiempos entre llegadas no parecen ser exponenciales, entonces se modelan, con frecuencia, con la distribución Erlang. Una distribución Erlang es una variable aleatoria continua (llámela T) cuya función de densidad $f(t)$ se especifica mediante dos parámetros: un parámetro de proporcionalidad R y un parámetro de forma k (k debe ser un entero positivo). Dados los valores de R y k, la densidad Erlang tiene la función de densidad de probabilidad siguiente:

$$f(t) = \frac{R(Rt)^{k-1}e^{-Rt}}{(k-1)!} \qquad (t \geq 0) \tag{8}$$

Si se aplica la integración por partes, es posible demostrar que si T es una distribución Erlang con parámetro de proporcionalidad R y parámetro de forma k, entonces

$$E(\mathrm{T}) = \frac{k}{R} \qquad \text{y} \qquad \mathrm{var}\ \mathrm{T} = \frac{k}{R^2} \tag{9}$$

Con el objeto de ver la manera en que al modificar el parámetro de forma cambia la forma de la distribución Erlang, considere, para un valor dado de λ, una familia de distribuciones Erlang con parámetro de proporcionalidad $k\lambda$ y parámetro de forma k. Según (9), cada una de estas Erlangs tiene una media de $\frac{1}{\lambda}$. A medida que k varía, la distribución Erlang adopta varias formas. Por ejemplo, en la figura 4 se puede ver que para un valor dado de λ, las funciones de densidad de las distribuciones Erlang tienen parámetros de forma 1, 2, 4, 6 y 20. Para $k = 1$, la densidad Erlang es similar a una distribución exponencial; de hecho, si se fija $k = 1$ en (8), se encuentra que para $k = 1$, la distribución Erlang es una distribución exponencial con parámetro R. A medida que k aumenta, la distribución Erlang se comporta cada vez más y más como una distribución normal. Para valores extremadamente grandes de k, la distribución Erlang se aproxima a una variable aleatoria con varianza cero (es decir, un tiempo constante entre llegadas). Por lo tanto, al variar k nos podríamos aproximar tanto a la distribución sesgada como a la simétrica.

Se puede demostrar que una distribución Erlang con parámetro de forma k y parámetro de proporcionalidad $k\lambda$ tiene la misma distribución que la variable aleatoria $\mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_k$, donde cada $\mathbf{A}_i$ es una variable aleatoria exponencial con parámetro $k\lambda$, y las $\mathbf{A}_i$ son variables aleatorias independientes.

Si modelamos los tiempos entre llegadas como una distribución Erlang con parámetro de forma k, lo que realmente decimos es que el proceso entre llegadas equivale a que un cliente pase por k fases (cada una de las cuales posee la propiedad de carencia de memo-

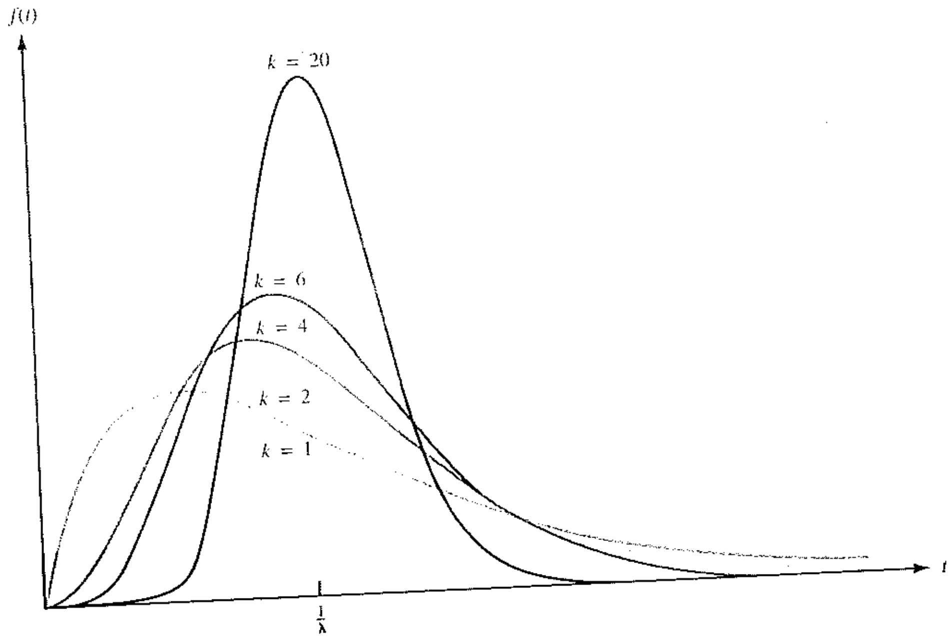


FIGURA 4
Funciones de densidad para las distribuciones Erlang

ria) antes de llegar. Por esta razón, el parámetro de forma se conoce con frecuencia como el *número de fases* de la distribución Erlang.

Modelado del proceso de servicio

Dirija su atención al modelado del proceso de servicio. Suponga que los tiempos de servicio de clientes distintos son variables aleatorias independientes, y que cada tiempo de servicio para cada uno de los clientes está regido por una variable aleatoria **S** cuya función de densidad es $s(t)$. Sea $\frac{1}{\mu}$ el tiempo de servicio medio de un cliente. Naturalmente,

$$\frac{1}{\mu} = \int_0^\infty ts(t)dt$$

Las unidades de la variable $\frac{1}{\mu}$ son horas por cliente, de modo que las unidades de μ son clientes por hora. Por esta razón, a μ se le llama tasa de servicio. Por ejemplo, $\mu = 5$ significa que si siempre hubiera clientes presentes, el servidor podría atender a un promedio de cinco clientes por hora, y el tiempo promedio de servicio a cada cliente sería $\frac{1}{5}$ hora. Al igual que en los tiempos entre llegadas, esperamos que sea posible modelar con precisión los tiempos de servicio como variables aleatorias exponenciales. Si es posible modelar el tiempo de servicio de un cliente como una variable aleatoria exponencial, entonces usted podrá determinar la distribución del tiempo de servicio restante de un cliente sin tener que rastrear cuánto tiempo estuvo en servicio el cliente. Observe también que si los tiempos de servicio siguen una densidad exponencial $s(t) = \mu e^{-\mu t}$, entonces el tiempo de servicio medio de un cliente será $\frac{1}{\mu}$.

Con el objeto de ejemplificar el modo como la suposición de los tiempos de servicio exponenciales simplifica los cálculos, considere un sistema de tres servidores en el cual cada tiempo de servicio del cliente está regido por una distribución exponencial $s(t) = \mu e^{-\mu t}$. Suponga que los tres servidores están ocupados, y que un cliente está esperando (véase figura 5). ¿Cuál es la probabilidad de que el cliente que está esperando sea el último de los cuatro clientes para completar el servicio? Según la figura 5, es evidente que ocurrirá lo siguiente: Uno de los clientes 1 a 3 (digamos, el cliente 3) será el primero en completar el servicio. Luego el cliente 4 entrará al servicio. Por la propiedad de carencia de memoria, el tiempo de servicio del cliente 4 tiene la misma distribución que los tiempos de servicio restantes de los clientes 1 y 2. Por lo tanto, por simetría, los clientes 4, 1 y 2 tendrán la misma oportunidad de ser el último cliente en completar el servicio. Esto quiere decir que el cliente 4 tiene $\frac{1}{3}$ de probabilidad de ser el último en completar el ser-

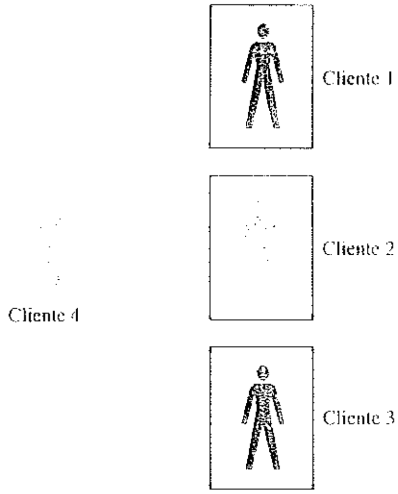


FIGURA 5
Ejemplo de la utilidad de la distribución exponencial

vicio. Sin la propiedad de carencia de memoria, sería difícil de resolver este problema porque habría dificultades para determinar la distribución de probabilidad del tiempo de servicio restante (después de que el cliente 3 completa el servicio) de los clientes 1 y 2.

Infortunadamente, los tiempos de servicio reales podrían no ser consistentes con la propiedad de carencia de memoria. Por esta razón suponemos, con frecuencia, que $s(t)$ es una distribución Erlang con parámetro de forma k y parámetro de proporcionalidad $k\mu$. De acuerdo con (9), tenemos un tiempo de servicio medio de $\frac{1}{\mu}$. Modelar los tiempos de servicio como una distribución Erlang con parámetro de forma k también implica que el tiempo de servicio de un cliente se podría considerar como un pasaje por k fases de servicio, en el cual el tiempo para terminar cada fase tiene la propiedad de carencia de memoria y una media de $\frac{1}{k\mu}$ (véase figura 6). En muchas situaciones, la distribución Erlang se puede ajustar estrechamente a los tiempos de servicio observados.

A veces los tiempos entre llegadas o de servicio se pueden modelar como si tuvieran varianza cero; en este caso, los tiempos entre llegadas o de servicio se consideran como **deterministas**. Por ejemplo, si los tiempos entre llegadas son deterministas, entonces cada tiempo entre llegadas será exactamente $\frac{1}{\lambda}$, y si los tiempos de servicio son deterministas, el tiempo de servicio de cada cliente será exactamente $\frac{1}{\mu}$.

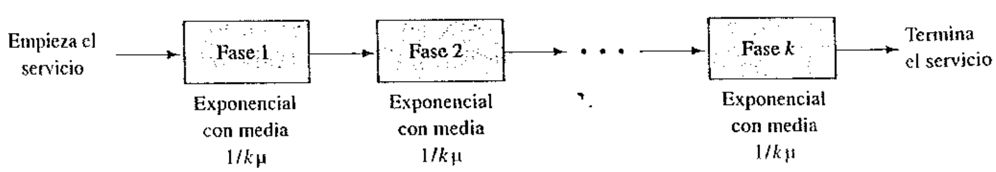
Notación de Kendall-Lee para los sistemas de líneas de espera

Ya se desarrolló terminología suficiente para describir la notación estándar utilizada con el fin de representar muchos sistemas de líneas de espera. La notación que se estudia en esta sección sirve para caracterizar un sistema de líneas de espera en el cual todas las llegadas esperan en una sola cola hasta que está libre uno de los s servidores paralelos idénticos. Luego el primer cliente en la cola entra al servicio, y así sucesivamente (véase figura 7). Por ejemplo, si el cliente en el servidor 3 es el siguiente para completar el servicio, entonces (si se supone una disciplina FCFS) el primer cliente en la cola entraría al servidor 3. El cliente siguiente en la cola entraría al servicio luego de finalizar el servicio siguiente, etcétera.

Kendall (1951) diseñó la notación siguiente para representar dicho sistema de líneas de espera. Cada sistema de líneas de espera se describe mediante seis características:

1/2/3/4/5/6

FIGURA 6
Representación del tiempo de servicio de Erlang



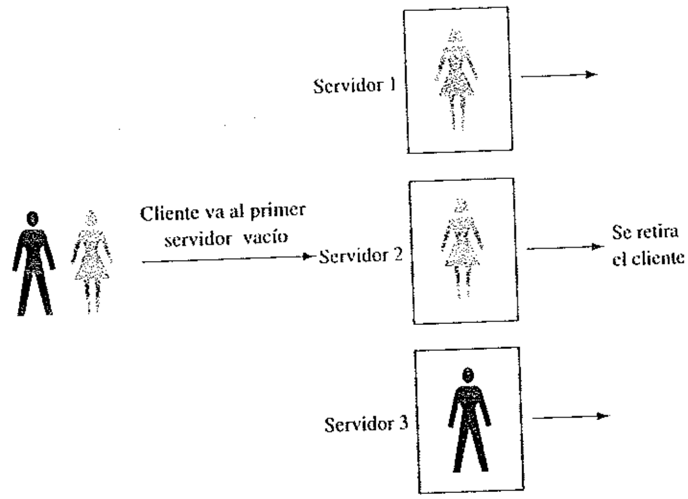


FIGURA 7
Sistema de una sola línea de espera con servidores paralelos

La primera característica especifica la naturaleza del proceso de llegada. Se utilizan las abreviaturas estándar siguientes:

M = Los tiempos entre llegadas son variables aleatorias independientes e idénticamente distribuidas (iid) cuya distribución es exponencial.
D = Los tiempos entre llegadas son iid y deterministas.
E_k = Los tiempos entre llegadas son Erlangs iid con parámetro de forma k.
GI = Los tiempos entre llegadas son iid y están regidos por alguna distribución general.

La segunda característica especifica la naturaleza de los tiempos de servicio:

M = Los tiempos de servicio son iid y están distribuidos exponencialmente.
D = Los tiempos de servicio son iid y deterministas.
E_k = Los tiempos de servicio son Erlangs iid con parámetro de forma k.
G = Los tiempos de servicio son iid y están regidos por alguna distribución general.

La tercera característica es la cantidad de servidores en paralelo. La cuarta característica es la disciplina de las líneas de espera:

FCFS = El primero en llegar, primero en ser atendido
LCFS = El último en entrar, primero en salir
SIRO = Servicio en orden aleatorio
GD = Disciplina general de líneas de espera

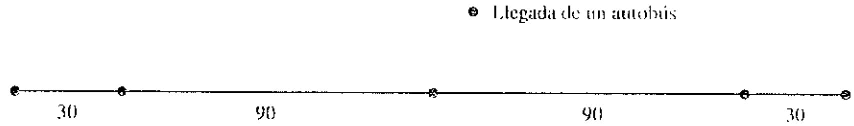
La quinta característica especifica el número máximo admisible de clientes en el sistema (incluidos los clientes que están esperando y los que están en servicio). La sexta característica da el tamaño de la población de donde se extraen los clientes. A menos que la cantidad de clientes potenciales sea del mismo orden de magnitud que el número de servidores, la población se considera infinita. En muchos modelos importantes 4/5/6 es $GD/\infty/\infty$. Si así sucede, entonces 4/5/6 se omite, a menudo.

Como ejemplo de esta notación, M/E_2/8/FCFS/10/∞ podría representar una clínica con ocho médicos, tiempos entre llegadas exponenciales, tiempos de servicio de Erlang de dos fases, una disciplina de líneas de espera FCFS y una capacidad total de 10 pacientes.

Paradoja del tiempo de espera

Esta sección finaliza con un breve análisis de una paradoja interesante conocida como paradoja del tiempo de espera.

FIGURA 8
Paradoja del tiempo de espera



Suponga que el tiempo entre la llegada de autobuses al centro de estudiantes está distribuida en forma exponencial, con una media de 60 minutos. Si arribamos al centro de estudiantes en un instante elegido al azar, ¿cuál es el tiempo promedio que tendremos que esperar el autobús?

De la propiedad de carencia de memoria de la distribución exponencial, se infiere que no importa cuánto tiempo pasó desde que el último autobús llegó, esperaríamos un promedio de 60 minutos hasta que el siguiente autobús llegue. Esta respuesta es, en realidad, correcta, pero parece contradecirse con el razonamiento siguiente. En promedio, alguien que arriba en un tiempo aleatorio debe hacerlo a la mitad de un intervalo representativo entre llegadas de autobuses sucesivos. Si llegamos en el punto medio de un intervalo representativo, y el tiempo promedio entre autobuses es 60 minutos, entonces tendremos que esperar, en promedio $(\frac{1}{2})60 = 30$ minutos, a que llegue el autobús siguiente. ¿Por qué este razonamiento es incorrecto? Simplemente porque el intervalo representativo entre autobuses es más de 60 minutos. La razón de esta anomalía es que lo más probable es arribar a un intervalo más amplio que a uno más corto. Simplifiquemos la situación suponiendo que la mitad de todos los autobuses van separados 30 minutos, y que la mitad de todos los autobuses van 90 minutos separados. Uno podría pensar que como el tiempo promedio entre autobuses es 60 minutos, la espera promedio para que llegue un autobús sería $(\frac{1}{2})60 = 30$ minutos, pero esto es incorrecto. Examine una secuencia representativa de los tiempos entre llegadas del autobús (véase figura 8). La mitad de los tiempos entre llegadas son de 30 minutos, y la mitad son de 90 minutos. Evidentemente, hay $\frac{90}{30+90} = \frac{3}{4}$ de probabilidad de que un autobús llegará durante un tiempo entre llegadas de 90 minutos, y una probabilidad de $\frac{30}{30+90} = \frac{1}{4}$ de que llegue uno en un tiempo entre llegadas de 30 minutos. Por lo tanto, la dimensión promedio del tiempo entre llegadas en el cual un cliente llega es $(\frac{3}{4})(90) + (\frac{1}{4})(30) = 75$ minutos. Como llegamos, en promedio, a la mitad de un tiempo entre llegadas, la espera promedio es $(\frac{3}{4})(\frac{1}{2})90 + (\frac{1}{4})(\frac{1}{2})30 = 37.5$ minutos, la cual es mayor de 30 minutos.

Regresemos al caso donde los tiempos entre llegadas son exponenciales con media de 60 minutos, entonces el tamaño promedio de un tiempo entre llegadas representativo es de 120 minutos. Por lo tanto, el tiempo promedio que tendremos que esperar un autobús es $(\frac{1}{2})(120) = 60$ minutos. Observe que si los autobuses llegan *siempre* cada 60 minutos, entonces el tiempo promedio que una persona tendría que esperar un autobús sería $(\frac{1}{2})(60) = 30$ minutos. En general, se puede demostrar que si A es la variable aleatoria para el tiempo entre autobuses, entonces el tiempo promedio hasta el autobús siguiente (según se ve por una llegada que es igualmente probable que ocurra en cualquier momento) se obtiene con

$$\frac{1}{2}\left(E(\mathrm{A}) + \frac{\mathrm{var\ A}}{E(\mathrm{A})}\right)$$

Por lo que se refiere al ejemplo de los autobuses, $\lambda = \frac{1}{60}$, de modo que las ecuaciones (3) y (4) muestran que $E(\mathrm{A}) = 60$ minutos y var A = 3 600 minutos. Al sustituir en la fórmula se tiene,

$$\text{Tiempo de espera previsto} = \tfrac{1}{2}\left(60 + \tfrac{3\,600}{60}\right) = 60 \text{ min}$$

PROBLEMAS

Grupo A

1 Suponga que llego a un sistema de líneas de espera $M/M/7$/FCFS/8/∞ cuando todos los servidores están ocupados. ¿Cuál es la probabilidad de que complete el servicio antes de por lo menos uno de los siete clientes que están en servicio?

TABLA 2

Tiempo entre autobuses	Probabilidad
30 min	$\frac{1}{4}$
1 h	$\frac{1}{4}$
2 h	$\frac{1}{2}$

2 El tiempo entre autobuses sigue la función de masa de la tabla 2. ¿Cuál es el tiempo promedio que uno debe esperar un autobús?

3 Hay cuatro secciones del tercer grado en la escuela primaria Jefferson. Los alumnos que hay en cada sección se señalan enseguida: sección 1, 20 estudiantes; sección 2, 25 estudiantes; sección 3, 35 estudiantes; sección 4, 40 estudiantes. ¿Cuál es el tamaño promedio de una sección del tercer grado? Suponga que el consejo de educación selecciona en forma aleatoria una alumna de tercer grado de la escuela Jefferson. En promedio, ¿cuántos estudiantes estarán en esa clase?

4 El tiempo entre llegadas de autobuses sigue una distribución exponencial con una media de 60 minutos.

a ¿Cuál es la probabilidad de que exactamente cuatro autobuses lleguen durante las 2 horas siguientes?

b ¿Y de que por lo menos dos autobuses arriben durante las 2 horas siguientes?

c ¿Y de que no llegue ningún autobús en las 2 horas siguientes?

d Acaba de llegar un autobús. ¿Cuál es la probabilidad de que sea 30 o 90 minutos antes de que llegue el siguiente autobús?

5 Durante el año 2000 hubo un promedio de .022 accidentes automovilísticos por persona en Estados Unidos. Aplique lo que sabe de la variable aleatoria tipo Poisson, y explique la verdad del enunciado: "La mayoría de los conductores es mejor que el promedio".

6 Suponga que es igualmente probable que un vuelo de avión esté 50, 60, 70, 80 o 90% lleno.

a ¿Qué fracción de asientos está llena en un vuelo representativo? A esto se le conoce como el *factor de carga del vuelo*.

b Siempre nos quejamos que nunca hay asientos en nuestros vuelos. Dada la información anterior, ¿cuál es el factor de carga promedio en un viaje por avión que yo tome?

7 Llega un promedio de 12 trabajos por hora a nuestra impresora del departamento.

a Determine la probabilidad de que ningún trabajo llegue durante los siguientes 15 minutos mediante dos cálculos distintos (uno que utilice la variable de Poisson y otro la variable aleatoria exponencial).

b ¿Cuál es la probabilidad de que cinco o menos trabajos lleguen durante los siguientes 30 minutos?

20.3 Procesos de nacimiento y muerte

La idea importante del proceso de nacimiento y muerte se analiza en esta sección. Posteriormente se usan los procesos para contestar preguntas relacionadas con diferentes tipos de sistemas de líneas de espera.

Definimos el número de personas presentes en un sistema de líneas de espera en el tiempo t como el **estado** del sistema de líneas de espera en el tiempo t. Para $t = 0$, el estado del sistema es igual al número de personas que están inicialmente en el sistema. La cantidad $P_{ij}(t)$ es de gran interés para este análisis, y se define como la probabilidad de que j personas estén presentes en el sistema de líneas de espera en el tiempo t, dado que en el tiempo 0 están presentes i personas. Observe que $P_{ij}(t)$ es análoga a la probabilidad de transición de n pasos $P_{ij}(n)$ (la probabilidad de que después de n transiciones, una cadena de Markov esté en el estado j, dado que la cadena empezó en el estado i), tratada en el capítulo 17. Recuerde que para la mayor parte de las cadenas de Markov, la $P_{ij}(n)$ se aproxima al límite π_j, el cual es independiente del estado inicial i. De igual manera, resulta que para muchos sistemas de líneas de espera, $P_{ij}(t)$ se aproximará al límite π_j para t grandes; π_j es independiente del estado inicial i. Denominamos a π_j como **estado estable**, o probabilidad de equilibrio, del estado j.

Por lo que toca a los sistemas de líneas de espera que trataremos, se podría pensar que π_j es la probabilidad de que j clientes estén presentes en un instante en el futuro distante. Otra posibilidad es que podríamos considerar a π_j (para un tiempo en el futuro distante) como la fracción del tiempo en que j clientes están presentes. En la mayor parte de los sistemas de líneas de espera, el valor de $P_{ij}(t)$ para t pequeñas dependerá en forma decisiva de i, la cantidad de clientes presentes al inicio. Por ejemplo, si t es pequeña, entonces esperaríamos que $P_{50,1}(t)$ y $P_{1,1}(t)$ fueran muy distintas. No obstante, si existen las probabilidades de estado estable, entonces para una t grande, tanto $P_{50,1}(t)$ como $P_{1,1}(t)$ estarán cercanas a π_1. Es difícil de contestar cuán grande debe ser t antes de alcanzar en forma aproximada el estado estable. El comportamiento de $P_{ij}(t)$ antes de alcanzar el estado estable se llama **comportamiento transitorio** del sistema de líneas de espera. El análisis del comportamiento transitorio del sistema se trata en la sección 20.16. Por ahora, al analizar el comportamiento de un sistema de líneas de espera, supondremos que el estado estable ya se alcanzó. De esta manera podemos trabajar con π_j y no con $P_{ij}(t)$.

Se trata enseguida una cierta clase de procesos estocásticos de tiempo continuo, denominados procesos de nacimiento-muerte, los cuales comprenden muchos sistemas interesantes de sistemas de líneas de espera. Por lo que se refiere a un proceso de nacimiento-muerte, es fácil determinar las probabilidades de estado estable (si acaso existen).

Un **proceso de nacimiento-muerte** es un proceso estocástico de tiempo continuo para el cual el estado del sistema, en cualquier momento, es un entero no negativo (véase una definición de proceso estocástico de tiempo continuo en la sección 17.1). Si un proceso de nacimiento-muerte está en el estado j en el tiempo t, entonces el movimiento del proceso se rige por las leyes siguientes.

Leyes de movimiento para los procesos de nacimiento-muerte

Ley 1 Un nacimiento ocurre entre el tiempo t y el tiempo $t + \Delta t$ con probabilidad $\lambda_j \Delta t + o(\Delta t)$.[†] El estado del sistema incrementa en 1, hasta $j + 1$, con un nacimiento. La variable λ_j se denomina **tasa de nacimientos** en el estado j. En la mayor parte de los sistemas de líneas de espera un nacimiento es simplemente una llegada.

Ley 2 Una muerte ocurre entre el tiempo t y el tiempo $t + \Delta t$ con probabilidad $\mu_j \Delta t + o(\Delta t)$. El estado del sistema disminuye en 1, hasta $j - 1$, con una muerte. La variable μ_j se denomina **tasa de mortalidad** en el estado j. En la mayor parte de los sistemas de líneas de espera una muerte es la finalización de un servicio. Obsérvese que debe cumplirse $\mu_0 = 0$ o podría presentarse un estado negativo.

Ley 3 Los nacimientos y las muertes son independientes entre sí.

Las leyes 1 a 3 se pueden aplicar para demostrar que la probabilidad de que más de un evento suceda (nacimiento o muerte) entre t y $t + \Delta t$ es $o(\Delta t)$. Nótese que cualquier proceso de nacimiento-muerte está completamente especificado al conocerse las tasas de nacimiento λ_j y las tasas de mortalidad μ_j. Como no puede haber un estado negativo, cualquier proceso de nacimiento-muerte debe tener $\mu_0 = 0$.

Relación de la distribución exponencial con los procesos de nacimiento-muerte

La mayor parte de los sistemas de líneas de espera con tiempos entre llegadas exponenciales y tiempos de servicio exponenciales se podrían modelar como si fueran procesos de nacimiento-muerte. Para ejemplificar por qué es así, considere un sistema de colas $M/M/1$/FCFS/∞/∞, en el cual los tiempos entre llegadas son exponenciales con parámetro λ y los tiempos de servicio están distribuidos en forma exponencial y su parámetro es μ. Si el estado (cantidad de personas presentes) en el tiempo t es j, entonces de la propiedad de carencia de memoria de la distribución exponencial se infiere que la probabilidad de un nacimiento durante el tiempo $[t, t + \Delta t]$ no depende de cuánto tiempo ha permanecido el sistema en el estado j. Esto quiere decir que la probabilidad de que se presente un nacimiento durante $[t, t + \Delta t]$ no depende cuánto tiempo el sistema ha permanecido en el estado j y, por lo tanto, se podría determinar como si una llegada se hubiera producido justamente en el tiempo t. Entonces, la probabilidad de que un nacimiento ocurra durante $[t, t + \Delta t]$ es

$$\int_0^{\Delta t} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda \Delta t}$$

Según la expansión de las series de Taylor dada en la sección 11.1,

$$e^{-\lambda \Delta t} = 1 - \lambda \Delta t + o(\Delta t)$$

[†]Refiérase a la sección 20.2; ahí se establece que $o(\Delta t)$ significa que $\lim_{\Delta t \to 0} \dfrac{o(\Delta t)}{\Delta t} = 0$.

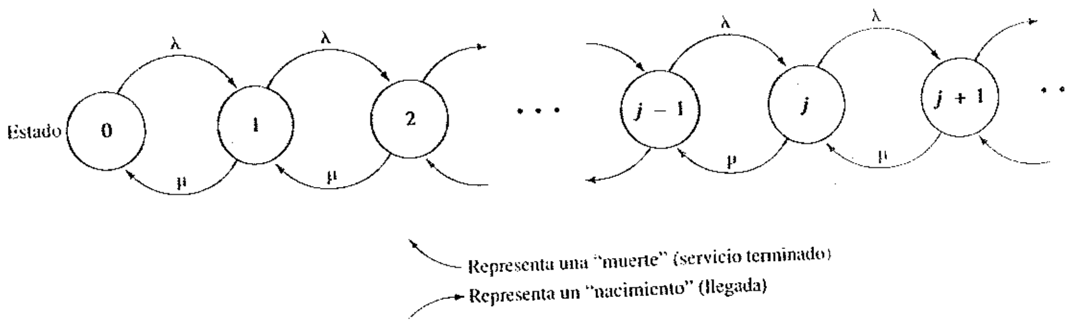


FIGURA 9
Diagrama de tasas para un sistema de colas *M/M/*1/FCFS/∞/∞

Esto quiere decir que la probabilidad de que se presente un nacimiento durante $[t, t + \Delta t]$ es $\lambda\Delta t + o(\Delta t)$. A partir de todo lo anterior podríamos concluir que la tasa de nacimiento en el estado j es simplemente la tasa de llegadas λ.

Con el fin de determinar la tasa de nacimientos en el tiempo t, obsérvese que si el estado es cero en el tiempo t, entonces nadie está en servicio, de modo que no puede haber finalización de ningún servicio entre el tiempo t y $t + \Delta t$. Por lo tanto, $\mu_0 = 0$. Si el estado en el tiempo t es $j \geq 1$, entonces sabemos que (puesto que hay sólo un servidor) exactamente un cliente estará en servicio. Por lo tanto, según la propiedad de carencia de memoria de la distribución exponencial, la probabilidad de que un cliente finalice el servicio entre t y $t + \Delta t$ es

$$\int_0^{\Delta t} \mu e^{-\mu t} dt = 1 - e^{-\mu \Delta t} = \mu\Delta t + o(\Delta t)$$

Por lo tanto, para $j \geq 1$, $\mu_j = \mu$. En resumen, si suponemos que las finalizaciones del servicio y las llegadas se presentan en forma independiente, entonces, el sistema de colas *M/M/*1/FCFS/∞/∞ es un proceso de nacimiento-muerte. Las tasas de nacimiento y de mortalidad para el sistema de colas *M/M/*1/FCFS/∞/∞ se podría representar mediante un diagrama de tasas como el de la figura 9.

Los sistemas de colas más complicados con tiempos entre llegadas exponenciales y tiempos de servicio exponenciales se podrían modelar como procesos de nacimiento-muerte si se añaden tasas de servicio para servidores ocupados y se suman tasas de llegada para corrientes de llegadas distintas. Por ejemplo, considere un sistema de colas *M/M/*3/FCFS/∞/∞ en el cual los tiempos entre llegadas son exponenciales con $\lambda = 4$ y los tiempos de servicio son exponenciales con $\mu = 5$. Para modelar este sistema como un proceso de nacimiento-muerte, podríamos usar los parámetros siguientes (véase figura 10):

$$\lambda_j = 4 \qquad (j = 0, 1, 2, \ldots)$$
$$\mu_0 = 0, \quad \mu_1 = 5, \quad \mu_2 = 5 + 5 = 10, \quad \mu_j = 5 + 5 + 5 = 15 \qquad (j = 3, 4, 5, \ldots)$$

Si los tiempos entre llegadas o los tiempos de servicio no son exponenciales, entonces no es apropiado el modelo del proceso nacimiento-muerte.† Suponga, por ejemplo, que los tiempos de servicio no son exponenciales y que estamos considerando un sistema de colas *M/G/*1/FCFS/∞/∞. Como los tiempos de servicio para un sistema *M/G/*1/FCFS/∞/∞ podrían ser no exponenciales, la probabilidad de que ocurra una muerte (final del servicio) entre t y $t + \Delta t$ depende del tiempo desde que se completó el último servicio. Esto viola la ley 2, de modo que no podemos modelar un sistema *M/G/*1/FCFS/∞/∞ como un proceso de nacimiento-muerte.

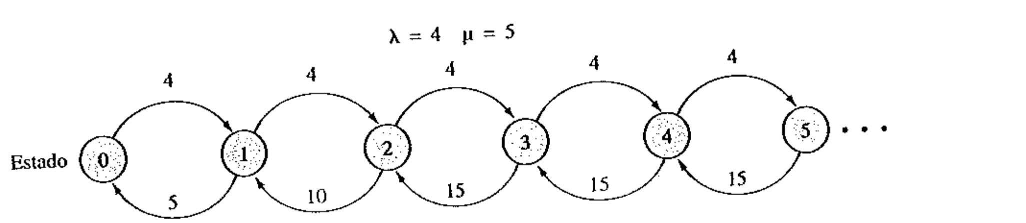


FIGURA 10
Diagrama de tasas para el sistema de colas *M/M/*3/FCFS/∞/∞

†Es posible desarrollar un modelo modificado de nacimiento-muerte si el tiempo de servicio y los tiempos entre llegadas son distribuciones Erlang.

Derivación de las probabilidades de estado estable para procesos de nacimiento-muerte

Ahora mostramos cómo las π_j se podrían determinar para un proceso arbitrario de nacimiento-muerte. La clave es relacionar (para Δt pequeñas) $P_{ij}(t + \Delta t)$ con $P_{ij}(t)$. La manera de hacerlo es observar que hay cuatro modos para el estado, en el tiempo $t + \Delta t$, sea j. Para $j \geq 1$, los cuatro modos se señalan en la tabla 3. Por lo que se refiere a $j \geq 1$, la probabilidad de que el estado del sistema sea $j - 1$ en el tiempo t y j en el tiempo $t +$ t es (véase figura 11)

$$P_{i,j-1}(t)(\lambda_{j-1}\Delta t + o(\Delta t))$$

Con argumentos similares se obtiene (II) y (III). Se infiere (IV) porque si el sistema está en un estado distinto a j, $j - 1$ o $j + 1$ en el tiempo t, entonces para concluir en el estado j en el tiempo $t + \Delta t$, más de un evento (nacimiento o muerte) debe ocurrir entre t y $t + \Delta t$. Por lo que toca a la ley 3, su probabilidad es $o(\Delta t)$. Por lo tanto,

$$P_{ij}(t + \Delta t) = (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}) + (\mathrm{IV})$$

Después de reagrupar los términos de esta ecuación se tiene

$$\begin{aligned} P_{ij}(t + \Delta t) = P_{ij}(t) & \\ & + \Delta t(\lambda_{j-1}P_{i,j-1}(t) + \mu_{j+1}P_{i,j+1}(t) - P_{ij}(t)\mu_j - P_{ij}(t)\lambda_j) \\ & + \underline{o(\Delta t)(P_{i,j-1}(t) + P_{i,j+1}(t) + 1 - 2P_{ij}(t))} \end{aligned} \tag{10}$$

Puesto que el término subrayado se podría expresar como $o(\Delta t)$, se vuelve a escribir (10) como

$$P_{ij}(t + \Delta t) - P_{ij}(t) = \Delta t(\lambda_{j-1}P_{i,j-1}(t) + \mu_{j+1}P_{i,j+1}(t) - P_{ij}(t)\mu_j - P_{ij}(t)\lambda_j) + o(\Delta t)$$

Luego de dividir ambos miembros de esta ecuación entre Δt y de hacer que Δt se aproxime a cero se observa que para toda i y $j \geq 1$,

$$P'_{ij}(t) = \lambda_{j-1}P_{i,j-1}(t) + \mu_{j+1}P_{i,j+1}(t) - P_{ij}(t)\mu_j - P_{ij}(t)\lambda_j \tag{10'}$$

Ya que para $j = 0$, $P_{i,j-1}(t) = 0$ y $\mu_j = 0$, se tiene, para $j = 0$,

$$P'_{i,0}(t) = \mu_1 P_{i,1}(t) - \lambda_0 P_{i,0}(t)$$

Esto es un sistema infinito de ecuaciones diferenciales. (Una ecuación diferencial es simplemente una ecuación en la que aparece una derivada.) Desde el punto de vista teórico, estas ecuaciones se podrían resolver para $P_{ij}(t)$. Pero la realidad es que este sistema de ecuaciones es extremadamente difícil de resolver. No todo está perdido. Se puede usar (10′) para obtener las probabilidades de estado estable π_j ($j = 0, 1, 2, \ldots$). Al igual que con las cadenas de Markov, la probabilidad de estado estable π_j se define como

$$\lim_{t \to \infty} P_{ij}(t)$$

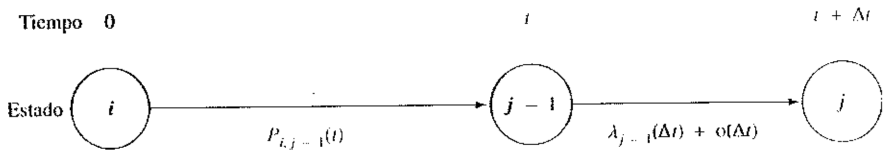
Entonces, para t grande y cualquier estado inicial i, $P_{ij}(t)$ no cambia mucho y se podría considerar como una constante. Por lo tanto, en el estado estable (t grande), $P'_{ij}(t) = 0$.

TABLA 3

Cálculos de que la probabilidad que el estado en el tiempo $t + \Delta t$ sea j

Estado en el tiempo t	Estado en el tiempo $t + \Delta t$	Probabilidad de esta secuencia de eventos
$j - 1$	j	$P_{i,j-1}(t)\,(\lambda_{j-1}\Delta t + o(\Delta t)) = (\mathrm{I})$
$j + 1$	j	$P_{i,j+1}(t)\,(\mu_{j+1}\Delta t + o(\Delta t)) = (\mathrm{II})$
j	j	$P_{i,j}(t)\,(1 - \mu_j\,\Delta t - \lambda_j\Delta t - 2o(\Delta t)) = (\mathrm{III})$
Cualquier otro estado	j	$o(\Delta t) = (\mathrm{IV})$

FIGURA 11
Probabilidad de que el estado sea $j-1$ en el tiempo t y j en el tiempo $t+\Delta t$ es $P_{i,j-1}(t)(\lambda_{j-1}(\Delta t)+o(\Delta t))$



En el estado estable, se conservarán, asimismo, $P_{i,j-1}(t) = \pi_{j-1}$, $P_{i,j+1}(t) = \pi_{j+1}$ y $P_{ij}(t) = \pi_j$. Cuando se sustituyen estas relaciones en (10′), se tiene, para $j \geq 1$,

$$\lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1} - \pi_j\mu_j - \pi_j\lambda_j = 0 \qquad (10'')$$

$$\lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1} = \pi_j(\lambda_j + \mu_j) \qquad (j = 1, 2, \ldots)$$

Para $j = 0$, se tiene

$$\mu_1\pi_1 = \pi_0\lambda_0$$

Las ecuaciones (10″) son un sistema infinito de ecuaciones lineales con las que se puede encontrar con facilidad las π_j. Antes de explicar cómo resolver (10″), ofrecemos una deducción intuitiva de (10″) con base en la observación siguiente: *en cualquier tiempo t que observemos un proceso nacimiento-muerte, debe ser cierto que, para cada estado j, la cantidad de veces que hemos entrado al estado j difiere cuando mucho en una unidad del número de veces que hemos dejado el estado j.*

Suponga que para el tiempo t, hemos entrado tres veces al estado 6. Entonces, debe haber ocurrido uno de los casos de la tabla 4. Por ejemplo, si se presenta el caso 2, empezamos en el estado 6 y concluimos en algún otro estado. Como hemos observado tres transiciones en el estado 6 dentro del tiempo t, deben haber ocurrido los eventos siguientes (entre otros):

Iniciar en el estado 6	Entrar al estado 6 (por segunda vez)
Salir del estado 6 (por primera vez)	Salir del estado 6 (por tercera vez)
Entrar al estado 6 (por primera vez)	Entrar al estado 6 (por tercera vez)
Dejar el estado 6 (por segunda vez)	Salir del estado 6 (por cuarta vez)

Por lo tanto, si se presenta el caso 2, entonces para el tiempo t debemos haber salido cuatro veces del estado 6.

Esta observación sugiere que para t grande y para $j = 0, 1, 2, \ldots$ (y para condiciones iniciales cualesquiera) es cierto que

$$\frac{\text{Número de salidas esperadas desde el estado } j}{\text{Unidad de tiempo}} = \frac{\text{Número de entradas esperadas al estado } j}{\text{Unidad de tiempo}} \qquad (11)$$

Si se supone que el sistema se ha afirmado en el estado estable, sabemos que el sistema pasa una fracción π_j de su tiempo en el estado j. Ahora podemos aplicar (11) para determinar las probabilidades de estado estable π_j. Para $j \geq 1$, sólo podemos salir del estado j si vamos al estado $j+1$ o al estado $j-1$, de modo que para $j \geq 1$, obtenemos

$$\frac{\text{Número esperado de salidas desde el estado } j}{\text{Unidad de tiempo}} = \pi_j(\lambda_j + \mu_j) \qquad (12)$$

Ya que para $j \geq 1$ sólo podemos entrar al estado j desde el estado $j-1$ o desde $j+1$,

$$\frac{\text{Número esperado de entradas al estado } j}{\text{Unidad de tiempo}} = \pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1} \qquad (13)$$

Si se sustituyen (12) y (13) en (11) se tiene

$$\pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1} = \pi_j(\lambda_j + \mu_j) \qquad (j = 1, 2, \ldots) \qquad (14)$$

TABLA 4

Relación entre el número de transiciones que entran y salen de un estado en el tiempo t

Estado inicial	Estado en el tiempo t	Número de transiciones que salen del estado 6 en el tiempo t
Caso 1: estado 6	Estado 6	3
Caso 2: estado 6	Cualquier estado excepto 6	4
Caso 3: cualquier estado excepto el estado 6	Estado 6	2
Caso 4: cualquier estado excepto el estado 6	Cualquier estado excepto 6	3

Para $j = 0$, sabemos que $\mu_0 = \pi_{-1} = 0$, de modo que también tenemos

$$\pi_1\mu_1 = \pi_0\lambda_0 \tag{14'}$$

Las ecuaciones (14) y (14′) se denominan **ecuaciones de balance de flujo** o **ecuaciones de conservación del flujo** para un proceso de nacimiento–muerte. Observe que (14) expresa el hecho de que, en el estado estable, la tasa a la cual se presentan las transiciones en cualquier estado i debe ser igual a la tasa a la cual ocurren las transiciones fuera del estado i. Si (14) no se cumple para todos los estados, entonces la probabilidad se "acumularía" en algún estado, y no existiría el estado estable.

Al escribir las ecuaciones para (14) y (14′) obtenemos las ecuaciones de balance de flujo siguientes para un proceso de nacimiento-muerte:

$$\begin{array}{ll}
(j = 0) & \pi_0\lambda_0 = \pi_1\mu_1 \\
(j = 1) & (\lambda_1 + \mu_1)\pi_1 = \lambda_0\pi_0 + \mu_2\pi_2 \\
(j = 2) & (\lambda_2 + \mu_2)\pi_2 = \lambda_1\pi_1 + \mu_3\pi_3 \\
\vdots & \\
(j\text{-ésima ecuación}) & (\lambda_j + \mu_j)\pi_j = \lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1}
\end{array} \tag{15}$$

Solución de las ecuaciones de balance de flujo en un proceso nacimiento-muerte

Para resolver (15), empezamos por expresar todas las π_j en términos de π_0. A partir de la ecuación ($j = 0$) obtenemos

$$\pi_1 = \frac{\pi_0\lambda_0}{\mu_1}$$

Al sustituir este resultado en la ecuación ($j = 1$) tenemos

$$\lambda_0\pi_0 + \mu_2\pi_2 = \frac{(\lambda_1 + \mu_1)\pi_0\lambda_0}{\mu_1}$$

$$\mu_2\pi_2 = \frac{\pi_0(\lambda_0\lambda_1)}{\mu_1}$$

Por lo tanto,

$$\pi_2 = \frac{\pi_0(\lambda_0\lambda_1)}{\mu_1\mu_2}$$

Podríamos usar ahora la ecuación ($j = 3$) para encontrar π_3 en términos de π_0, y así sucesivamente. Si definimos

$$c_j = \frac{\lambda_0\lambda_1 \cdots \lambda_{j-1}}{\mu_1\mu_2 \cdots \mu_j}$$

entonces es posible demostrar que

$$\pi_j = \pi_0 c_j \tag{16}$$

(Véase problema 1 al final de esta sección). Puesto que en cualquier tiempo dado debemos estar en algún estado, las probabilidades de estado estable deben sumarse a 1.

$$\sum_{j=0}^{j=\infty} \pi_j = 1 \tag{17}$$

Al sustituir (16) en (17) se llega a

$$\pi_0\left(1 + \sum_{j=1}^{j=\infty} c_j\right) = 1 \tag{18}$$

Si $\sum_{j=1}^{j=\infty} c_j$ es finita, podemos aplicar (18) para determinar π_0:

$$\pi_0 = \frac{1}{1 + \sum_{j=1}^{j=\infty} c_j} \tag{19}$$

Luego se usa (16) para encontrar $\pi_1, \pi_2, \ldots$. Es posible demostrar que si $\sum_{j=1}^{j=\infty} c_j$ es infinita, entonces, no existe distribución de estado estable. La razón más común para que no exista el estado estable es que la tasa de llegadas es por lo menos igual a la tasa máxima a la cual los clientes pueden ser atendidos.

Uso de una hoja de cálculo para estimar las probabilidades de estado estable

Mediante el ejemplo siguiente se ilustra el modo en que una hoja de cálculo se usa para determinar las probabilidades de estado estable en un proceso de nacimiento-muerte.

EJEMPLO 2 Indiana Bell

Los representantes del servicio a clientes de Indiana Bell reciben un promedio de 1 700 llamadas por hora. El tiempo entre éstas sigue una distribución exponencial. Un representante del servicio a clientes puede atender un promedio de 30 llamadas por hora. El tiempo necesario para atender una llamada también sigue una distribución exponencial. Indiana Bell puede tener esperando hasta 25 personas. Si 25 personas están en espera, se pierde una llamada para el sistema. Indiana Bell tiene 75 representantes del servicio.

1 ¿En qué fracción del tiempo están todos los operadores ocupados?

2 ¿Qué fracción de todas las llamadas pierde el sistema?

Bell.xls

Solución En la figura 12 (archivo Bell.xls) se presenta una hoja de cálculo para determinar las probabilidades de estado estable para este proceso de nacimiento-muerte. Hagamos el estado i en cualquier tiempo igual al número de personas cuyas llamadas están siendo atendidas o que están en espera. Tenemos que para $i = 0, 1, 2, \ldots, 99$, $\lambda_i = 1\,700$. Del hecho de que se pierden las llamadas recibidas cuando $75 + 25 = 100$ llamadas están en el sistema se infiere que $\lambda_{100} = 0$. Entonces, no puede ocurrir ningún estado $i > 100$ (¿por qué?). Tenemos $u_0 = 0$ y para $i = 1, 2, \ldots, 75$, $\mu_i = 30i$. Para $i > 75$, $\mu_i = 30(75) = 2\,250$.

Para contestar las partes (1) y (2), necesitamos calcular las probabilidades de estado estable π_i = fracción del tiempo en que el estado es i. Escribimos los posibles estados del sistema (0–100) en las celdas A4:A104. Para hacerlo, escribimos 0 en la celda A4 y 1 en A5. Luego seleccionamos el intervalo A4:A5 y arrastramos el cursor hasta A6:A104. Se escribe la tasa de llegadas de 1 700 en B4, y lleve el curso hacia abajo a B5:B104 pa-

	A	B	C	D	E	F
1						Prob(i>=75)
2	INDIANA	BELL	EXAMPLE		0	.012759326
3	STATE	LAMBDA	MU	CJ	PROB	
4	0	1700	0	1	2.451E-25	
5	1	1700	30	56.6666667	1.3889E-23	
6	2	1700	60	1605.55556	3.9352E-22	
7	3	1700	90	30327.1605	7.4332E-21	
8	4	1700	120	429634.774	1.053E-19	
9	5	1700	150	4869194.1	1.1934E-18	
10	6	1700	180	45986833.2	1.1271E-17	
11	7	1700	210	372274364	9.1244E-17	
12	8	1700	240	2636943411	6.4631E-16	
13	9	1700	270	1.6603E+10	4.0694E-15	
14	10	1700	300	9.4084E+10	2.306E-14	
15	11	1700	330	4.8467E+11	1.1879E-13	
16	12	1700	360	2.2887E+12	5.6097E-13	
17	13	1700	390	9.9765E+12	2.4452E-12	
18	14	1700	420	4.0381E+13	9.8974E-12	
19	15	1700	450	1.5255E+14	3.739E-11	
20	16	1700	480	5.4029E+14	1.3242E-10	
21	17	1700	510	1.801E+15	4.4141E-10	
22	18	1700	540	5.6697E+15	1.3896E-09	
23	19	1700	570	1.691E+16	4.1445E-09	
24	20	1700	600	4.791E+16	1.1743E-08	
25	21	1700	630	1.2928E+17	3.1687E-08	
26	22	1700	660	3.33E+17	8.1618E-08	
27	23	1700	690	8.2043E+17	2.0109E-07	
28	24	1700	720	1.9371E+18	4.7479E-07	
29	25	1700	750	4.3908E+18	1.0762E-06	
30	26	1700	780	9.5697E+18	2.3455E-06	
31	27	1700	810	2.0085E+19	4.9227E-06	
32	28	1700	840	4.0648E+19	9.9627E-06	
33	29	1700	870	7.9426E+19	1.9467E-05	
34	30	1700	900	1.5003E+20	3.6772E-05	
35	31	1700	930	2.7424E+20	6.7217E-05	
36	32	1700	960	4.8564E+20	0.00011903	
37	33	1700	990	8.3393E+20	0.00020439	
38	34	1700	1020	1.3899E+21	0.00034066	
39	35	1700	1050	2.2503E+21	0.00055154	
40	36	1700	1080	3.5421E+21	0.00086817	
41	37	1700	1110	5.4248E+21	0.00132962	
42	38	1700	1140	8.0897E+21	0.00198277	
43	39	1700	1170	1.1754E+22	0.00288095	
44	40	1700	1200	1.6652E+22	0.00408134	
45	41	1700	1230	2.3015E+22	0.00564088	
46	42	1700	1260	3.1052E+22	0.00761072	
47	43	1700	1290	4.0921E+22	0.01002963	
48	44	1700	1320	5.2701E+22	0.01291694	
49	45	1700	1350	6.6364E+22	0.01626578	
50	46	1700	1380	8.1753E+22	0.02003755	
51	47	1700	1410	9.8567E+22	0.02415875	
52	48	1700	1440	1.1636E+23	0.02852075	
53	49	1700	1470	1.3457E+23	0.03298318	
54	50	1700	1500	1.5251E+23	0.03738094	
55	51	1700	1530	1.6946E+23	0.04153437	
56	52	1700	1560	1.8467E+23	0.04526182	
57	53	1700	1590	1.9744E+23	0.04839314	
58	54	1700	1620	2.0719E+23	0.05078292	
59	55	1700	1650	2.1347E+23	0.0523218	
60	56	1700	1680	2.1601E+23	0.05294468	

FIGURA 12
Indiana Bell

A	A	B	C	D	E	F
61	57	1700	1710	2.1E+23	0.0526351	
62	58	1700	1740	2.1E+23	0.0514251	
63	59	1700	1770	2.0E+23	0.0493913	
64	60	1700	1800	1.9E+23	0.0466473	
65	61	1700	1830	1.8E+23	0.0433336	
66	62	1700	1860	1.6E+23	0.039606	
67	63	1700	1890	1.5E+23	0.0356244	
68	64	1700	1920	1.3E+23	0.0315425	
69	65	1700	1950	1.1E+23	0.0274985	
70	66	1700	1980	9.6E+22	0.0236099	
71	67	1700	2010	8.1E+22	0.0199685	
72	68	1700	2040	6.8E+22	0.0166405	
73	69	1700	2070	5.6E+22	0.0136661	
74	70	1700	2100	4.5E+22	0.011063	
75	71	1700	2130	3.6E+22	0.0088296	
76	72	1700	2160	2.8E+22	0.0069492	
77	73	1700	2190	2.2E+22	0.0053944	
78	74	1700	2220	1.7E+22	0.0041308	
79	75	1700	2250	1.3E+22	0.0031211	
80	76	1700	2250	9.6E+21	0.0023581	
81	77	1700	2250	7.3E+21	0.0017817	
82	78	1700	2250	5.5E+21	0.0013462	
83	79	1700	2250	4.1E+21	0.0010171	
84	80	1700	2250	3.1E+21	0.0007685	
85	81	1700	2250	2.4E+21	0.0005806	
86	82	1700	2250	1.8E+21	0.0004387	
87	83	1700	2250	1.4E+21	0.0003315	
88	84	1700	2250	1.0E+21	0.0002504	
89	85	1700	2250	7.7E+20	0.0001892	
90	86	1700	2250	5.8E+20	0.000143	
91	87	1700	2250	4.4E+20	0.000108	
92	88	1700	2250	3.3E+20	0.0000816	
93	89	1700	2250	2.5E+20	0.0000617	
94	90	1700	2250	1.9E+20	0.0000466	
95	91	1700	2250	1.4E+20	0.0000352	
96	92	1700	2250	1.1E+20	0.0000266	
97	93	1700	2250	8.2E+19	0.0000201	
98	94	1700	2250	6.2E+19	0.0000152	
99	95	1700	2250	4.7E+19	0.0000115	
100	96	1700	2250	3.5E+19	0.0000087	
101	97	1700	2250	2.7E+19	0.0000065	
102	98	1700	2250	2.0E+19	0.0000049	
103	99	1700	2250	1.5E+19	0.0000037	
104	100	0	2250	1.2E+19	0.0000028	

FIGURA 12
(Continuación)

ra generar las tasas de llegadas para todos los estados. Para producir las tasas de servicio, escriba 0 en C4. Luego escriba 30 en C5 y 60 en la celda C6. Seleccione después el intervalo C5:C6, y arrastre el cursor hacia abajo hasta C79. De este modo se generan las tasas de servicio para los estados 0 a 75. Escriba 2 250 en C80, y arrastre ese resultado hacia abajo hasta C81:C104. Esta operación genera la tasa de servicio (2 250) para los estados 76 a 100. En el intervalo de la celda D4:D104 calculamos las c_j que se requieren para estimar las probabilidades de estado estable. Entonces, para empezar, escribimos 1 en D4. Como $c_1 = \lambda_0/\mu_1$, introducimos =B4/C5 en la celda D5. Luego, como $c_2 = c_1\lambda_1/\mu_2$, escribimos =D5*B5/C6 en D6. Al copiar desde D6 hasta D7:D104 generamos ahora las c_j que faltaban. En E4 calculamos π_0 mediante =SUM(D$4:D$104). En E5 determinamos π_1 al escribir =D5*E$4. Cuando copiamos desde el intervalo E5 hasta el intervalo E5:E104, generamos las probabilidades de estado estable que faltaban. Ahora ya podemos contestar a las preguntas (1) y (2).

1 Determinamos $\pi_{75} + \pi_{76} + \cdots + \pi_{100}$. Para hacerlo, introducimos el comando =SUMA(E79:E104) [=SUM(E79:E104) en el programa en inglés] en la celda F2 y encontramos .013.

2 Una llamada que entra se pierde si el estado es igual a 100. Se pierde una fracción $\pi_{100} = .0000028$ de todas las llegadas. Por lo tanto, la compañía telefónica ¡proporciona un muy buen servicio!

En las secciones 20.4 a 20.6 y 20.9 a 20.10 se aplica la teoría de los procesos de nacimiento-muerte con el objeto de determinar las distribuciones de las probabilidades de estado estable para una variedad de sistemas de líneas de espera. Después se utilizaron las distribuciones de probabilidades de estado estacionario para estimar otras cantidades de interés (como el tiempo de espera esperado y el número esperado de clientes en el sistema).

Los modelos de nacimiento-muerte se utilizan para modelar fenómenos distintos a los sistemas de líneas de espera. Por ejemplo, el número de compañías en una industria que se pueda modelar como un proceso de nacimiento-muerte: el estado de la industria en cualquier tiempo dado es la cantidad de compañías que están en el negocio; un nacimiento corresponde a una firma que entra a la industria; y una muerte corresponde a una compañía que deja el negocio.

PROBLEMAS

Grupo A

1 Demuestre que los valores de π_j dados en (16) realmente satisfacen las ecuaciones de balance de flujo (14) y (14′).

2 En mi casa hay dos luminarias. En promedio, una luminaria o foco dura 22 días (distribución exponencial). Cuando una luminaria se quema, me tardo un promedio de dos días (distribución exponencial) para reemplazarla.

a Formule un modelo de nacimiento-muerte de tres estados de esta situación.

b Determine la fracción de tiempo que ambas luminarias están en funcionamiento.

c Encuentre la fracción de tiempo que ninguna luminaria funciona.

Grupo B

3 Usted está realizando un análisis industrial de la industria de pizzas en Bloomington. La tasa (por año) a la cual las pizzerías ingresan a la industria está dada por p, donde p = precio de una pizza en dólares. Se supone que el precio de una pizza es $\max(0, 16 - .5F)$, donde F = cantidad de pizzerías en Bloomington. Durante un año determinado, la probabilidad de que una pizzería fracase es $1/(10 + p)$. Genere un modelo de nacimiento-muerte de esta situación.

a Estime el promedio de pizzerías en Bloomington en el estado estable.

b ¿En qué fracción de tiempo habrá más de 20 pizzerías en Bloomington?

20.4 Sistema de líneas de espera $M/M/1/GD/\infty/\infty$ y la fórmula de colas $L = \lambda W$

Como ya se explicó, la metodología del nacimiento-muerte en la sección anterior, ahora la utilizaremos para analizar las propiedades del sistema de colas $M/M/1/GD/\infty/\infty$. Recuerde que los tiempos de llegadas del sistema de colas $M/M/1/GD/\infty/\infty$ son exponenciales (suponemos que la tasa de llegadas por unidad de tiempo es λ) y tiene un solo servidor con tiempos de servicio exponenciales (suponemos que cada tiempo de servicio del cliente es exponencial con tasa μ). En la sección 20.3 demostramos que un sistema de líneas de espera $M/M/1/GD/\infty/\infty$ se podría modelar como un proceso de nacimiento-muerte con los parámetros siguientes:

$$\begin{aligned} \lambda_j &= \lambda \qquad (j = 0, 1, 2, \ldots) \\ \mu_0 &= 0 \\ \mu_j &= \mu \qquad (j = 1, 2, 3, \ldots) \end{aligned} \tag{20}$$

Deducción de las probabilidades de estado estable

Es posible utilizar las ecuaciones (15) a (19) para encontrar π_j, la probabilidad de estado estable de que estén presentes j clientes. Cuando se sustituye (20) en (16) se obtiene

$$\pi_1 = \frac{\lambda\pi_0}{\mu}, \qquad \pi_2 = \frac{\lambda^2\pi_0}{\mu^2}, \qquad \ldots, \qquad \pi_j = \frac{\lambda^j\pi_0}{\mu^j} \tag{21}$$

Definamos $\rho = \frac{\lambda}{\mu}$. Por razones que muy pronto serán evidentes, ρ es la **intensidad de tráfico** del sistema de líneas de espera. Al sustituir (21) en (17) tenemos

$$\pi_0(1 + \rho + \rho^2 + \cdots) = 1 \tag{22}$$

Ahora supongamos que $0 \leq \rho < 1$. Entonces evaluamos la suma $S = 1 + \rho + \rho^2 + \cdots$ como sigue: multiplicamos S por ρ, lo cual da $\rho S = \rho + \rho^2 + \rho^3 + \cdots$. Entonces $S - \rho S = 1$, y

$$S = \frac{1}{1 - \rho} \tag{23}$$

Si sustituimos (23) en (22) tenemos

$$\pi_0 = 1 - \rho \qquad (0 \leq \rho < 1) \tag{24}$$

Luego de sustituir (24) en (21) encontramos

$$\pi_j = \rho^j(1 - \rho) \qquad (0 \leq \rho < 1) \tag{25}$$

Ahora bien, si $\rho \geq 1$ la suma infinita en (22) se "amplifica" (intente $\rho = 1$, por ejemplo, y obtendrá $1 + 1 + 1 + \cdots$). Por consiguiente, si $\rho \geq 1$ no existe ninguna distribución de estado estable. Puesto que $\rho = \frac{\lambda}{\mu}$, vemos que si $\lambda \geq \mu$ (es decir, la tasa de llegadas es por lo menos igual a la tasa de servicio), entonces no existe distribución de estado estable.

Si $\rho > 1$, es fácil ver por qué no existe distribución de estado estable. Suponga que $\lambda = 6$ clientes por hora y $\mu = 4$ clientes por hora. Incluso si el servidor estuviera trabajando todo el tiempo, sólo puede atender un promedio de cuatro personas por hora. Por lo tanto, la cantidad promedio de clientes en el sistema aumentaría por lo menos $6 - 4 = 2$ clientes por hora. Esto quiere decir que después de mucho tiempo, la cantidad de clientes presentes se "amplificaría", y ninguna distribución de estado estable podría haber. Si $\rho = 1$, la inexistencia de un estado estable no es tan obvia, pero este análisis sí indica que no existe estado estable.

Deducción de *L*

En lo que resta de esta sección supondremos que $\rho < 1$, lo cual asegura que sí existe una distribución de probabilidades de estado estable, como la de (25). Utilizamos ahora la distribución de probabilidades de estado estable de (25) para determinar varias cantidades que interesan. Por ejemplo, si suponemos que el estado estable ya se alcanzó, la cantidad de clientes promedio presente en el sistema de colas (llámelo L) se obtiene con

$$\begin{aligned} \sum_{j=0}^{j=\infty} j\pi_j &= \sum_{j=0}^{j=\infty} j\rho^j(1 - \rho) \\ &= (1 - \rho)\sum_{j=0}^{j=\infty} j\rho^j \end{aligned}$$

Si definimos

$$\sum_{j=0}^{j=\infty} j\rho^j = \rho + 2\rho^2 + 3\rho^3 + \cdots$$

observamos que $\rho S' = \rho^2 + 2\rho^3 + 3\rho^4 + \cdots$. Si hacemos una sustracción, obtenemos

$$S' - \rho S' = \rho + \rho^2 + \cdots = \frac{\rho}{1 - \rho}$$

Por lo tanto,

$$S' = \frac{\rho}{(1 - \rho)^2}$$

y,

$$L = (1 - \rho)\frac{\rho}{(1 - \rho)^2} = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda} \tag{26}$$

Deducción de L_q

En algunas circunstancias, interesa conocer la cantidad esperada de personas en la cola. A esta cantidad la denotaremos con L_q. Observe que si están presentes 0 o 1 cliente en el sistema, entonces nadie está en la cola, pero si j personas están presentes ($j \geq 1$), habrá $j - 1$ en la línea de espera. Por lo tanto, si estamos en el estado estable,

$$\begin{aligned} L_q &= \sum_{j=1}^{j=\infty} (j - 1)\pi_j = \sum_{j=1}^{j=\infty} j\pi_j - \sum_{j=1}^{j=\infty} \pi_j \\ &= L - (1 - \pi_0) = L - \rho \end{aligned}$$

donde la última ecuación se infiere de (24). Como $L = \frac{\rho}{1-\rho}$, entonces

$$L_q = \frac{\rho}{1 - \rho} - \rho = \frac{\rho^2}{1 - \rho} = \frac{\lambda^2}{\mu(\mu - \lambda)} \tag{27}$$

Deducción de L_s

También es importante L_s, la cantidad esperada de clientes en servicio. Para un sistema de colas $M/M/1/GD/\infty/\infty$,

$$L_s = 0\pi_0 + 1(\pi_1 + \pi_2 + \cdots) = 1 - \pi_0 = 1 - (1 - \rho) = \rho$$

Como cada cliente que está presente en la cola o en servicio, se deduce que para cualquier sistema de colas (no justamente un sistema $M/M/1/GD/\infty/\infty$), $L = L_s + L_q$. Por lo tanto, luego de aplicar las fórmulas para L y L_s, podríamos haber determinado L_q a partir de

$$L_q = L - L_s = \frac{\rho}{1 - \rho} - \rho = \frac{\rho^2}{1 - \rho}$$

Fórmula de líneas de espera $L = \lambda W$

El interés se centra a menudo en el tiempo que un cliente representativo pasa en un sistema de líneas de espera. Definimos W como el tiempo previsto que un cliente pasa en el sistema de líneas de espera, que comprende el tiempo en la cola y el tiempo en servicio; y W_q como el tiempo previsto que un cliente pasa formado en la cola. Tanto W como W_q se calculan suponiendo que ya se alcanzó el estado estable. Mediante el uso de un resultado eficaz que se conoce como **fórmula de líneas de espera de Little**, W y W_q se podrían calcular con facilidad a partir de L y L_q. Primero definimos (para cualquier sistema de líneas de espera o cualquier subconjunto de un sistema de líneas de espera) las cantidades siguientes:

λ = número promedio de llegadas al sistema por unidad de tiempo

L = número promedio de clientes presentes en el sistema de colas

L_q = número promedio de clientes formados en la cola

L_s = número promedio de clientes en servicio

W = tiempo promedio que un cliente pasa en el sistema

W_q = tiempo promedio que un cliente pasa en la cola

W_s = tiempo promedio que un cliente pasa en servicio

Todos los promedios son de estado estable en estas definiciones. En la mayor parte de los sistemas de líneas de espera, la fórmula de líneas de espera de Little se podría resumir como en el teorema 3.

TEOREMA 3

Para *cualquier* sistema de colas en el cual existe una distribución de estado estable, se cumplen las relaciones siguientes:

$$L = \lambda W \tag{28}$$

$$L_q = \lambda W_q \tag{29}$$

$$L_s = \lambda W_s \tag{30}$$

Antes de usar estos resultados importantes, presentamos una justificación intuitiva de (28). Primero observe que ambos miembros de (28) tienen las mismas unidades (suponemos que la unidad de tiempo es la hora). A esto se llega luego de ver que L se expresa en términos de cantidad de clientes, λ se expresa en clientes por hora y W está en horas. Por lo tanto, λW tiene las mismas unidades (clientes) que L. Si desea revisar una demostración rigurosa del teorema de Little, refiérase a Ross (1970). Nos damos por satisfechos con el análisis heurístico siguiente.

Considere un sistema de líneas de espera en el cual los clientes son atendidos de acuerdo a como van llegando. Una llegada arbitraria entra al sistema (suponga que ya se alcanzó el estado estable). Este cliente permanece en el sistema hasta que completa su servicio, y hasta su partida habrá (en promedio) L clientes presentes en el sistema. Pero cuando este cliente se va, ¿quién se quedará en el sistema? Sólo aquellos clientes que llegan durante el tiempo que el cliente inicial pasa en el sistema. Puesto que el cliente inicial pasa un promedio de W horas en el sistema, un promedio de λW clientes llegará durante la estancia del cliente en el sistema. Por lo tanto, $L = \lambda W$. La demostración "real" de $L = \lambda W$ es virtualmente independiente de la cantidad de servidores, de la distribución del tiempo entre llegadas, de la disciplina del servicio y de la distribución del tiempo de servicio. Por lo tanto, siempre que exista un estado estable podremos aplicar las ecuaciones (28) a (30) a cualquier sistema de líneas de espera.

Con el fin de ilustrar el uso de (28) y (29) determinemos W y W_q para un sistema de colas $M/M/1/GD/\infty/\infty$. De acuerdo con (26),

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

Entonces, con (28) se tiene

$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu - \lambda} \tag{31}$$

Con (27) se llega a

$$L_q = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

y de (29) se infiere

$$W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu - \lambda)} \tag{32}$$

Observe que (como se esperaba) cuando ρ se aproxima a 1, tanto W como W_q se vuelven muy grandes. Para ρ cercana a cero, W_q se aproxima a cero, pero para ρ pequeña, W se aproxima a $\frac{1}{\mu}$, el tiempo medio de servicio.

En los tres ejemplos que siguen se muestran aplicaciones de las fórmulas que se han desarrollado.

EJEMPLO 3 Banco que ofrece servicio en su automóvil

Un promedio de 10 automóviles por hora llegan a un cajero con un solo servidor que proporciona servicio sin que uno descienda del automóvil. Suponga que el tiempo de servicio promedio por cada cliente es 4 minutos, y que tanto los tiempos entre llegadas y los tiempos de servicio son exponenciales. Conteste las preguntas siguientes:

1 ¿Cuál es la probabilidad de que el cajero esté ocioso?

2 ¿Cuál es el número promedio de automóviles que están en la cola del cajero? (Se considera que un automóvil que está siendo atendido no está en la cola esperando).

3 ¿Cuál es la cantidad promedio de tiempo que un cliente pasa en el estacionamiento del banco (incluyendo el tiempo en servicio)?

4 ¿Cuántos clientes atenderá en promedio el cajero por hora?

Solución De acuerdo con las premisas, estamos trabajando con un sistema de colas de $M/M/1/GD/\infty/\infty$ para el cual $\lambda = 10$ automóviles por hora y $\mu = 15$ automóviles por hora. Por lo tanto, $\rho = \frac{10}{15} = \frac{2}{3}$.

1 Según (24), $\pi_0 = 1 - \rho = 1 - \frac{2}{3} = \frac{1}{3}$. Por lo tanto, el cajero estará ocioso un promedio de un tercio del tiempo.

2 Determinemos L_q. A partir de (27),

$$L_q = \frac{\rho^2}{1-\rho} = \frac{(\frac{2}{3})^2}{1-\frac{2}{3}} = \frac{4}{3} \quad \text{clientes}$$

3 Estimemos W. A partir de (28), $W = \frac{L}{\lambda}$. Entonces, según (26),

$$L = \frac{\rho}{1-\rho} = \frac{\frac{2}{3}}{1-\frac{2}{3}} = 2 \quad \text{clientes}$$

Por lo tanto, $W = \frac{2}{10} = \frac{1}{5}$ h $= 12$ minutos (W tiene las mismas unidades de λ).

4 Si el cajero siempre estuviera ocupado, atendería un promedio de $\mu = 15$ clientes por hora. Según la parte (1), sabemos que el cajero está ocupado sólo dos tercios del tiempo. Por lo tanto, durante cada hora, el cajero atenderá un promedio de $(\frac{2}{3})(15) = 10$ clientes. Éste debe ser el caso porque, en el estado estable, 10 clientes llegan cada hora, de modo que 10 clientes deben dejar cada hora el sistema.

EJEMPLO 4 Estación de servicio

Suponga que todos los dueños de automóvil acuden a la gasolinera cuando sus tanques están a la mitad.† En el momento actual llega un promedio de 7.5 clientes por hora a una

†Este ejemplo está basado en Erickson (1973).

gasolinera que tiene una sola bomba. Se requiere un promedio de 4 minutos para servir a un automóvil. Suponga que los tiempos entre llegadas y los tiempos de servicio son exponenciales.

1 Calcule L y W para las circunstancias actuales.

2 Suponga que hay un déficit de gasolina y que hay compras de pánico. Para modelar este fenómeno, suponga que todos los dueños de automóvil compran ahora gasolina cuando sus tanques tienen $\frac{3}{4}$ de combustible. Como cada dueño ahora pone menos gasolina en el tanque cada vez que acude a la gasolinera, suponemos que el tiempo de servicio promedio se reduce a 3 minutos y un tercio. ¿Qué tanto afectan a L y W las compras de pánico?

Solución **1** Tenemos un sistema $M/M/1/GD/\infty/\infty$ con $\lambda = 7.5$ automóviles por hora y $\mu = 15$ automóviles por hora. Por lo tanto, $\rho = \frac{7.5}{15} = .50$. De acuerdo con (26), $L = \frac{.50}{1-.50} = 1$, y con (28), $W = \frac{L}{\lambda} = \frac{1}{7.5} = 0.13$ horas. Por lo tanto, en estas circunstancias, todo está bajo control, por lo que, al parecer, son improbables las líneas de espera largas.

2 Se tiene ahora un sistema $M/M/1/GD/\infty/\infty$ con $\lambda = 2(7.5) = 15$ automóviles por hora. (Esto se infiere porque cada dueño de automóvil llenará su tanque dos veces). Ahora $\mu = \frac{60}{3.333} = 18$ automóviles por hora, y $\rho = \frac{15}{18} = \frac{5}{6}$. Entonces,

$$L = \frac{\frac{5}{6}}{1 - \frac{5}{6}} = 5 \text{ automóviles} \qquad \text{y} \qquad W = \frac{L}{\lambda} = \frac{5}{15} = \frac{1}{3} \text{ horas} = 20 \text{ minutos}$$

Por lo tanto, las compras de pánico ocasionan largas líneas de espera.

En el ejemplo 4 se ilustra el hecho de que cuando ρ se aproxima a 1, L y, por consiguiente, W se incrementan con rapidez. Este hecho se muestra en la tabla 5.

Un modelo de optimización de colas

En el ejemplo 5 se ilustra cómo se usa la teoría de colas para auxiliar en la toma de decisiones.

TABLA 5
Relación entre ρ y L para un sistema $M/M/1/GD/\infty/\infty$

ρ	L para un sistema $M/M/1/GD/\infty/\infty$
0.30	0.43
0.40	0.67
0.50	1.00
0.60	1.50
0.70	2.33
0.80	4.00
0.90	9.00
0.95	19.00
0.99	99.00

EJEMPLO 5 Centro de herramienta

Los mecánicos que trabajan en una planta troqueladora deben solicitar su herramienta en un centro de herramienta.[†] Un promedio de 10 mecánicos por hora llega pidiendo su equipo. Por el momento, un empleado atiende este centro; su salario es de 6 dólares por hora y tarda un promedio de 5 minutos en cumplir con cada pedido de herramienta solicitada. Como cada mecánico produce 10 dólares en valor de bienes por hora, cada hora que un mecánico tarda en el centro de herramienta cuesta a la compañía 10 dólares. La compañía está pensando si valdría la pena o no contratar (a 4 dólares la hora) un ayudante para el empleado. Si se contrata al ayudante, el empleado tardará un promedio de sólo 4 minutos en reunir el equipo que le solicita cada mecánico. Suponga que los tiempos de servicio y de llegadas son exponenciales. ¿Se debe contratar al ayudante?

Solución Los problemas en los que un analista debe elegir entre varios sistemas de colas reciben el nombre de **problemas de optimización de colas**. En el problema presente, el objetivo de la compañía es minimizar la suma del costo de servicio por hora y el costo esperado por hora debido a los tiempos muertos de los mecánicos. En los problemas de optimización de colas, el componente del costo debido a los clientes que esperan en la cola se denomina costo por demora. Por lo tanto, la compañía desea minimizar

$$\frac{\text{Costo esperado}}{\text{hora}} = \frac{\text{costo de servicio}}{\text{hora}} + \frac{\text{costo esperado de demora}}{\text{hora}}$$

El cálculo del costo de servicio por hora es, por lo regular, simple. La manera más fácil de calcular el costo por la demora por hora es observar que

$$\frac{\text{Costo esperado por demora}}{\text{Hora}} = \left(\frac{\text{costo esperado por demora}}{\text{cliente}}\right)\left(\frac{\text{clientes esperados}}{\text{hora}}\right)$$

En el problema presente,

$$\frac{\text{Costo esperado por demora}}{\text{Cliente}} = \left(\frac{\$10}{\text{hora-mecánico}}\right)\left(\begin{matrix}\text{horas promedio que el}\\ \text{mecánico pasa en el sistema}\end{matrix}\right)$$

Por lo tanto,

$$\frac{\text{Costo esperado por demora}}{\text{Cliente}} = 10W \quad \text{y} \quad \frac{\text{costo esperado por demora}}{\text{hora}} = 10W\lambda$$

Ya podemos comparar el costo esperado por hora si no se contrata al ayudante con el costo esperado por hora si se contrata al ayudante. Si no se contrata al ayudante, $\lambda = 10$ mecánicos por hora y $\mu = 12$ mecánicos por hora. De acuerdo con (31), $W = \frac{1}{12-10} = \frac{1}{2}$ hora. Como el empleado gana 6 dólares por hora, tenemos que

$$\frac{\text{Costo de servicio}}{\text{Hora}} = \$6 \quad \text{y} \quad \frac{\text{costo esperado por demora}}{\text{hora}} = 10(\tfrac{1}{2})10 = \$50$$

[†]Este ejemplo se basa en Brigham (1955).

Por lo tanto, sin el ayudante, el costo esperado por hora es 6 + 50 = 56 dólares. Con el ayudante, μ = 15 clientes por hora. Entonces $W = \frac{1}{15-10} = \frac{1}{5}$ de hora y

$$\frac{\text{Costo esperado por demora}}{\text{Hora}} = 10(\tfrac{1}{5})(10) = \$20$$

Como el costo de servicio por hora es ahora 6 + 4 = 10 dólares por hora, el costo esperado por hora con el ayudante es 20 + 10 = 30 dólares. Por lo tanto, se debe contratar al ayudante porque con él se ahorran 50 − 20 = 30 dólares por hora en costos por demora, que es más que el salario de 4 dólares por hora.

La fórmula para la cola $L = \lambda W$ es muy general y se puede aplicar en muchas situaciones que no parecen ser problemas de líneas de espera. Piense en cualquier situación donde una cantidad (tal como solicitudes de préstamos hipotecarios, papas en McDonald's, ingresos por ventas de computadoras) fluye por un sistema. Si hacemos

L = cantidad promedio de clientes presentes en el sistema de colas

λ = tasa de llegadas al sistema

W = tiempo promedio que una unidad de la cantidad pasa en el sistema

entonces, $L = \lambda W$, o bien, $W = L/\lambda$.

Siguen algunos ejemplos de $L = \lambda W$ en situaciones que no son líneas de espera.

EJEMPLO 6 Papas en McDonald's

Una sucursal de McDonald's usa un promedio de 10 000 libras de papas por semana. La cantidad promedio de libras de papas disponibles es 5 000. ¿En cuánto tiempo estarán, en promedio, las papas en el restaurante antes de que las utilicen?

Solución Sabemos que L = 5 000 libras y λ = 10 000 libras/semana. Por lo tanto, W = 5 000 libras/(10 000 libras/semana) = .5 semana.

EJEMPLO 7 Cuentas por cobrar

Una tienda que comercializa computadoras vende un valor de 300 000 dólares de computadoras por año. Las cuentas por cobrar son, en promedio, 45 000 dólares. ¿Cuánto tiempo toma desde el momento en que un cliente es facturado hasta que la tienda recibe el pago?

Solución Tenemos que L = 45 000 dólares y λ = 300 000 dólares/año. Por lo tanto, W = 45 000 dólares/(300 000 dólares/año) = .15 año.

Hoja de cálculo para el sistema de colas $M/M/1/GD/\infty/\infty$

MM1.xls

En la figura 13 (archivo MM1.xls) se muestra una plantilla que es posible usar para determinar cantidades importantes del sistema de líneas de espera $M/M/1/GD/\infty/\infty$. Escriba simplemente λ en la celda A4 y μ en la celda B4. L, L_q, L_s, W, W_q y W_s se calculan en los renglones 6 y 8. La columna B da las probabilidades de estado estable (calculadas a partir de (24) y (25)). Estamos suponiendo que λ y μ son tales que es muy pequeña la probabilidad de que más de 1 000 clientes estén presentes. En la figura 13 se han introducido los valores de λ y μ para el ejemplo 3.

	A	B	C
1	M/M/1	QUEUE	
2			
3	LAMBDA?	MU?	RO
4	10	15	0.66666667
5	L	LQ	LS
6	2	1.33333333	0.66666667
7	W	WQ	WS
8	0.2	0.13333333	0.06666667
9	J	PI(J)	
10	0	0.33333333	
11	1	0.22222222	
12	2	0.14814815	
13	3	0.09876543	
14	4	0.06584362	
15	5	0.04389575	
16	6	0.02926383	
17	7	0.01950922	
18	8	0.01300615	
19	9	0.00867076	
20	10	0.00578051	
21	11	0.00385367	
22	12	0.00256912	
23	13	0.00171274	
24	14	0.00114183	
25	15	0.00076122	
26	16	0.00050748	
27	17	0.00033832	
28	18	0.00022555	
29	19	0.00015036	
30	20	0.00010024	
31	21	6.6829E-05	
32	22	4.4552E-05	
33	23	2.9702E-05	
34	24	1.9801E-05	
35	25	1.3201E-05	
36	26	8.8005E-06	
37	27	5.867E-06	
38	28	3.9113E-06	
39	29	2.6075E-06	
40	30	1.7384E-06	

FIGURA 13
Cola *M/M/*1

A	A	B	C
41	31	0.0000012	
42	32	0.0000008	
43	33	0.0000005	
44	34	0.0000003	
45	35	0.0000002	
46	36	0.0000002	
47	37	0.0000001	
48	38	6.8E-08	
49	39	4.5E-08	
50	40	3.0E-08	
51	41	2.0E-08	
52	42	1.3E-08	
53	43	8.9E-09	
54	44	6.0E-09	
55	45	4.0E-09	
56	46	2.6E-09	
57	47	1.8E-09	
58	48	1.2E-09	
59	49	7.8E-10	
60	50	5.2E-10	
61	51	3.5E-10	
62	52	2.3E-10	
63	53	1.5E-10	
64	54	1.0E-10	
65	55	6.9E-11	
66	56	4.6E-11	
67	57	3.1E-11	
68	58	2.0E-11	
69	59	1.4E-11	
70	60	9.1E-12	
71	61	6.0E-12	
72	62	4.0E-12	
73	63	2.7E-12	
74	64	1.8E-12	
75	65	1.2E-12	
76	66	8.0E-13	
77	67	5.3E-13	
78	68	3.5E-13	
79	69	2.4E-13	
80	70	1.6E-13	

FIGURA 13
(Continuación)

PROBLEMAS

Grupo A

1[†] Todos los pasajeros y su equipaje tienen que ser revisados para investigar si no llevan armas. Suponga que 10 pasajeros por minuto, en promedio, llegan al aeropuerto de Gotham City (los tiempos entre llegadas son exponenciales). Para investigar si los pasajeros llevan armas, el aeropuerto debe contar con un punto de revisión que consta de un detector de metales y un aparato de rayos X. Se requieren dos empleados siempre que el punto de revisión está en operación. Un punto de revisión puede verificar un promedio de 12 pasajeros por minuto (el tiempo para revisar a los pasajeros es exponencial). Si se supone que el aeropuerto tiene sólo un punto de revisión, conteste las preguntas siguientes:

a ¿Cuál es la probabilidad de que un pasajero tenga que esperar antes de ser revisado en busca de armas?

b ¿Cuántos pasajeros, en promedio, hacen fila para pasar al punto de revisión?

[†]Basado en Gilliam (1979).

c ¿Cuánto tiempo pasará el pasajero en el punto de revisión, en promedio?

2 El Departamento de Ciencias de Decisión pretende determinar si renta una copiadora lenta o una rápida. El departamento opina que el tiempo de un empleado vale 15 dólares por hora. La renta de la copiadora lenta es de 4 dólares la hora, y un empleado requiere un promedio de 10 minutos para completar el copiado (tiempo con distribución exponencial). La renta de la copiadora rápida es de 5 dólares por hora y a un empleado le toma un promedio de 6 minutos terminar el copiado. Un promedio de 4 empleados por hora necesita usar la copiadora (los tiempos entre llegadas son exponenciales). ¿Cuál copiadora debe rentar el departamento?

3 Para un sistema de líneas de espera $M/M/1/GD/\infty/\infty$, suponga que λ y μ son el doble.

a ¿Qué tanto cambia L?

b ¿Qué tanto cambia W?

c ¿Qué tanto cambia la distribución de probabilidades de estado estable?

4 Un restaurante especializado en bocadillos tiene una ventanilla desde la cual da servicio a los automovilistas en su vehículo. Llegan en promedio 40 clientes por hora a la ventanilla. Se requiere 1 minuto en promedio atender a un cliente. Suponga que los tiempos entre llegadas y de servicio son exponenciales.

a ¿Cuántos clientes están en la cola, en promedio?

b ¿Cuánto tiempo, en promedio, pasa un cliente en el restaurante (desde el tiempo de llegada hasta que el tiempo de servicio se completa)?

c ¿En qué fracción de tiempo hay más de tres automóviles esperando servicio [esto comprende el automóvil (si acaso hay alguno) en la ventanilla]?

5 En un sábado cualquiera, el Red Lobster atiende a 1 000 clientes. El restaurante está abierto 12 horas. En promedio, hay 150 clientes. ¿Cuánto tiempo, en promedio, pasa un cliente en el restaurante?

6 La sala de maternidad local recibe 1 500 bebés al año. Cinco camas de la sala de maternidad están ocupadas, en promedio. ¿Cuánto tiempo la madre promedio permanece en la sala?

7 Suponga que un promedio de 125 paquetes por segundo de información llegan a un selector de vía y que se necesita un promedio de .002 segundos procesar cada paquete. Suponga tiempos entre llegadas y de servicio exponenciales, y conteste las preguntas siguientes:

a ¿Cuál es el número promedio de paquetes que están esperando entrar al selector de vía?

b ¿Cuál es la probabilidad que 10 o más paquetes estén presentes?

Grupo B

8 Refiérase al problema 1. Suponga que la aerolínea desea determinar cuántos puntos de revisión operar para minimizar los costos de operación y los costos por demora en un periodo de 10 años. Suponga que el costo por retrasar a un pasajero una hora es de 10 dólares y que el aeropuerto está abierto todos los días durante 16 horas. Cuesta un millón de dólares comprar un detector de metales y un aparato de rayos X, contratar al personal que lo opere y el mantenimiento durante 10 años. Por último, suponga que todos los pasajeros tienen la misma probabilidad de pasar por un punto de revisión.

9† Cada una de las máquinas de la línea de ensamble de Widgetco deja de funcionar, en promedio, una vez al minuto. Hay trabajadores asignados para reiniciar una máquina que se paró. La compañía paga a cada trabajador c_s dólares por hora, y estima que cada hora en que una máquina está inactiva cuesta a la empresa c_m dólares en producción perdida. La información señala que el tiempo entre paros sucesivos de una máquina y el tiempo para reiniciarla son exponenciales. Widgetco planea asignar a cada trabajador un cierto número de máquinas para que las vigilen y las reparen. Sea M = número total de máquinas de Widgetco, w = número de trabajadores contratados por Widgetco y $R = \frac{M}{w}$ = máquinas asignadas a cada trabajador.

a Exprese los costos por hora de Widgetco en términos de R y M.

b Demuestre que el valor óptimo de R no depende del valor de M.

c Utilice el cálculo para demostrar que los costos se minimizan cuando se escoge

$$R = \frac{\frac{\mu}{60}}{1 + \left(\frac{c_m}{c_s}\right)^{1/2}}$$

d Suponga que c_m = 78 centavos y c_s = 2.75 dólares. Widgetco tiene 200 máquinas, y un trabajador puede reiniciar una máquina en un promedio de 7.8 segundos. ¿Cómo puede la compañía minimizar los costos?

e En los incisos (a) a (d) hemos supuesto tácitamente que, en cualquier punto en el tiempo, la tasa a la cual se paran las máquinas asignadas a un trabajador no depende de la cantidad de máquinas asignadas al trabajador que están trabajando ahora en forma adecuada. ¿Parece razonable esta suposición?

10 Imagine un aeropuerto donde los taxis y los clientes llegan (tiempos entre llegadas exponenciales) con tasas respectivas de uno y dos por minuto. No importa cuántos taxis estén presentes, un taxi tiene que esperar. Si un cliente que llega no encuentra un taxi, se retira de inmediato.

a Modele este sistema como un proceso de nacimiento-muerte. (*Sugerencia*: determine cuál es el estado del sistema en cualquier tiempo dado y trace un diagrama de tasas.)

b Encuentre la cantidad promedio de taxis que están esperando un cliente.

c Suponga que todos los clientes que usan un taxi pagan una tarifa de 2 dólares. Durante una hora cualquiera, ¿qué ingreso recibirán los taxis?

11 Un banco pretende determinar cuál máquina debe rentar de las dos que hay para procesar cheques. La renta de la máquina 1 es 10 000 dólares por año y procesa 1000 cheques por hora. La renta de la máquina 2 es de 15 000 dólares por año y elabora 1600 cheques por hora. Suponga que las máquinas trabajan 8 horas al día, cinco días a la semana, 50 semanas al año. El banco debe procesar un promedio de 800 cheques por hora, y el cheque promedio procesado es por 100 dólares. Suponga que la tasa de interés anual es de 20%. Después determine el costo que representa para el banco (en intereses perdidos) cada hora que un cheque pasa esperando el proceso y el tiempo en procesarlo. Si se supone, además, que los tiempos entre llegadas y los tiempos de servicio son exponenciales, ¿qué máquina debe rentar el banco?

†Basado en Vogel (1979).

12[‡] Una planta de neumáticos debe fabricar un promedio de 100 neumáticos por día. La planta produce los neumáticos en un lote de tamaño x. El gerente debe determinar el tamaño del lote x que minimice el tiempo en que un lote pasa en la planta. Desde el momento en que un lote de neumáticos llega, se requiere un promedio de $\frac{1}{20}$ de día arrancar la planta para fabricar neumáticos. Una vez que la planta ya está en funciones, toma un promedio de $\frac{1}{150}$ día producir un neumático. Suponga que el tiempo para producir un lote de neumáticos está distribuido en forma exponencial, y que el tiempo para que un lote de neumáticos "arribe" también está exponencialmente distribuido. Determine el tamaño del lote que minimice el tiempo previsto que un lote pasa en la planta (desde la llegada del lote hasta el momento en que se completa la producción del lote).

13 Un empleado de la Oficina estatal de desempleo es el responsable de procesar las formas de una compañía cuando ésta abre sus puertas. El trabajador es capaz de procesar un promedio de cuatro formas por semana. En 2002, un promedio de 1.8 compañías por semana presentó formas para su proceso, y el empleado tenía trabajo pendiente acumulado de .45 de semana. En 2003, un promedio de 3.9 compañías por semana presentó formas para procesarlas, por lo que el empleado tenía trabajo pendiente acumulado de cinco semanas. El pobre empleado fue despedido, pero puso pleito para recobrar su puesto de trabajo. La corte estableció que como la cantidad de trabajo enviado al trabajador casi se había duplicado, el trabajo pendiente del empleado también se debería haber duplicado. Como su trabajo pendiente se incrementó por más de un factor de 10, él debió haber estado flojeando, así que el Estado tenía justificación por haberlo despedido. Aplique la teoría de colas para defender al empleado (¡está basado en un caso real!)

[‡]Basado en Karmarkar (1985).

14 En relación con el modelo $M/M/1/GD/\infty/\infty$ demuestre que los resultados siguientes se cumplen

a $W = (L + 1)W_s$.

b $W_q = LW_s$.

c Interprete los resultados en (a) y (b).

15 Desde el momento en que se presenta una petición de información hasta el momento en que se entrega, una base de datos se tarda un promedio de tres segundos en responder. Encontramos que la base de datos está inactiva alrededor de 20% del tiempo. Conteste las preguntas siguientes suponiendo que la base de datos se puede modelar como un sistema $M/M/1$.

a ¿Cuál es el tiempo promedio de servicio por pregunta a la base de datos?

b ¿Cuál es el número promedio de preguntas en el sistema?

c ¿Cuál es la probabilidad de que cinco o más preguntas estén presentes?

20.5 Sistema de colas $M/M/1/GD/c/\infty$

En esta sección se trata el sistema de colas $M/M/1/GD/c/\infty$. Recuerde que este sistema de colas es un sistema $M/M/1/GD/\infty/\infty$ con una capacidad total de c clientes. El sistema $M/M/1/GD/c/\infty$ es idéntico al sistema $M/M/1/GD/\infty/\infty$, excepto por el hecho de que cuando c clientes están presentes, a todas las llegadas se le niega la entrada, y el sistema las pierde por siempre. Al igual que en la sección 20.4, suponemos que los tiempos entre llegadas son exponenciales con tasa λ, y los tiempos de servicio son exponenciales con tasa μ. Entonces el sistema $M/M/1/GD/c/\infty$ se podría modelar (véase figura 14) como un proceso de nacimiento-muerte con los parámetros siguientes:

$$\begin{aligned} \lambda_j &= \lambda \qquad (j = 0, 1, \ldots, c-1) \\ \lambda_c &= 0 \\ \mu_0 &= 0 \\ \mu_j &= \mu \qquad (j = 1, 2, \ldots, c) \end{aligned} \tag{33}$$

Como $\lambda_c = 0$, el sistema nunca alcanzará el estado $c + 1$ (o cualquier estado superior). Conviene, como en la sección 20.4, definir $\rho = \frac{\lambda}{\mu}$. Luego podemos aplicar las ecuaciones (16) a (19) para encontrar que si $\lambda \neq \mu$, las probabilidades de estado estable para el modelo $M/M/1/GD/c/\infty$ se obtienen con

$$\begin{aligned} \pi_0 &= \frac{1-\rho}{1-\rho^{c+1}} \\ \pi_j &= \rho^j \pi_0 \qquad (j = 1, 2, \ldots, c) \\ \pi_j &= 0 \qquad (j = c+1, c+2, \ldots) \end{aligned} \tag{34}$$

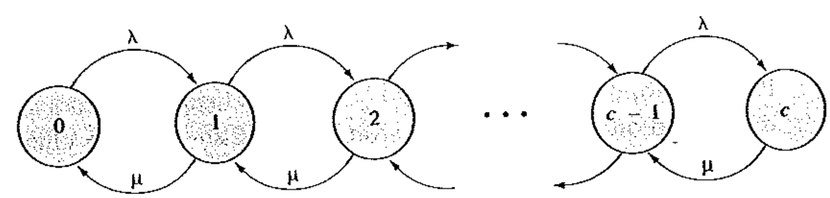


FIGURA 14
Diagrama de tasas para el sistema de colas $M/M/1/GD/c/\infty$

Al combinar (34) con el hecho de que $L = \sum_{j=0}^{j=c} j\pi_j$, es posible demostrar que cuando $\lambda \neq \mu$,

$$L = \frac{\rho[1 - (c + 1)\rho^c + c\rho^{c+1}]}{(1 - \rho^{c+1})(1 - \rho)} \tag{35}$$

Si $\lambda = \mu$, entonces todas las c_j en (16) son iguales a 1, y todas las π_j deben ser iguales. Por lo tanto, si $\lambda = \mu$, las probabilidades de estado estable para el sistema $M/M/1/GD/c/\infty$ son

$$\pi_j = \frac{1}{c + 1} \qquad (j = 0, 1, \ldots, c)$$
$$L = \frac{c}{2} \tag{36}$$

Al igual que con el sistema $M/M/1/GD/\infty/\infty$, $L_s = 0\pi_0 + 1(\pi_1 + \pi_2 + \cdots) = 1 - \pi_0$. Como antes, podríamos determinar L_q a partir de $L_q = L - L_s$.

La determinación de W y W_q a partir de (28) y (29) es cuestión engañosa. Recuerde que en (28) y (29), λ representa la cantidad promedio de clientes por unidad de tiempo que *entra realmente* al sistema. En el modelo de capacidad finita, llega un promedio de λ llegadas por unidad de tiempo, pero $\lambda\pi_c$ de estas llegadas encuentran el sistema lleno en toda su capacidad y se retiran. Por lo tanto, un promedio de $\lambda - \lambda\pi_c = \lambda(1 - \pi_c)$ llegadas por unidad de tiempo en realidad entran al sistema. Al combinar este hecho con (28) y (29) se obtiene

$$W = \frac{L}{\lambda(1 - \pi_c)} \qquad \text{y} \qquad W_q = \frac{L_q}{\lambda(1 - \pi_c)} \tag{37}$$

Para un sistema $M/M/1/GD/c/\infty$, existirá un estado estable incluso si $\lambda \geq \mu$. La razón es que incluso si $\lambda \geq \mu$, la capacidad finita del sistema evita que la cantidad de personas en el sistema se "amplifique".

EJEMPLO 8 Una peluquería

Una peluquería que atiende una sola persona tiene un total de 10 sillas. Los tiempos entre llegadas siguen una distribución exponencial, y un promedio de 20 posibles clientes llega cada hora a la peluquería. Los clientes que al llegar a la peluquería la encuentran llena, ya no entran. El peluquero se tarda un promedio de 12 minutos en cortar el cabello a cada cliente. Los tiempos del corte de cabello están distribuidos en forma exponencial.

1 En promedio, ¿cuántos cortes de cabello por hora completará el peluquero?

2 ¿Cuánto tiempo pasará en promedio un cliente en la peluquería?

Solución **1** Una fracción π_{10} de todas las llegadas encontrará que la peluquería está llena. Por lo tanto, un promedio de $\lambda(1 - \pi_{10})$ entrará cada hora a la peluquería. Todos los clientes que entran obtendrán su corte de cabello, de modo que el peluquero hará un promedio de $\lambda(1 - \pi_{10})$ cortes por hora. Según este problema, $c = 10$, $\lambda = 20$ clientes por hora y $\mu = 5$ clientes por hora. Entonces, $\rho = \frac{20}{5} = 4$, y (34) da

$$\pi_0 = \frac{1 - 4}{1 - 4^{11}}$$

y

$$\pi_{10} = 4^{10}\left(\frac{1 - 4}{1 - 4^{11}}\right) = \frac{-3(4^{10})}{1 - 4^{11}} = .75$$

Por lo tanto, un promedio de $20(1 - \frac{3}{4}) = 5$ clientes por hora conseguirá su corte de cabello. Esto quiere decir que un promedio de $20 - 5 = 15$ clientes por hora no entrarán a la peluquería.

2 Para determinar W, utilice (35) y (37). Según (35),

$$L = \frac{4[1 - 11(4^{10}) + 10(4^{11})]}{(1 - 4^{11})(1 - 4)} = 9.67 \text{ clientes}$$

Entonces con (37) se obtiene

$$W = \frac{9.67}{20(1 - \frac{3}{4})} = 1.93 \text{ horas}$$

¡Esta peluquería está llena! El peluquero haría bien en contratar por lo menos a otro peluquero.

Hoja de cálculo para el sistema de líneas de espera $M/M/1/GD/c/\infty$

MM1CAP.xls

En la figura 15 (archivo MM1CAP.xls) se presenta una plantilla que se puede usar para calcular cantidades importantes para el sistema de colas $M/M/1/GD/c/\infty$. Introduzca λ en la celda B2, μ en la celda C2 y c (suponemos que $c \le 1\,000$) en la celda D2. En la celda F2 se proporciona la probabilidad de estado estable de que el estado es c. Ésta es la fracción de todos los clientes que encuentran el sistema lleno. En el renglón 4 se calculan las cantidades L, L_s, L_q, W, W_s y W_q. En la columna E se calculan las probabilidades de estado estable a partir de las ecuaciones (16) a (18). En la figura 15 se presentan los datos del ejemplo 8.

PROBLEMAS

Grupo A

1 Una instalación de servicio consta de un servidor, el cual puede atender a un promedio de 2 clientes por hora (tiempos de servicio exponenciales). Un promedio de 3 clientes por hora llega a la instalación (se supone que los tiempos entre llegadas son exponenciales). La capacidad del sistema es de 3 clientes.

a ¿Cuántos clientes potenciales entran, en promedio, al sistema cada hora?

b ¿Cuál es la probabilidad de que el servidor esté ocupado?

2 Un promedio de 40 automóviles por hora (los tiempos entre llegadas siguen una distribución exponencial) está tentado a pasar por el servicio para automovilistas del restaurante Hot Dog King. Si un total de más de 4 automóviles están haciendo cola (incluso el auto al que están atendiendo) un automóvil no entrará a la cola. Se requiere un promedio de 4 minutos (distribución exponencial) para atender a un automóvil.

a ¿Cuál es la cantidad promedio de automóviles que está esperando atención? (No se incluye un vehículo al que están atendiendo.?

b ¿Cuántos vehículos serán atendidos, en promedio, por hora?

c Apenas me he formado en la cola para que me atiendan. En promedio, ¿cuánto tiempo esperaré antes de que me sirvan mis bocadillos?

3 Llega un promedio de 125 paquetes de información por minuto a un selector de vía para internet. Se requiere un promedio de .002 segundos para procesar un paquete de información. El diseño del selector de vía permite tener una memoria temporal limitada para almacenar mensajes en espera. Cualquier mensaje que llegue cuando esta memoria está llena, se pierde. Si suponemos que los tiempos entre llegadas y de servicio siguen una distribución exponencial, ¿de qué tamaño debe ser la memoria temporal para tener la certeza de que se pierde, cuando mucho, un mensaje en un millón?

Grupo B

4 Demuestre que si $\rho \neq 1$

$$1 + \rho + \rho^2 + \cdots + \rho^c = \frac{1 - \rho^{c+1}}{1 - \rho}$$

(*Sugerencia*: recuerde cómo se evaluó $1 + \rho + \rho^2 + \cdots$)

5 Utilice la respuesta del problema 3 para derivar las probabilidades de estado estable para el sistema $M/M/1/GD/c/\infty$ dado en la ecuación (34).

6 Hay dos peluquerías situadas lado a lado, cada una atendida por un solo peluquero, en Dunkirk Square. En cada una pueden estar 4 clientes, como máximo; cualquier cliente potencial que encuentre una peluquería llena no espera el corte de cabello. El peluquero 1 cobra 11 dólares por corte de cabello y tarda un promedio de 12 minutos en terminar el corte. El peluquero 2 cobra 5 dólares por corte de cabello y tarda un promedio de 6 minutos en terminar su trabajo. Un promedio de 10 clientes potenciales por hora llega a cada peluquería. Naturalmente, un cliente potencial se vuelve un cliente real sólo si encuentra que la peluquería no está llena. Si suponemos que los tiempos entre llegadas y los tiempos para el corte de cabello son exponenciales, ¿qué peluquero ganará más dinero?

7 Seas Beginnings, una pequeña compañía de pedidos por correo tiene una línea telefónica. Un promedio de 60 personas por hora llama para hacer pedidos, y se necesita un mi-

FIGURA 15

	A	B	C	D	E	F	G
1	M/M/1/GD/c	LAMBDA?	MU?	c?	RO	PI(c)	TURNED AWAY
2		20	5	10	4	0.75000018	15.00000358
3		L	LS	LQ	W	WS	WQ
4		9.66666929	0.99999928	8.66667	1.93333524	0.2	1.733335241
5							
6							
7							
8							
9							
10							
11							
12	STATE	LAMBDA(J)	MU(J)	CJ	PROB	#IN QUEUE	COLA*COLE
13	0	20	0	1	7.1526E-07	0	0
14	1	20	5	4	2.861E-06	0	2.86102E-06
15	2	20	5	16	1.1444E-05	1	2.28882E-05
16	3	20	5	64	4.5776E-05	2	0.000137329
17	4	20	5	256	0.00018311	3	0.000732422
18	5	20	5	1024	0.00073242	4	0.00366211
19	6	20	5	4096	0.00292969	5	0.017578129
20	7	20	5	16384	0.01171875	6	0.08203127
21	8	20	5	65536	0.04687501	7	0.375000089
22	9	20	5	262144	0.18750004	8	1.687500402
23	10	0	5	1048576	0.75000018	9	7.500001788
24	11	0	5	0	0	10	0
25	12	0	5	0	0	11	0
26	13	0	5	0	0	12	0
27	14	0	5	0	0	13	0
28	15	0	5	0	0	14	0
29	16	0	5	0	0	15	0
30	17	0	5	0	0	16	0
31	18	0	5	0	0	17	0
32	19	0	5	0	0	18	0
33	20	0	5	0	0	19	0
34	21	0	5	0	0	20	0
35	22	0	5	0	0	21	0
36	23	0	5	0	0	22	0
37	24	0	5	0	0	23	0
38	25	0	5	0	0	24	0
39	26	0	5	0	0	25	0
40	27	0	5	0	0	26	0
41	28	0	5	0	0	27	0

nuto para atender una llamada. El tiempo entre llamadas y el tiempo para atender una llamada siguen una distribución exponencial. Si la línea está ocupada, Seas Beginnings puede poner hasta $c - 1$ en espera. Si $c - 1$ personas están en espera, una persona que llame escucha una señal de ocupado y llama a un competidor (Air End). Seas Beginnings desea que sólo 1% de todos los que llamen escuche una señal de ocupado. ¿A cuántas personas debe mantener en espera la compañía?

20.6 Sistema de colas $M/M/s/GD/\infty/\infty$

Consideremos ahora el sistema $M/M/s/GD/\infty/\infty$. Suponemos que los tiempos entre llegadas son exponenciales (con tasa λ), los tiempos de servicio son exponenciales (con tasa μ) y que hay sólo una cola de clientes que esperan ser atendidos en uno de los servidores en paralelo. Si están presentes $j \le s$ clientes, entonces los j clientes están en servicio; si $j > s$ clientes están presentes, entonces los s servidores están ocupados, y $j - s$ clientes están haciendo cola. Cualquier cliente que llegue y encuentre un servidor desocupado entra al servicio de inmediato, pero un cliente que llegue y no encuentre un servidor desocupado se une a la cola de clientes que esperan servicio. Los bancos y las oficinas de correos en los cuales todos los clientes hacen una sola cola en espera del servicio se pueden modelar como sistemas de líneas de espera $M/M/s/GD/\infty/\infty$.

Para describir el sistema $M/M/s/GD/\infty/\infty$ como un modelo de nacimiento-muerte, obsérvese que (como en el modelo $M/M/1/GD/\infty/\infty$ $\lambda_j = \lambda$ $(j = 0, 1, 2, \ldots)$. Si j servidores están ocupados, entonces los servicios terminan a una tasa de

$$\underbrace{\mu + \mu + \cdots}_{j\mu} = j\mu$$

Siempre que j clientes estén presentes, min (j, s) servidores estarán ocupados. Por lo tanto, $\mu_j = \min(j, s)\mu$. En resumen, encontramos que el sistema $M/M/s/GD/\infty/\infty$ se puede modelar como un proceso de nacimiento-muerte (véase figura 16) con parámetros

$$\begin{aligned} \lambda_j &= \lambda \qquad && (j = 0, 1, \ldots) \\ \mu_j &= j\mu \qquad && (j = 0, 1, \ldots, s) \\ \mu_j &= s\mu \qquad && (j = s+1, s+2, \ldots) \end{aligned} \tag{38}$$

definimos $\rho = \frac{\lambda}{s\mu}$. For $\rho < 1$, al sustituir (38) en (16) a (19) se obtienen las siguientes probabilidades de estado estable:

$$\pi_0 = \frac{1}{\displaystyle\sum_{i=0}^{i=(s-1)} \frac{(s\rho)^i}{i!} + \frac{(s\rho)^s}{s!(1-\rho)}} \tag{39}$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{j!} \qquad (j = 1, 2, \ldots, s) \tag{39.1}$$

$$\pi_j = \frac{(s\rho)^j \pi_0}{s!s^{j-s}} \qquad (j = s, s+1, s+2, \ldots) \tag{39.2}$$

Si $\rho \ge 1$, no existe estado estable. En otras palabras, si la tasa de llegadas es por lo menos de la misma magnitud que la tasa de servicio máxima posible ($\lambda \ge s\mu$), el sistema se "amplifica".

De acuerdo con (39.2), se puede demostrar que la probabilidad de estado estable de que todos los servidores estén ocupados está dada por

$$P(j \ge s) = \frac{(s\rho)^s \pi_0}{s!(1-\rho)} \tag{40}$$

FIGURA 16
Diagrama de tasas para el sistema de colas $M/M/s/GD/\infty/\infty$

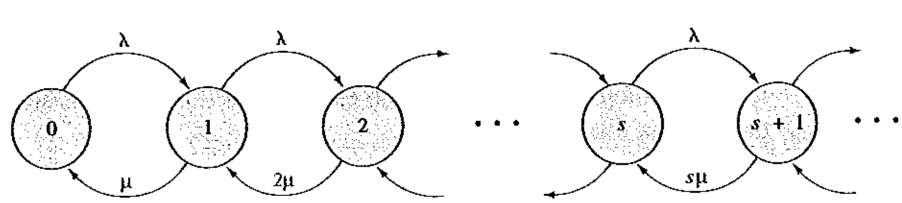


TABLA 6

$P(j \geq s)$ para el sistema de colas $M/M/s/GD/\infty/\infty$

ρ	$s = 2$	$s = 3$	$s = 4$	$s = 5$	$s = 6$	$s = 7$
.10	.02	.00	.00	.00	.00	.00
.20	.07	.02	.00	.00	.00	.00
.30	.14	.07	.04	.02	.01	.00
.40	.23	.14	.09	.06	.04	.03
.50	.33	.24	.17	.13	.10	.08
.55	.39	.29	.23	.18	.14	.11
.60	.45	.35	.29	.24	.20	.17
.65	.51	.42	.35	.30	.26	.21
.70	.57	.51	.43	.38	.34	.30
.75	.64	.57	.51	.46	.42	.39
.80	.71	.65	.60	.55	.52	.49
.85	.78	.73	.69	.65	.62	.60
.90	.85	.83	.79	.76	.74	.72
.95	.92	.91	.89	.88	.87	.85

En la tabla 6 se calcula $P(j \geq s)$ para diversas situaciones. También se puede demostrar que

$$L_q = \frac{P(j \geq s)\rho}{1 - \rho} \tag{41}$$

Entonces, con (28) se obtiene

$$W_q = \frac{L_q}{\lambda} = \frac{P(j \geq s)}{s\mu - \lambda} \tag{42}$$

Para determinar L (y luego W), se aplica el hecho de que $L = L_q + L_s$. Puesto que $W_s = \frac{1}{\mu}$, la ecuación (30) muestra que $L_s = \frac{\lambda}{\mu}$. Entonces,

$$L = L_q + \frac{\lambda}{\mu} \tag{43}$$

Asimismo,

$$\begin{aligned} W &= \frac{L}{\lambda} \\ &= \frac{L_q}{\lambda} + \frac{1}{\mu} \\ &= W_q + \frac{1}{\mu} \\ &= \frac{P(j \geq s)}{s\mu - \lambda} + \frac{1}{\mu} \end{aligned} \tag{44}$$

Cuando necesitamos determinar L, L_q, W o W_q, empezamos por buscar $P(j \geq s)$ en la tabla 6. Luego usamos (41) a (44) para estimar la cantidad que deseamos. Si nos interesa la distribución de probabilidad de estado estable, encontramos $P(j \geq s)$ en la tabla 6 y, luego, aplicamos (40) para determinar π_0. Después, mediante (39.1) y (39.2) se obtiene la distribución completa de estado estable. Los dos ejemplos siguientes ilustran la aplicación de las fórmulas anteriores.

EJEMPLO 1 Cajeros de bancos

Imagine un banco con dos cajeros. Un promedio de 80 clientes por hora llega al banco, y esperan en una sola cola que se desocupe algún cajero. El tiempo promedio que se requiere para atender a un cliente es 1.2 minutos. Suponga que los tiempos entre llegadas y los tiempos de servicio son exponenciales. Determine.

1 Número esperado de clientes presentes en el banco.

2 Tiempo esperado que un cliente pasa en el banco.

3 Fracción de tiempo que un cajero en particular está desocupado.

Solución **1** Tenemos un sistema $M/M/2/GD/\infty/\infty$ con $\lambda = 80$ clientes por hora y $\mu = 50$ clientes por hora. Por lo tanto $\rho = \frac{80}{2(50)} = 0.80 < 1$, de modo que sí existe un estado estable. (Para $\lambda \geq 100$ no existiría estado estable.) De acuerdo con la tabla 6, $P(j \geq 2) = .71$. Entonces, con (41) se tiene

$$L_q = \frac{.80(.71)}{1 - .80} = 2.84 \text{ clientes}$$

y de (43), $L = 2.84 + \frac{80}{50} = 4.44$ clientes.

2 Como $W = \frac{L}{\lambda}$, $W = \frac{4.44}{80} = 0.055$ hora $= 3.3$ minutos.

3 Para determinar la fracción de tiempo que está desocupado un cajero en particular, observe que está desocupado durante el tiempo total en que $j = 0$ y la mitad del tiempo (por simetría) que $j = 1$. La probabilidad de que un servidor esté desocupado está dada por $\pi_0 + 0.5\pi_1$. Apoyándonos en el hecho de que $P(j \geq 2) = .71$, obtenemos π_0 a partir de (40):

$$\pi_0 = \frac{s!P(j \geq s)(1 - \rho)}{(s\rho)^2} = \frac{2!(.71)(1 - .80)}{(1.6)^2} = .11$$

Ahora (39.1) da

$$\pi_1 = \frac{(1.6)^1 \pi_0}{1!} = .176$$

Por lo tanto, la probabilidad de que un cajero en particular esté desocupado es $\pi_0 + 0.5\pi_1 = .11 + 0.5(.176) = .198$. Podríamos haber encontrado π_0 directamente a partir de (39):

$$\pi_0 = \frac{1}{1 + \frac{[2(.80)]^1}{1!} + \frac{[2(.80)]^2}{2!(1 - .80)}} = \frac{1}{1 + 1.6 + 6.4} = \frac{1}{9}$$

Este valor es consistente con el cálculo efectuado de $\pi_0 = .11$.

EJEMPLO 10 Personal para un banco

El gerente de un banco debe determinar cuántos cajeros deben trabajar los viernes. Por cada minuto que un cliente permanece en la fila, el gerente opina que se genera un costo de 5 centavos por la demora. Dos clientes por minuto, en promedio, llegan al banco. Toma, en promedio, dos minutos completar la transacción de un cliente. Al banco le cuesta 9 dólares por hora contratar un cajero. Los tiempos entre llegadas y los tiempos de servicio son exponenciales. ¿Cuántos cajeros debe tener el banco trabajando los viernes con el fin de minimizar la suma de los costos de servicio y los costos de demora?

Solución Como $\lambda = 2$ clientes por minuto y $\mu = 0.5$ cliente por minuto, $\frac{\lambda}{s\mu} < 1$ requiere que $\frac{4}{s} < 1$ es decir, $s \geq 5$. Por lo tanto, debe haber por lo menos 5 cajeros, o la cantidad de clientes presentes se "amplificará". Ahora calculemos, para $s = 5, 6, \ldots,$

$$\frac{\text{Costo esperado de servicio}}{\text{Minuto}} + \frac{\text{costo esperado por demora}}{\text{minuto}}$$

Puesto que cada cajero recibe $\frac{9}{60} = 15$ centavos por minuto,

$$\frac{\text{Costo esperado del servicio}}{\text{Minuto}} = 0.15s$$

Al igual que en el ejemplo 4,

$$\frac{\text{Costo esperado por demora}}{\text{Minuto}} = \left(\frac{\text{clientes esperados}}{\text{minuto}}\right)\left(\frac{\text{costo esperado por demora}}{\text{cliente}}\right)$$

Pero

$$\frac{\text{Costo esperado por demora}}{\text{Cliente}} = 0.05W_q$$

Como llega un promedio de dos clientes por minuto,

$$\frac{\text{Costo esperado por demora}}{\text{Minuto}} = 2(0.05W_q) = 0.10W_q$$

Para $s = 5$, $\rho = \frac{2}{.5(5)} = .80$ y $P(j \geq 5) = .55$. De (42),

$$W_q = \frac{.55}{5(.5) - 2} = 1.1 \text{ minutos}$$

por lo tanto, para $s = 5$,

$$\frac{\text{Costo esperado por demora}}{\text{Minuto}} = 0.10(1.1) = 11¢$$

y para $s = 5$,

$$\frac{\text{Costo total esperado}}{\text{Minuto}} = 0.15(5) + 0.11 = 86¢$$

Como $s = 6$ tiene un costo de servicio por minuto de $6(0.15) = 90$ centavos, 6 cajeros no pueden tener un costo total inferior a 5 cajeros. Por lo tanto, lo óptimo es tener 5 cajeros en servicio. En otras palabras, si se pone otro cajero el banco puede ahorrar cuando mucho 11 centavos por minuto en costos por demora. Como otro cajero cuesta 15 centavos por minuto, no puede ser óptimo contratar más de 5 cajeros.

Además del tiempo previsto del cliente en el sistema, la distribución del tiempo de espera del cliente es importante. Por ejemplo, si todos los clientes que tienen que esperar más de 5 minutos en una caja del supermercado, deciden cambiarse a otra tienda, la probabilidad de que un cliente dado se cambie a otra tienda es igual a $P(\mathbf{W} > 5)$. Para determinar esta probabilidad, necesitamos conocer la distribución del tiempo de espera de un cliente. Para un sistema de colas $M/M/s/FCFS/\infty/\infty$, se puede demostrar que

$$P(\mathbf{W} > t) = e^{-\mu t}\left\{1 + P(j \geq s)\,\frac{1 - \exp\left[-\mu t(s - 1 - s\rho)\right]}{s - 1 - s\rho}\right\}^{\dagger} \tag{45}$$

$$P(\mathbf{W}_q > t) = P(j \geq s)\exp\left[-s\mu(1 - \rho)t\right] \tag{46}$$

Para ilustrar el uso de las ecuaciones (45) y (46), refiérase al ejemplo 7. Suponga que (para $s = 5$) el gerente del banco quiere saber la probabilidad de que un cliente tenga que esperar en la fila más de 10 minutos. Para $s = 5$, $\rho = .80$, $P(j \geq 5) = .55$ y $\mu = 0.5$ cliente por minuto, la ecuación (46) proporciona

$$P(\mathbf{W}_q > 10) = .55\exp\left[-5(0.5)(1 - .80)(10)\right] = .55\,e^{-5} = .004$$

†Si $s - 1 = s\rho$, entonces $P(\mathbf{W} > t) = e^{-\mu t}(1 + P(j \geq s)\mu t)$.

Por lo tanto, el gerente del banco puede tener la certeza que la probabilidad de que un cliente tenga que esperar más de 10 minutos es muy pequeña.

Hoja de cálculo para el sistema de colas *M/M/s/GD/∞/∞*

Multiple.xls

En la figura 17 (archivo Multiple.xls) se presenta un plantilla que se puede usar para calcular cantidades importantes para el sistema de colas $M/M/s/GD/\infty/\infty$. Escriba λ en la celda B2, μ en la celda C2, y en la celda D2, s. Calcule $P(j \geq s)$ en la celda B6. En el renglón 4 se calculan las cantidades L, L_s, L_q, W, W_s y W_q. Estime $P(\mathbf{W}_q > t)$ en A8 para el valor de la entrada t en la celda B8. En la celda C8 calcule $P(\mathbf{W} > t)$ para el valor de t en la celda B8. Las probabilidades de estado estable se calculan en la columna E (estamos suponiendo que hay una pequeña probabilidad de que más de 1 000 clientes estén presentes). Los datos del ejemplo 10 se proporcionan en la figura 17 (con cinco servidores).

Si contamos con una hoja de cálculo para determinar cantidades de interés para el sistema $M/M/s$ podremos aplicar técnicas para hoja de cálculo, como Tablas y Buscar objetivo para contestar preguntas de interés. Por ejemplo, reconsidere el ejemplo 10. Para determinar la cantidad de servidores que minimiza el costo esperado por minuto nos gustaría variar el número de servidores (empezar con cinco) y calcular el costo esperado por minuto para diferentes cantidades de servidores. Esto se efectúa con una **tabla de datos unidireccional.** (Véase figura 18.)

Paso 1 Escriba la cantidad posible de servidores (5 a 8) en las celdas J5:J8.

Paso 2 Escriba la fórmula del costo esperado por minuto en una columna a la derecha y un reglón por arriba de donde se listan las cantidades posibles de servidores. Esto es en la celda K4.

$$=0.15*D2+B2*G4*0.05$$

Paso 3 Seleccione *el intervalo de la tabla*. Éste comprende los valores, la fórmula calculada y el intervalo donde se sitúan los valores de la fórmula calculada. En el ejemplo, el intervalo de la tabla es J4:K8.

Paso 4 Elija Tabla de Datos y seleccione Tabla Unidireccional (porque modificamos sólo un valor, el número de servidores).

Paso 5 Llene el cuadro de diálogo como se indica en la figura 19. De esta manera se instruye a Excel a colocar en forma repetida los valores de entrada en la columna de la izquierda del intervalo de la tabla en la celda D2 (número de servidores) y vuelva a calcular la fórmula (costo esperado por minuto, el cual se introduce en la celda K4). Luego obtenga el costo esperado por minuto para 5 a 8 servidores. Al igual que la vez anterior, encontrará que cinco servidores consiguen el más bajo costo esperado por minuto.

El siguiente es otro ejemplo de cómo podemos usar herramientas efectivas de hoja de cálculo con el fin de contestar preguntas importantes acerca de las líneas de espera. Suponga que queremos saber (en el caso de cinco servidores) el nonagésimo percentil del tiempo de un cliente en el sistema. Es decir, deseamos saber el valor de t que hace $P(\mathbf{W} > t)$ igual a .10. Esto se calcula fácilmente con Buscar objetivo de Excel. Buscar objetivo permite encontrar qué valor de una celda (la celda *que cambia*) ocasiona que una fórmula en otra celda (la celda *definida*) asuma un valor deseado (llamada *con el valor*).

Para usar Buscar objetivo en la búsqueda del nonagésimo percentil del tiempo de un cliente en el sistema, seleccione Buscar objetivo en Herramientas, y se llena el cuadro de diálogo como se indica en la figura 20. Este cuadro de diálogo encuentra el valor para t en B8 que hace $P(\mathbf{W} > t)$ igual a .1 (calculado en C8). Encontramos que con cinco servidores, 10% de todos los clientes pasará por lo menos 6.7 min en el banco. (Véase figura 21.)

Observe que la precisión de Buscar objetivo se mejora si selecciona Opciones en Herramientas, luego elija Calcular, y en Cambio máximo escriba un número más pequeño que el que ya está señalado, que es de .001. Por ejemplo, un Cambio máximo de .000001 da la certeza que al terminar la operación de Buscar objetivo, $P(\mathbf{W} > q)$ estará dentro de .000001 de .10.

FIGURA 17

	A	B	C	D	E	F	G	H
1	M/M/s/GD	LAMBDA?	MU?	s?	RO			
2		2	0.5	5	0.8			
3		L	LS	LQ	W	WS	WQ	
4		6.21645022	4	2.21645022	3.10822511	1.999999999	1.108225109	
5	STATE	P(j>=s)						
6	1	0.55411255						
7	P(Wq>t)	t?	P(W>t)					
8	0.019390014	6.70521931	0.10000006					
9								
10								
11								
12	STATE	LAMBDA(J)	MU(J)	CJ	PROB	#IN QUEUE	COLA*COLE	COLE*COL
13	0	2	0	1	0.01298701	0	0	0
14	1	2	0.5	4	0.05194805	0	0.051948052	0
15	2	2	1	8	0.1038961	0	0.207792208	0
16	3	2	1.5	10.6666667	0.13852814	0	0.415584416	0
17	4	2	2	10.6666667	0.13852814	0	0.554112554	0
18	5	2	2.5	8.53333333	0.11082251	0	0.554112554	0
19	6	2	2.5	6.82666667	0.08865801	1	0.531948052	0.08865801
20	7	2	2.5	5.46133333	0.07092641	2	0.496484848	0.14185281
21	8	2	2.5	4.36906667	0.05674113	3	0.453929004	0.17022338
22	9	2	2.5	3.49525333	0.0453929	4	0.408536104	0.1815716
23	10	2	2.5	2.79620267	0.03631432	5	0.363143203	0.1815716
24	11	2	2.5	2.23696213	0.02905146	6	0.319566019	0.17430874
25	12	2	2.5	1.78956971	0.02324117	7	0.27889398	0.16268816
26	13	2	2.5	1.43165577	0.01859293	8	0.241708116	0.14874346
27	14	2	2.5	1.14532461	0.01487435	9	0.208240839	0.13386911
28	15	2	2.5	0.91625969	0.01189948	10	0.178492147	0.11899476
29	16	2	2.5	0.73300775	0.00951958	11	0.152313299	0.10471539
30	17	2	2.5	0.5864062	0.00761566	12	0.129466304	0.09138798
31	18	2	2.5	0.46912496	0.00609253	13	0.109665575	0.07920292
32	19	2	2.5	0.37529997	0.00487403	14	0.092606486	0.06823636
33	20	2	2.5	0.30023998	0.00389922	15	0.077984409	0.05848831
34	21	2	2.5	0.24019198	0.00311938	16	0.065506904	0.04991002
35	22	2	2.5	0.19215358	0.0024955	17	0.054901024	0.04242352
36	23	2	2.5	0.15372287	0.0019964	18	0.04591722	0.03593522
37	24	2	2.5	0.12297829	0.00159712	19	0.038330897	0.03034529
38	25	2	2.5	0.09838264	0.0012777	20	0.031942414	0.02555393
39	26	2	2.5	0.07870611	0.00102216	21	0.026576088	0.0214653
40	27	2	2.5	0.06296489	0.00081773	22	0.022078597	0.01798997
41	28	2	2.5	0.05037191	0.00065418	23	0.018317058	0.01504615
42	29	2	2.5	0.04029753	0.00052334	24	0.015176991	0.01256027
43	30	2	2.5	0.03223802	0.00041868	25	0.012560268	0.01046689
44	31	2	2.5	0.02579042	0.00033494	26	0.010383155	0.00870845
45	32	2	2.5	0.02063233	0.00026795	27	0.008574476	0.00723471
46	33	2	2.5	0.01650587	0.00021436	28	0.007073943	0.00600213
47	34	2	2.5	0.01320469	0.00017149	29	0.005830644	0.0049732
48	35	2	2.5	0.01056376	0.00013719	30	0.004801707	0.00411575
49	36	2	2.5	0.008451	0.00010975	31	0.003951119	0.00340235
50	37	2	2.5	0.0067608	8.7803E-05	32	0.003248698	0.00280968
51	38	2	2.5	0.00540864	7.0242E-05	33	0.0026692	0.00231799

	J	K
2		
3		
4	Servidores	0.86082251
5	5	0.86082251
6	6	0.92847608
7	7	1.05900734
8	8	1.2029522

FIGURA **18**

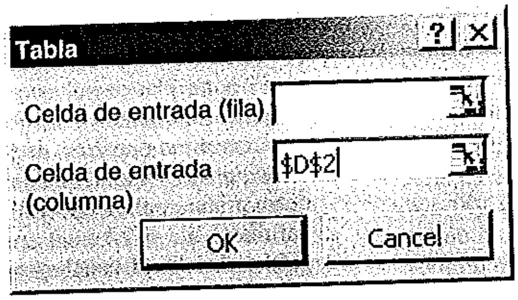


FIGURA **19**

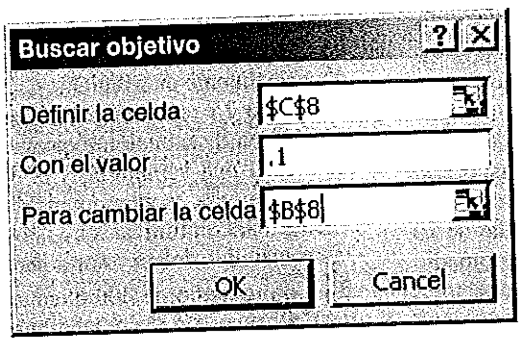


FIGURA **20**

	A	B	C
7	P(Wq>t)	t?	P(W>t)
8	0.019390014	6.70521931	0.10000006

FIGURA **21**

Aplicación de LINGO para los cálculos de $M/M/s/GD/\infty/\infty$

La función **@PEB**( ) de LINGO obtiene la probabilidad de que todos los servidores estén ocupados ($P(j \geq s)$) para un sistema $M/M/s/GD/\infty/\infty$. La función **@PEB** tiene dos argumentos: el primero es el valor de λ/μ, y el segundo es el número de servidores. Por lo tanto, en el ejemplo 9, **@PEB**(80/50,2) $=$ .711111 da $P(j \geq 2)$.

Es posible usar la función **@PEB** para resolver problemas de optimización de líneas de espera con LINGO. Por ejemplo, para determinar el número que minimiza el costo de los servidores en el ejemplo 10 introduciríamos el problema siguiente en LINGO:

```
MODEL:
1) MIN=.10*@PEB(4,S)/(.5*S-2) + .15*S;
2) S>5;
END
```

En el renglón 1 .10*@PEB(4,S)/(.5*S−2) es el costo esperado por minuto debido a la espera de los clientes, en tanto que .15*S es el costo de servicio por minuto. Se infiere el renglón 2, porque necesitamos por lo menos cinco servidores para que exista un estado estable. LINGO da por resultado $S = 5$ con un valor de la función objetivo de .860823 (éste es el costo esperado por minuto).

PROBLEMAS

Grupo A

1 Un supermercado pretende decidir cuántas cajas registradoras mantener abiertas. Suponga que un promedio de 18 clientes llega cada hora, y que el promedio de tiempo en que un cliente salda su cuenta es de 4 minutos. Los tiempos entre llegadas y los tiempos de servicio son exponenciales; además, el sistema se podría modelar como un sistema de líneas de espera $M/M/s/GD/\infty/\infty$. Cuesta 20 dólares por hora operar una caja registradora, y se fija un costo de 25 centavos por cada minuto que el cliente pasa en el área de las cajas registradoras. ¿Cuántas cajas debe mantener abiertas la tienda?

2 Un pequeño banco trata de determinar cuántos cajeros contratar. El costo total por emplear a un cajero es 100 dólares por día, y un cajero puede atender un promedio de 60 clientes por día. Un promedio de 50 clientes llega en promedio al día al banco. Los tiempos entre llegadas y los tiempos de servicio son exponenciales. Si el costo por demora por día-cliente es 100 dólares, ¿cuántos cajeros debe contratar el banco?

3 Todos los tiempos entre llegadas y de servicio son exponenciales en este problema.

a En este momento, tanto el departamento de finanzas como el departamento de comercialización tienen una mecanógrafa. Cada mecanógrafa puede escribir 25 cartas por día. Finanzas requiere que se escriban un promedio de 20 cartas por día, y comercialización necesita un promedio de 15 cartas al día. Determine para cada departamento el tiempo promedio que transcurre entre el momento en que se solicita una carta y la terminación de la misma.

b Suponga que las dos mecanógrafas están mancomunadas; es decir, cada mecanógrafa estaría disponible para hacer las cartas de cualquiera de los dos departamentos. De acuerdo con este arreglo, calcule el tiempo promedio que transcurre entre el momento en que se pide una carta y la terminación de la misma.

c Comente los resultados de los incisos (a) y (b).

d Según el arreglo mancomunado, ¿cuál es la probabilidad de que más de .200 de día transcurra entre la petición de elaborar una carta y la terminación de la misma?

4 MacBurger desea determinar cuántos servidores (o colas) deben estar disponibles durante el turno del desayuno. Un promedio de 100 clientes llega durante una hora al restaurante. Cada cola o servidor puede atender un promedio de 50 clientes por hora. Un servidor cuesta 5 dólares por hora, y el costo de un cliente que espera en la fila durante una hora es 20 dólares. Suponga que es aplicable un modelo $M/M/s/GD/\infty/\infty$, y determine el número de colas que minimizan la suma de los costos de demora y de servicio.

5 Un promedio de 100 clientes llega al banco de Gotham City. El tiempo promedio de servicio por cada cliente es un minuto. Los tiempos de servicio y los tiempos entre llegadas son exponenciales. El gerente quiere asegurarse que no más de 1% de todos los clientes tendrá que esperar en la cola durante más de 5 minutos. Si el banco sigue la estrategia de formar a todos los clientes en una sola cola, ¿cuántos cajeros debe contratar el banco?

6 Un promedio de 90 huéspedes por hora llega al vestíbulo de un hotel (los tiempos entre llegadas son exponenciales) y esperan para registrarse. En este momento hay cinco empleados, y los huéspedes están formados en una sola cola, esperando que esté disponible un empleado. Un empleado tarda un tiempo promedio de 3 minutos (distribuido exponencialmente) en atender a un huésped. Los empleados ganan 10 dólares por hora, y el hotel fija un costo de tiempo de espera de 20 dólares por cada hora que un huésped espera en la cola.

a Calcule el costo esperado por hora del sistema actual.

b El hotel planea poner en lugar de un empleado una máquina automática para registrar a los huéspedes. El gerente estima que 20% de todos los huéspedes utilizará la máquina. Una máquina de este tipo tarda un promedio de un minuto para atender al huésped. Cuesta 48 dólares por día (1 día = 8 horas) operar una de estas máquinas. ¿El hotel debería instalar la máquina? Suponga que todos los clientes que están dispuestos a usar la máquina hacen una sola fila.

7 Un promedio de 50 clientes por hora llega a una pequeña oficina de correos. Los tiempos entre llegadas están exponencialmente distribuidos. Cada ventanilla atiende un promedio de 25 clientes por hora. Los tiempos de servicio siguen una distribución exponencial. Cuesta 25 dólares por hora abrir una ventanilla, y la oficina valora el tiempo que un cliente pasa esperando en la fila en 15 dólares por hora-cliente. Para minimizar los costos esperados por hora, ¿cuántas ventanillas deben estar abiertas?

8 Un promedio de 300 clientes llega a una enorme sucursal del banco 2. Se requiere un promedio de 2 minutos para atender a cada cliente. Cuesta 10 dólares por hora mantener abierta una ventanilla de cajero, por lo que el banco estima que perderá 50 dólares en ganancias futuras por cada hora que un cliente espera en la cola. ¿Cuántas ventanillas de cajeros debe abrir el banco 2?

9 Un promedio de 40 estudiantes por hora llega al laboratorio de computación de la Maestría de Administración de Empresas. El estudiante promedio utiliza una computadora durante 20 minutos. Suponga tiempos entre llegadas y de servicio exponenciales.

a Si queremos que, cuando mucho, el tiempo promedio que un estudiante espere una computadora sea 10 minutos, ¿cuántas computadoras debe tener el laboratorio?

b Si queremos que 95% de todos los estudiantes espere durante 5 minutos o menos una computadora, ¿cuántas computadoras debe tener el laboratorio?

10 Un sistema de almacenamiento de datos consta de 3 unidades de disco que comparte una cola común. Llega un promedio de 50 peticiones de almacenamiento por segundo. El tiempo promedio necesario para atender una solicitud es de .03 s. Suponga que los tiempos entre llegadas y los tiempos de servicio son exponenciales, y determine

a La probabilidad de que una unidad de disco esté ocupada.

b La probabilidad de que ninguna unidad de disco esté ocupada.

c La probabilidad de que un trabajo tenga que esperar.

d La cantidad promedio de trabajos presente en el sistema de almacenamiento.

11 Un contador de boletos de la Northwest Airlines pronostica que necesitarán registrarse 200 personas por hora. Se requiere un promedio de 2 minutos atender a un cliente. Suponga que los tiempos entre llegadas y los tiempos de servicio son exponenciales, y que todos los clientes hacen una sola fila, esperando al primer agente disponible.

a Si queremos que el tiempo promedio que un cliente pasa en la cola y en el servicio sea de 30 minutos o menos, ¿cuántos empleados deben estar disponibles?

b Si queremos que 95% de todos los clientes esperen 45 minutos o menos, ¿cuántos empleados debe haber disponibles?

Grupo B

12 Un promedio de 100 clientes llega por hora al banco de Gotham City. Un cajero tarda un promedio de 2 minutos en atender a un cliente. Los tiempos de servicio y entre llegadas son exponenciales. En la actualidad, cuatro cajeros trabajan en el banco. El gerente del banco desea comparar los dos sistemas siguientes respecto al número promedio de clientes presentes en el banco y la probabilidad de que un cliente pase más de 8 minutos en el banco:

Sistema 1 Cada cajero tiene su propia cola, y no se permite cambiarse de cola.

Sistema 2 Todos los clientes se forman en una sola cola y esperan al primer cajero disponible.

Si usted fuera el gerente del banco, ¿qué sistema preferiría?

13 Un taller de instalación de mofles tiene tres mecánicos. Cada mecánico tarda un promedio de 45 minutos en instalar un nuevo mofle. Suponga que llega un promedio de un cliente por hora. ¿Cuál es el número esperado de mecánicos que están ocupados en cualquier momento dado? Dé respuesta a la pregunta sin suponer que los tiempos de servicio y los tiempos entre llegadas son exponenciales.

14 Considere los dos sistemas siguientes de líneas de espera:

Sistema 1 Un sistema $M/M/1$ con tasa de llegadas λ y tasa de servicio 3μ.

Sistema 2 Un sistema $M/M/3$ con tasa de llegadas λ y en el que cada servidor trabaja a una tasa μ.

Sin hacer cálculos extensos, ¿cuál sistema tendrá la W y L más pequeñas? (*Sugerencia*: exprese los parámetros de nacimiento-muerte de cada sistema. Luego determine cuál es más efectivo).

15 (Requiere usar una hoja de cálculo de LINGO.) La planta Carco ubicada en Bedford fabrica limpiadores para parabrisas para automóviles Ford. En un día específico, cada máquina de la planta es capaz de producir 1 000 limpiadores. La planta trabaja 250 días al año, y Ford necesita 3 millones de limpiadores al año. Cuesta 50 000 dólares por año operar una máquina. Por cada día que un limpiador es retrasado se genera un costo de 100 dólares (debido a un periodo de paralización en la producción de otras plantas). ¿Cuántas máquinas debe tener la planta de Ford? Suponga que los tiempos entre llegadas y los tiempos de servicio son exponenciales.

20.7 Modelos $M/G/\infty/GD/\infty/\infty$ y $GI/G/\infty/GD/\infty/\infty$

Hay muchos ejemplos de sistemas en los cuales un cliente nunca tiene que esperar para que comience el servicio. En dichos sistemas, se podría pensar que la permanencia completa del cliente en el sistema es el tiempo de servicio. Como un cliente nunca tiene que esperar el servicio, hay, en esencia, un servidor disponible para cada llegada, y podríamos pensar que éste es un sistema **con servidores infinitos** (o atiéndase usted mismo). Dos ejemplos de un sistema con servidores infinitos se presentan en la tabla 7.

Un sistema con servidores infinitos, en el cual los tiempos entre llegadas y de servicio pueden apegarse a una distribución arbitraria de probabilidad, se podrían expresar como sistemas de colas $GI/G/\infty/GD/\infty/\infty$, aplicando la notación de Kendall-Lee. Dicho sistema opera como sigue:

1 Los tiempos entre llegadas son iid con distribución común **A**. Definimos $E(\mathbf{A}) = \frac{1}{\lambda}$. Por lo tanto, λ es la tasa de llegadas.

2 Cuando un cliente llega, entra inmediatamente al servicio. El tiempo de cada cliente en el sistema está regido por una distribución S que tiene $E(\mathbf{S}) = \frac{1}{\mu}$.

TABLA 7
Ejemplos de sistemas de líneas de espera con servidores infinitos

Situación	Llegadas	Tiempo de servicio (tiempo en el sistema)	Estado del sistema
Industria	La compañía entra a la industria	Tiempo hasta que la firma deja a la industria	Número de compañías en la industria
Programas universitarios	Los estudiantes entran a un programa	Tiempo en que el estudiante permanece en el programa	Número de estudiantes en el programa

Sea L el número esperado de clientes en el sistema en el estado estable, y W el tiempo previsto que un cliente pasa en el sistema. Por definición, $W = \frac{1}{\mu}$. Entonces, de la ecuación (30) se infiere que

$$L = \frac{\lambda}{\mu} \tag{47}$$

La ecuación (47) no requiere ninguna suposición de exponencialidad. Si los tiempos entre llegadas son exponenciales, se puede demostrar que (incluso para una distribución de tiempo de servicio arbitraria) la probabilidad de estado estable de que j clientes estén presentes (llámela π_j) sigue una distribución Poisson con media $\frac{\lambda}{\mu}$. De donde se infiere que

$$\pi_j = \frac{(\frac{\lambda}{\mu})^j e^{-\lambda/\mu}}{j!}$$

El ejemplo siguiente ilustra la aplicación característica de un sistema $GI/G/\infty/GD/\infty/\infty$.

EJEMPLO 11 Helados en Smalltown

Durante cada año, un promedio de tres heladerías o fuentes de soda abre sus puertas en Smalltown. El promedio de tiempo en que una fuente de sodas permanece en el negocio es 10 años. ¿Cuál es el número promedio de heladerías que se encontraría en Smalltown el primero de enero de 2525? Si el tiempo entre la apertura de las heladerías es exponencial, ¿cuál es la probabilidad de que el primero de enero de 2525 haya 25 heladerías en Smalltown?

Solución Sabemos que $\lambda = 3$ tiendas por año y que $\frac{1}{\mu} = 10$ años por heladería. Si suponemos que el estado estable ya se alcanzó, habrá un promedio de $L = \lambda(\frac{1}{\mu}) = 3(10) = 30$ heladerías en Smalltown. Si los tiempos entre llegadas de las heladerías son exponenciales, entonces,

$$\pi_{25} = \frac{(30)^{25}e^{-30}}{25!} = .05$$

Naturalmente, también podríamos calcular la probabilidad de que haya 25 heladerías con la fórmula de Excel

=POISSON(30,25,0)

Esta fórmula da .045.

PROBLEMAS

Grupo A

1 El club Columbus Record atrae cada semana 100 nuevos miembros. Los miembros permanecen activos durante un promedio de un año (1 año = 52 semanas). En promedio, ¿cuántos miembros tendrá el club de discos?

2 El programa para obtener el grado de doctor de la universidad estatal en el área de negocios, admite un promedio de 25 estudiantes de doctorado cada año. Si un estudiante de doctorado pasa un promedio de cuatro años como residente en la universidad, ¿cuántos estudiantes de doctorado esperaría uno encontrar ahí?

3 Hay en la actualidad 40 empresas en el estado de Indiana, que construyen casas en las que se utiliza la energía solar. Un promedio de 20 empresas de este tipo abre sus puertas cada año en ese estado. La compañía promedio permanece en el negocio durante 10 años. Si la tendencia actual continúa, ¿cuál es la cantidad esperada de compañías de este tipo que se encontrará en Indiana? Si el tiempo entre la entrada de las compañías a la industria se distribuye en forma exponencial, ¿cuál es la probabilidad de que (en el estado estable) haya más de 300 compañías relacionadas con la energía solar en el negocio? (*Sugerencia*: para una λ grande, la distribución Poisson se puede aproximar mediante una distribución normal).

20.8 Sistema de líneas de espera $M/G/1/GD/\infty/\infty$

En esta sección se estudia un sistema de líneas de espera con un solo servidor en el que los tiempos entre llegadas son exponenciales, pero la distribución de los tiempos de servicio (S) no requiere ser exponencial. Sea λ la tasa de llegadas (se supone que se mide en llegadas por hora). También definamos $\frac{1}{\mu} = E(S)$ y $\sigma^2 = \text{var } S$.

De acuerdo con la notación de Kendall, un sistema de líneas de espera de ese tipo se representa como un sistema de colas $M/G/1/GD/\infty/\infty$. Un sistema $M/G/1/GD/\infty/\infty$ *no* es un proceso de nacimiento-muerte porque la probabilidad de que se complete el servicio entre t y $t + \Delta t$ cuando el estado del sistema en el tiempo t es j depende del tiempo que ha transcurrido desde que se completó la última servicio (porque los tiempos de servicio ya no tienen la propiedad de carencia de memoria). Por lo tanto, no podemos expresar la probabilidad de que se complete un servicio entre t y $t + \Delta t$ en la forma $\mu\Delta t$, por lo que un modelo de nacimiento-muerte no es apropiado.

La determinación de las probabilidades de estado estable para un sistema de líneas de espera $M/G/1/GD/\infty/\infty$ es una cuestión difícil. Como ya no son válidas las ecuaciones de estado estable para el proceso de nacimiento-muerte, se debe emplear un método distinto. Se usa la teoría de las cadenas de Markov para determinar π_i', la probabilidad de que después que el sistema ha operado por largo tiempo, i clientes estarán presentes en el instante inmediatamente después que se complete un servicio (véase problema 5 al final de esta sección). Se puede demostrar que $\pi_i' = \pi_i$, donde π_i es la fracción de tiempo después que el sistema ha operado por mucho tiempo en que i clientes están presentes (véase Kleinrock (1975)).

Por fortuna, si utilizamos los resultados de Pollaczek y Khinchin, podemos determinar L_q, L, L_s, W_q, W y W_s . Pollaczek y Khinchin demostraron que para un sistema de colas $M/G/1/GD/\infty/\infty$

$$L_q = \frac{\lambda^2\sigma^2 + \rho^2}{2(1 - \rho)} \tag{48}$$

donde $\rho = \frac{\lambda}{\mu}$. Como $W_s = \frac{1}{\mu}$, de (30) se infiere que $L_s = \lambda(\frac{1}{\mu}) = \rho$. Como $L = L_s + L_q$, obtenemos

$$L = L_q + \rho \tag{49}$$

De (29) y (28) se infiere que

$$W_q = \frac{L_q}{\lambda} \tag{50}$$

$$W = W_q + \frac{1}{\mu} \tag{51}$$

Se puede demostrar, asimismo, que π_0, la fracción del tiempo que el servidor está desocupado, es $1 - \rho$. (Véase problema 2 al final de esta sección.) Este resultado es similar al del sistema $M/M/1/GD/\infty/\infty$.

Con el fin de ilustrar el uso de (48) a (51), imagine un sistema $M/M/1/GD/\infty/\infty$ con $\lambda = 5$ clientes por hora y $\mu = 8$ clientes por hora. De acuerdo con lo que sabemos de $M/M/1/GD/\infty/\infty$,

$$L = \frac{\lambda}{\mu - \lambda} = \frac{5}{8 - 5} = \frac{5}{3} \text{ clientes}$$

$$L_q = L - \rho = \frac{5}{3} - \frac{5}{8} = \frac{25}{24} \text{ clientes}$$

$$W = \frac{L}{\lambda} = \frac{\frac{5}{3}}{5} = \frac{1}{3} \text{ hora}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{25}{24}}{5} = \frac{5}{24} \text{ hora}$$

Según (3) y (4), sabemos que $E(S) = \frac{1}{8}$ hora y var $S = \frac{1}{64}$ hora^2. Entonces, con (48) se obtiene

$$L_q = \frac{\frac{(5)^2}{64} + \left(\frac{5}{8}\right)^2}{2\left(1 - \frac{5}{8}\right)} = \frac{25}{24} \text{ clientes}$$

$$L = L_q + \rho = \frac{25}{24} + \frac{5}{8} = \frac{40}{24} = \frac{5}{3} \text{ clientes}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{25}{24}}{5} = \frac{5}{24} \text{ hora}$$

$$W = \frac{L}{\lambda} = \frac{\frac{5}{3}}{5} = \frac{1}{3} \text{ hora}$$

Para demostrar cómo la varianza del tiempo de servicio puede afectar de manera importante la efectividad de un sistema de colas, considere un sistema de colas $M/D/1/GD/\infty/\infty$ cuyas λ y μ son idénticas a las del sistema $M/M/1/GD/\infty/\infty$ que apenas tratamos. Por lo que toca a este modelo $M/D/1/GD/\infty/\infty$, $E(S) = \frac{1}{8}$ hora y var $S = 0$. Entonces,

$$L_q = \frac{\left(\frac{5}{8}\right)^2}{2\left(1 - \frac{5}{8}\right)} = \frac{25}{48} \text{ clientes}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{25}{48}}{5} = \frac{5}{48} \text{ hora}$$

En este sistema $M/D/1/GD/\infty/\infty$, un cliente representativo pasará sólo la mitad del tiempo en la cola como en un sistema de colas $M/M/1/GD/\infty/\infty$ con tasas de llegada y de servicio idénticas. Como lo señala este ejemplo, aún cuando los tiempos medios de servicio no disminuyan, al disminuir la variabilidad de los tiempos de servicio, se reduce de manera sustancial la dimensión de la cola y el tiempo de espera del cliente.

PROBLEMAS

Grupo A

1 Un promedio de 20 automóviles por hora llega a la ventanilla de "servicio en su auto" de un restaurante de bocadillos. Si el tiempo de servicio para cada automóvil es de 2 minutos, ¿cuántos automóviles (en promedio) estarán formados en la cola? Suponga tiempos entre llegadas exponenciales.

2 Aplique el hecho de que $L_s = \frac{\lambda}{\mu}$ para demostrar que, para un sistema de colas $M/G/1/GD/\infty/\infty$, la probabilidad de que el servidor está ocupado es $\rho = \frac{\lambda}{\mu}$.

3 Un promedio de 40 automóviles por hora llega a un taller de pintura de un solo servidor GM para ser pintado. El 95% de los automóviles requiere un minuto para que lo pinte; el 5% se pinta dos veces y el trabajo requiere 2.5 minutos. Suponga que los tiempos entre llegadas son exponenciales.

a ¿Cuánto tiempo espera, en promedio, un automóvil antes de que lo pinten?

b Si los automóviles nunca tuvieran que ser repintados, ¿qué tanto se modificaría la respuesta del inciso (a)?

Grupo B

4 Considere un sistema de colas $M/G/1/GD/\infty/\infty$ en el cual se presenta un promedio de 10 llegadas cada hora. Suponga que cada tiempo de servicio al cliente sigue la distribución Erlang, con parámetro de proporcionalidad de un cliente por minuto y parámetro de forma de cuatro.

a Encuentre el número previsto de clientes que esperan en la cola.

b Determine el tiempo previsto que un cliente pasará en el sistema.

c ¿Qué fracción del tiempo estará desocupado el servidor?

5 Imagine un sistema de colas $M/G/1/GD/\infty/\infty$ en el cual los tiempos entre llegadas se distribuyen exponencialmente con parámetro λ y los tiempos de servicio tienen una función de densidad de probabilidad $s(t)$. Sea X_i el número de clientes presentes en el instante después de que el i-ésimo cliente completa su servicio.

a Explique por qué $X_1, X_2, \ldots, X_k, \ldots$ es una cadena de Markov.

b Explique por qué $P_{ij} = P(X_{k+1} = j|X_k = i)$ es cero para $j < i - 1$.

c Dé la razón de que $i > 0$, $P_{i,i-1}$ = (probabilidad de que ninguna llegada ocurra durante un tiempo de servicio); P_{ii} = (probabilidad de que ocurra una llegada durante un tiempo de servicio); y que para $j \geq i$, P_{ij} = (probabilidad de que $j - i + 1$ llegadas ocurran durante un tiempo de servicio.

d Explique por qué, para $j \geq i - 1$ y $i > 0$,

$$P_{ij} = \int_0^\infty \frac{s(x)e^{-\lambda x}(\lambda x)^{j-i+1}}{(j-i+1)!}\,dx$$

Sugerencia: la probabilidad que un tiempo de servicio esté entre $x + \Delta x$ es $\Delta x s(x)$. Dado que el tiempo de servicio es igual a x, la probabilidad de que $j - i + 1$ llegadas ocurran durante el tiempo de servicio es

$$\frac{e^{-\lambda x}(\lambda x)^{j-i+1}}{(j-i+1)!}$$

20.9 Modelos de origen finito: modelo de reparación de máquinas

Con excepción del modelo $M/M/1/GD/c/\infty$, todos los modelos estudiados muestran tasas de llegada independientes del estado del sistema. Como ya se estudió, hay dos situaciones en las que la suposición de tasas de llegada de estado independientes podría ser inválida.

1 Si los clientes no desean formar largas colas, la tasa de llegadas podría ser una función decreciente de la cantidad de personas presentes en el sistema de líneas de espera. Véanse los problemas 4 y 5 al final de la sección, ya que ejemplifican esta situación.

2 Si las llegadas a un sistema son extraídos de una población pequeña, la tasa de llegadas podría depender en gran medida del estado del sistema. Por ejemplo, si un banco sólo tiene 10 depositantes, entonces en un instante en que todos los depositantes están en el banco, la tasa de llegada debe ser cero, pero que si menos de 10 personas están en el banco, la tasa de llegadas será positiva.

Los modelos en los cuales las llegadas son extraídas de una población pequeña se denominan **modelos de origen finito**. Enseguida se estudia un modelo de origen finito conocido como *modelo de reparación de máquinas* (o modelo de *interferencia de las máquinas*).

En el problema de la reparación de máquinas, el sistema consiste en K máquinas y R personas que reparan máquinas o reparadores. En cualquier momento, una máquina específica está en buenas o malas condiciones. El tiempo en que una máquina permanece en buenas condiciones sigue una distribución exponencial con tasa λ. Siempre que una máquina se descompone, es enviada a un centro de reparación que consta de R personas encargadas de reparar las máquinas. El centro de reparaciones da servicio a las máquinas descompuestas como si llegaran a un sistema $M/M/R/GD/\infty/\infty$.

Por lo tanto, si $j \leq R$ máquinas están en malas condiciones, entonces, una máquina que apenas se descompuso será asignada de inmediato a reparación; si $j > R$ máquinas se descomponen, $j - R$ máquinas estarán esperando en una sola cola a que un trabajador se desocupe para que las repare. Se supone que el tiempo que se requiere para completar la reparación de una máquina es exponencial con una tasa μ (el tiempo medio de reparación es $\frac{1}{\mu}$). Tras reparar una máquina, ésta vuelve a estar en buenas condiciones y nuevamente es susceptible de descomponerse. Este modelo de la reparación de máquinas se podría modelar como un proceso de nacimiento muerte, donde el estado j en cualquier momento es la cantidad de máquinas en malas condiciones. Mediante la notación de Kendall-Lee, el modelo descrito se podría expresar como un modelo $M/M/R/GD/K/K$. La primera K indica que podrían estar presentes, en cualquier momento, no más de K clientes (o máquinas), y la segunda K significa que las llegadas tienen un origen finito de tamaño K.

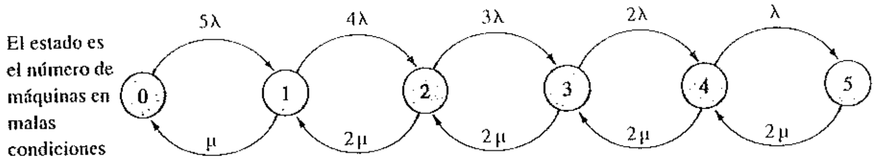
En la tabla 8 se muestra la interpretación de cada estado para un modelo de reparación de máquinas que tiene $K = 5$ y $R = 2$ (G = máquina en buenas condiciones; B = má-

TABLA 8

Estados posibles en el problema de reparación de máquinas cuando $K = 5$ y $R = 2$

Estado	Número de máquinas buenas	Cola de reparación	Número de reparadores ocupados
0	$G\ G\ G\ G\ G$		0
1	$G\ G\ G\ G$		1
2	$G\ G\ G$		2
3	$G\ G$	B	2
4	G	$B\ B$	2
5		$B\ B\ B$	2

FIGURA 22
Diagrama de tasas para un sistema de líneas de espera $M/M/R/GD/K/K$ cuando $R = 2$, $K = 5$



quina descompuesta). Para determinar los parámetros de nacimiento-muerte para el modelo de reparación de máquinas (véase figura 22), observe que un nacimiento corresponde a una descompostura de una máquina, y una muerte corresponde a una máquina que apenas fue reparada. Para deducir la tasa de nacimientos en el estado j, debemos determinar la tasa a la cual las máquinas se descomponen cuando el estado del sistema es j. Cuando el sistema es j, hay $K - j$ máquinas en buenas condiciones. Como cada máquina se descompone a una tasa λ, la tasa total a la cual se presentan las descomposturas cuando el estado es j es

$$\lambda_j = \underbrace{\lambda + \lambda + \cdots + \lambda}_{(K-j)\lambda\text{'s}} = (K - j)\lambda$$

Para determinar la tasa de muertes para el modelo de reparación de máquinas, procedemos como lo hicimos en el análisis del modelo $M/M/s/GD/\infty/\infty$. Cuando el estado es j, min (j, R) reparadores estarán ocupados. Puesto que cada reparador ocupado termina la compostura a una tasa μ, la tasa de muertes μ_j está dada por

$$\mu_j = j\mu \qquad (j = 0, 1, \ldots, R)$$
$$\mu_j = R\mu \qquad (j = R + 1, R + 2, \ldots, K)$$

Si definimos $\rho = \frac{\lambda}{\mu}$, al aplicar (16) a (18) se obtiene la distribución de probabilidades de estado estable siguiente:

$$\begin{aligned}\pi_j &= \binom{K}{j}\rho^j\pi_0 \qquad (j = 0, 1, \ldots, R) \\ &= \frac{\binom{K}{j}\rho^j j!\,\pi_0}{R!R^{j-R}} \qquad (j = R + 1, R + 2, \ldots, K)\end{aligned} \tag{52}$$

En (52)

$$\binom{K}{j} = \frac{K!}{j!(K - j)!}$$

donde $0! = 1$, y para $n \geq 1$, $n! = n(n-1)\cdots(2)(1)$. Para aplicar (52), empezamos por determinar π_0 a partir del hecho de que $\pi_0 + \pi_1 + \cdots + \pi_k = 1$. Al utilizar las probabilidades de estado estable en (52) ya es posible estimar las cantidades siguientes que interesan:

L = Cantidad esperada de máquinas descompuestas.

L_q = Cantidad esperada de máquinas que esperan servicio.

W = Tiempo promedio que una máquina pasa descompuesta (tiempo de paralización).

W_q = Tiempo promedio que una máquina pasa esperando servicio.

No hay, infortunadamente, fórmulas sencillas para L, L_q, W y W_q. Lo mejor que podemos hacer es expresar estas cantidades en términos de las π_j:

$$L = \sum_{j=0}^{j=K} j\pi_j \tag{53}$$

$$L_q = \sum_{j=R}^{j=K} (j-R)\pi_j \tag{54}$$

Enseguida utilizamos (28) y (29) para encontrar W y W_q. Como la tasa de llegadas depende del estado, la cantidad promedio de llegadas por unidad de tiempo se determina mediante $\bar{\lambda}$, donde

$$\bar{\lambda} = \sum_{j=0}^{j=K} \pi_j\lambda_j = \sum_{j=0}^{j=K} \lambda(K-j)\pi_j = \lambda(K-L) \tag{55}$$

Si aplicamos (28) a las máquinas que están en reparación y a aquéllas que están esperando reparación, tenemos

$$W = \frac{L}{\bar{\lambda}} \tag{56}$$

Al aplicar (29) a las máquinas que esperan reparación, obtenemos

$$W_q = \frac{L_q}{\bar{\lambda}} \tag{57}$$

Con el siguiente ejemplo se ilustra el uso de las fórmulas.

EJEMPLO 12 Patrullas

El Departamento de Policía de Gotham City tiene cinco patrullas. Una patrulla se descompone, y requiere servicio una vez cada 30 días. El departamento tiene dos mecánicos, cada uno de los cuales requiere tres días para reparar una patrulla. Los tiempos de descompostura y los tiempos de reparación son exponenciales.

1 Determine el número promedio de patrullas en buenas condiciones.

2 Encuentre el tiempo promedio de paralización para una patrulla que necesita reparación.

3 Estime la fracción de tiempo en que un mecánico en particular está desocupado.

Solución Es un problema de reparación de máquinas con $K = 5$, $R = 2$, $\lambda = \frac{1}{30}$ patrulla por día y $\mu = \frac{1}{3}$ de patrullas por día. Entonces,

$$\rho = \frac{\frac{1}{30}}{\frac{1}{3}} = \frac{1}{10}$$

Según (52)

$$
\begin{aligned}
\pi_1 &= \binom{5}{1}\left(\frac{1}{10}\right)\pi_0 = .5\pi_0 \\
\pi_2 &= \binom{5}{2}\left(\frac{1}{10}\right)^2 \pi_0 = .1\pi_0 \\
\pi_3 &= \binom{5}{3}\left(\frac{1}{10}\right)^3 \frac{3!}{2!2}\pi_0 = .015\pi_0 \\
\pi_4 &= \binom{5}{4}\left(\frac{1}{10}\right)^4 \frac{4!}{2!(2)^2}\pi_0 = .0015\pi_0 \\
\pi_5 &= \binom{5}{5}\left(\frac{1}{10}\right)^5 \frac{5!}{2!(2)^3}\pi_0 = .000075\pi_0
\end{aligned} \tag{58}
$$

Entonces $\pi_0(1 + .5 + .1 + .015 + .0015 + .000075) = 1$, o $\pi_0 = .619$. Ahora con (58) se tiene $\pi_1 = .310$, $\pi_2 = .062$, $\pi_3 = .009$, $\pi_4 = .001$ y $\pi_5 = 0$.

1 El número esperado de patrullas en buenas condiciones es $K - L$, que se obtiene mediante

$$
\begin{aligned}
K - \sum_{j=0}^{j=5} j\pi_j &= 5 - [0(.619) + 1(.310) + 2(.062) + 3(.009) + 4(.001) + 5(0)] \\
&= 5 - .465 = 4.535 \text{ patrullas en buenas condiciones}
\end{aligned}
$$

2 Determinamos $W = \frac{L}{\bar{\lambda}}$. De acuerdo con (55),

$$
\begin{aligned}
\bar{\lambda} &= \sum_{j=0}^{j=5} \lambda(5-j)\pi_j = \frac{1}{30}(5\pi_0 + 4\pi_1 + 3\pi_2 + 2\pi_3 + \pi_4 + 0\pi_5) \\
&= \frac{1}{30}[5(.619) + 4(.310) + 3(.062) + 2(.009) + 1(.001) + 0(0)] \\
&= 0.151 \text{ patrullas por día}
\end{aligned}
$$

o bien,

$$
\bar{\lambda} = \lambda(K - L) = \frac{4.535}{30} = 0.151 \text{ patrullas por día}
$$

Como $L = 0.465$ patrulla, entonces $W = \frac{0.465}{.0151} = 3.08$ días.

3 La fracción de tiempo en que un mecánico esta desocupado es $\pi_0 + 0.5\pi_1 = .619 + .5(.310) = .774$.

Si hubiera tres personas encargadas de la reparación, la fracción del tiempo en que un trabajador estaría desocupado sería $\pi_0 + (\frac{2}{3})\pi_1 + (\frac{1}{3})\pi_2$, y para un personal de R personas, la probabilidad de que un trabajador en particular esté desocupado es

$$
\pi_0 + \frac{(R-1)\pi_1}{R} + \frac{(R-2)\pi_2}{R} + \cdots + \frac{\pi_{R-1}}{R}
$$

Hoja de cálculo para el problema de reparación de máquinas

Machrep.xls

En la figura 23 (archivo Machrep.xls) se presenta una plantilla para hoja de cálculo para el modelo de reparación de máquinas. Se escribe λ en la celda B2; μ en C2 y la cantidad de mecánicos en D2. En la celda F2 se escribe el número de máquinas. En el renglón 4 se calculan L, L_q, L_s, W, W_q y W_s. L_s es igual al número esperado de máquinas (en estado

	A	B	C	D	E	F	G	H
1	MACHINE	LAMBDA?	MU?	R?	RO	K?		
2	REPAIR	0.03333333	0.33333333	3	0.1	5		
3	MODEL	L	LS	LQ	W	WS	WQ	
4		0.45494681	0.45450532	0.0004415	3.00291412	3	0.002914124	
5								
6								
7								
8								
9								
10								
11								
12	STATE	LAMBDA(J)	MU(J)	CJ	PROB	#IN QUEUE	COLA*COLE	COLE*COL
13	0	0.16666667	0	1	0.62085236	0	0	0
14	1	0.13333333	0.33333333	0.5	0.31042618	0	0.310426181	0
15	2	0.1	0.66666667	0.1	0.06208524	0	0.124170472	0
16	3	0.06666667	1	0.01	0.00620852	0	0.018625571	0
17	4	0.03333333	1	0.00066667	0.0004139	1	0.001655606	0.0004139
18	5	0	1	2.2222E-05	1.3797E-05	2	6.89836E-05	2.7593E-05
19	6	0	1	0	0	3	0	0
20	7	0	1	0	0	4	0	0
21	8	0	1	0	0	5	0	0

FIGURA 23

estable) que están en reparación y W_s es igual al tiempo previsto que una máquina descompuesta pasa en reparación. En la columna E se calculan las probabilidades de estado estable. Estamos suponiendo que $K \leq 1\,000$. La información del ejemplo 12 es la que aparece en la figura 23.

Uso de LINGO para los cálculos del modelo de reparación de máquinas

La función de LINGO @PFS($K*\lambda/\mu,R,K$) proporciona L, el número esperado (en el estado estable) de máquinas en malas condiciones. Las siglas FS significan *Finite Source* (origen finito). Por lo tanto, para el ejemplo 12, @PFS(5*(1/30)/(1/3),2,5) da el valor de .465.

PROBLEMAS

Grupo A

1 Una lavandería tiene cinco lavadoras. Una máquina se descompone una vez cada cinco días. Un mecánico es capaz de reparar la lavadora en un promedio de 2.5 días. Tres mecánicos están en servicio, por ahora. El dueño de la lavandería tiene la opción de reemplazarlos con un supertrabajador, que puede reparar una lavadora en un promedio de 5/6 de día. El salario del supertrabajador es igual al pago de los tres empleados regulares. Los tiempos de descomposturas y los tiempos de servicio son exponenciales. ¿Debe la lavandería reemplazar a los tres trabajadores por el supertrabajador?

2 Mi perra acaba de tener tres cachorros muy juguetones que saltan dentro y fuera de su canasto que les sirve de cama. Un cachorro pasa un promedio de 10 minutos (distribución exponencial) dentro de su canasto antes de saltar hacia fuera. Una vez fuera, el perrito pasa un promedio de 15 minutos (distribución exponencial) antes de regresar a su cama.

a ¿Cuál es la probabilidad de que más perritos estén afuera que adentro de su canasto en cualquier momento dado?

b ¿Cuántos cachorros estarán en promedio en su cama?

Grupo B

3† Gotham City tiene 10 000 luminarias en el alumbrado público. Los investigadores urbanos determinaron que, en cualquier momento dado, está fundido un promedio de 1 000 luminarias. Una luminaria se funde, en promedio, luego de 100 días de uso. La ciudad contrató a Mafia, Inc., para que cambie las luminarias fundidas. El contrato de Mafia, Inc., establece que la compañía reemplazará supuestamente una lámpara fundida del alumbrado público en un promedio de siete días. ¿Opina que Mafia, Inc., cumple con el contrato?

4 Con este ejemplo se ilustra el rechazo. La heladería Oryo Cookie de la plaza Dunkirk tiene tres competidores. Como a la gente no le gusta hacer largas líneas de espera para conseguir un helado, la tasa de llegadas de Oryo Cookie depende de la cantidad de personas que estén en el establecimiento. Más claro, cuando $j \leq 4$ personas están presentes en la heladería, los clientes llegan a una tasa de $(20 - 5j)$ por hora. Si más de cuatro personas están en la heladería Oryo, la tasa de llegadas es de cero. Por cada cliente, los ingresos menos costos de materia prima son 50 centavos. Cada servidor recibe como paga tres dólares por hora. Un servidor puede atender a un promedio de 10 clientes por hora. Para maximizar las ganancias esperadas (ingresos menos costos de materia prima y costos de mano de obra, ¿cuántos trabajado-

†Basado en Kolesar (1979).

res debe contratar Oryo? Suponga que los tiempos entre llegadas y los tiempos de servicio son exponenciales.

5 Suponga que todos los tiempos entre llegadas a un sistema de un solo servidor son exponenciales, pero cuando n clientes están presentes hay una probabilidad $\frac{n}{n+1}$ de que una llegada será rechazada y dejará el sistema antes de entrar al servicio. Asimismo, suponga tiempos de servicio exponenciales.

- **a** Determine la distribución de probabilidad del número de personas presentes en el estado estable.
- **b** Encuentre la cantidad esperada de personas presentes en el estado estable. *Sugerencia*: aplique el hecho de que podría ser útil:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

6 En el modelo de reparación de máquinas, demuestre que $W = K/\bar{\lambda} - (1/\lambda)$.

7 (Requiere el uso de una hoja de cálculo o LINGO.) El modelo de reparación de máquinas se podría usar con frecuencia para aproximar el comportamiento de la CPU (unidad central de proceso) de una computadora. Suponga que 20 terminales (asuma que siempre están ocupadas) alimentan a la unidad central. Después de que la unidad central responde al usuario, éste tiene una ventaja de 80 segundos antes de enviar otra petición a la unidad (a esto se le llama *tiempo para pensar*). La unidad central requiere un promedio de 2 s para contestar cualquier petición. ¿Cuánto tiempo tiene que esperar un usuario antes de que la unidad central resuelva su petición? ¿Qué tanto se modificaría su respuesta si hubiera 30 terminales? ¿Y si hubiera 40 terminales?. Naturalmente, usted debe plantear suposiciones apropiadas respecto a la exponencialidad para poder contestar a esta pregunta.

8 Allbest Airlines tiene 100 aviones. Las aeronaves se descomponen un promedio de dos veces al año, y se requiere una semana para ajustarlas. Si se supone que los tiempos entre las descomposturas y las reparaciones son exponenciales, ¿cuántos mecánicos se necesitan para asegurar que hay por lo menos 95% de probabilidad de que 90 o más aeronaves estén disponibles? (*Sugerencia*: utilice una tabla).

9 Un ejército tiene 200 tanques. Los tanques necesitan mantenimiento 10 veces al año, y para el mantenimiento se requiere un promedio de dos días. Al ejército le gustaría tener por lo menos un promedio de 180 tanques en buenas condiciones. ¿Cuántos mecánicos se requieren? Suponga que los tiempos entre llegadas y los de servicio son exponenciales. (*Sugerencia*: utilice una tabla).

Grupo C

10 Bectol, Inc., está construyendo una presa. Se requiere un total de 10 millones de pies cúbicos de tierra para la construcción. Se utiliza un cargador frontal o *bulldozer* para extraer tierra para la presa. Luego la tierra se envía a la presa en camiones de carga. Sólo hay un cargador frontal disponible y su renta es de 100 dólares por hora. Bectol puede rentar tantos camiones de carga como quiera, a 40 dólares por hora. Cada camión puede transportar 1 000 pies cúbicos de tierra. El cargador frontal se tarda un promedio de 12 minutos en cargar al camión con tierra, y cada camión, un promedio de cinco minutos en entregar la tierra a la presa y regresar al *bulldozer*. Proponga las suposiciones apropiadas respecto a la exponencialidad, y determine cómo Bectol puede minimizar el costo total esperado del movimiento de tierras necesario para la presa. (*Sugerencia*: ¡hay un problema de reparación de máquinas dondequiera!)

20.10 Líneas de espera exponenciales en serie y redes abiertas de líneas de espera

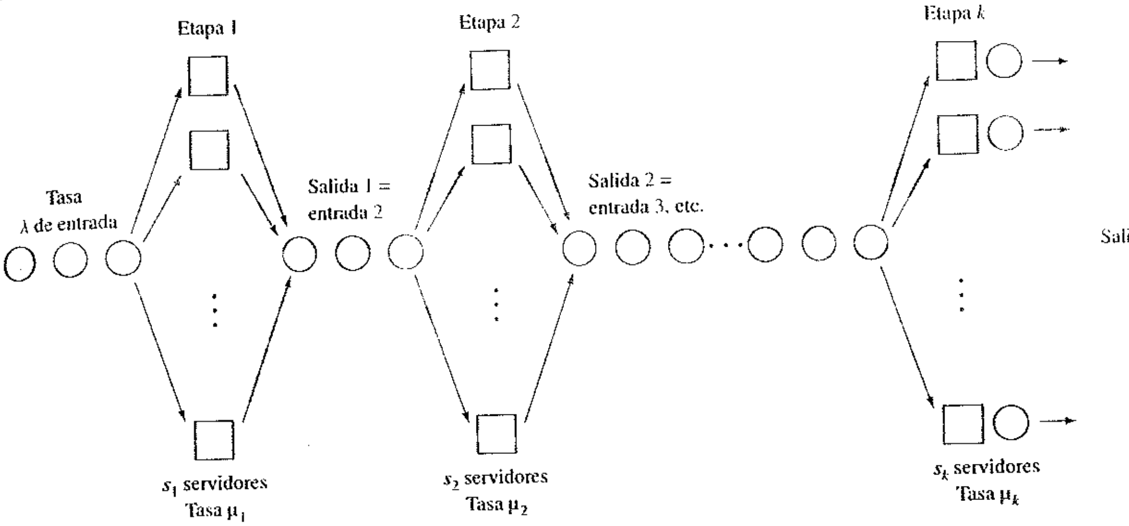
En los modelos de líneas de espera tratados hasta este momento, un tiempo de servicio completo al cliente transcurre en un solo servidor. En muchas situaciones (como en la fabricación de un producto en la línea de ensamble) el servicio al cliente no se completa hasta que el cliente ha sido atendido por más de un servidor (véase, por ejemplo, la figura 24).

Al entrar al sistema (figura 24), la llegada se somete al servicio de la etapa 1 (después de esperar en la cola, si todos los servidores de la etapa 1 están ocupados cuando llega). Después de completar el servicio de la etapa 1, el cliente espera y tiene que pasar por el servicio de la etapa 2. Este proceso continúa hasta que el cliente completa el servicio de la etapa k. Un sistema como el de la figura 24 se denomina **sistema de líneas de espera de k etapas en serie** (o en tándem). Un teorema notable, que se debe a Jackson (1957), es el siguiente (véase una demostración en Heyman y Sobel (1984)).

TEOREMA 4

Si (1) los tiempos entre llegadas para un sistema de líneas de espera en serie son exponenciales con tasa λ, (2) los tiempos de servicio por cada servidor de la etapa i son exponenciales y (3) cada etapa tiene una sala de espera de capacidad infinita, entonces los tiempos entre llegadas para las llegadas a cada etapa del sistema de líneas de espera son exponenciales con tasa λ.

FIGURA 24
Colas exponenciales en serie



Para que este resultado sea válido, cada etapa debe tener suficiente capacidad de servicio para atender a los clientes que arriban a una tasa λ; si no es así, la cola se "amplificará" en la etapa cuya capacidad sea insuficiente. De acuerdo con el estudio del sistema de líneas de espera $M/M/s/GD/\infty/\infty$ en la sección 20.6, vemos que cada etapa tiene capacidad suficiente para atender las llegadas que llegan con tasa λ si y sólo si para $j = 1, 2, \ldots, k$, $\lambda > s_j\mu_j$. Si $\lambda < s_j\mu_j$, el resultado de Jackson implica que la etapa j del sistema de la figura 24 se podría analizar como un sistema $M/M/s_j/GD/\infty/\infty$ con tiempos entre llegadas exponenciales cuya tasa es λ y tiempos de servicio exponenciales con un tiempo de servicio medio $\frac{1}{\mu_j}$. La utilidad del resultado de Jackson se ilustra mediante el ejemplo siguiente.

EJEMPLO 13 Ensamble de automóviles

Las últimas operaciones que se efectúan en un automóvil antes de que su manufactura esté completa son instalar el motor y poner los neumáticos. Un promedio de 54 automóviles que llegan por hora requiere estas dos operaciones. Hay un trabajador disponible para instalar el motor, y es capaz de atender un promedio de 60 automóviles por hora. Después de que ya se instaló el motor, el vehículo pasa a la estación de neumáticos y espera a que le pongan los propios. Hay tres trabajadores en esta estación. Cada uno trabaja en un automóvil a la vez, y es capaz de fijar los neumáticos a un vehículo en un promedio de tres minutos. Tanto los tiempos entre llegadas como los tiempos de servicio son exponenciales.

1 Determine la longitud media de la cola en cada estación de trabajo.

2 Estime el tiempo total previsto que un vehículo pasa en espera de servicio.

Solución Éste es un sistema de líneas de espera en serie con $\lambda = 54$ automóviles, $s_1 = 1$, $\mu_1 = 60$ automóviles por hora, $s_2 = 3$ y $\mu_2 = 20$ automóviles por hora (véase figura 25). Como $\lambda < \mu_1$ y $\lambda < 3\mu_2$, ninguna cola se "amplificará", por lo que es aplicable el teorema de Jackson. Por lo que toca a la etapa 1 (motor), $\rho = \frac{54}{60} = .90$. Entonces (27) proporciona

$$L_q \text{ (para el motor)} = \left(\frac{\rho^2}{1-\rho}\right) = \left[\frac{(.90)^2}{1-.90}\right] = 8.1 \text{ automóviles}$$

Ahora, de acuerdo con (32), se tiene

$$W_q \text{ (para el motor)} = \frac{L_q}{\lambda} = \frac{8.1}{54} = 0.15 \text{ hora}$$

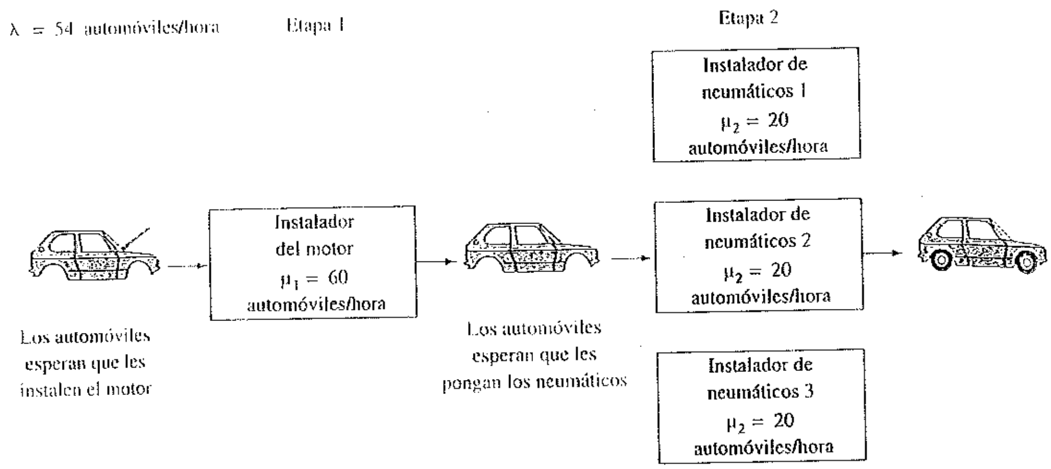


FIGURA 25
Sistema de líneas de espera en serie para automóviles

Por lo que se refiere a la etapa 2 (neumáticos), $\rho = \frac{54}{3(20)} = .90$. En la tabla 6 se encuentra que $P(j \geq 3) = .83$. Ahora (41) genera

$$L_q \text{ (para los neumáticos)} = \frac{.83(.90)}{1 - .90} = 7.47 \text{ automóviles}$$

Entonces,

$$W_q \text{ (para los neumáticos)} = \frac{L_q}{\lambda} = \frac{7.47}{54} = 0.138 \text{ hora}$$

Por lo tanto, el tiempo total previsto que un automóvil pasa en espera de la instalación del motor y los neumáticos es $0.15 + 0.138 = 0.288$ h.

Redes abiertas de líneas de espera

El tema que se trata enseguida es una generalización las de líneas de espera en serie. Al igual que en la figura 24, suponga que la estación j consiste en s_j servidores exponenciales, y cada uno opera a una tasa μ_j. Se supone que los clientes llegan a la estación j desde afuera del sistema de colas a una tasa r_j. También se supone que estos tiempos entre llegadas se apegan a una distribución exponencial. Una vez que se completa el servicio en la estación i, un cliente se forma en la cola de la estación j con probabilidad p_{ij} y termina el servicio con probabilidad

$$1 - \sum_{j=1}^{j=k} p_{ij}$$

Ahora definamos λ_j, la tasa a la cual los clientes llegan a la estación j (se incluyen las llegadas a la estación j desde afuera del sistema y desde otras estaciones). $\lambda_1, \lambda_2, \ldots, \lambda_k$ se pueden determinar al resolver el siguiente sistema de ecuaciones lineales:

$$\lambda_j = r_j + \sum_{i=1}^{i=k} p_{ij}\lambda_i \qquad (j = 1, 2, \ldots, k)$$

Esto se infiere porque una fracción p_{ij} de las llegadas λ_i a la estación i llegarán luego a la estación j. Suponga que $s_j\mu_j > \lambda_j$ se cumple para todas las estaciones. Entonces se puede demostrar que la distribución de probabilidad del número de clientes presentes en la estación j y el número esperado de clientes presentes en la estación j se puede determinar si se trata a la estación j como un sistema $M/M/s_j/GD/\infty/\infty$ con tasa de llegadas λ_j y tasa de servicio μ_j. Si para alguna j, $s_j\mu_j \leq \lambda_j$, entonces no existe distribución de estado estable de los clientes. Vale la pena hacer notar que la cantidad de clientes presentes en cada estación es una variable aleatoria independiente. Es decir, ¡conocer la cantidad de personas en todas las estaciones que no son la estación j no nos dice nada respecto a la distribución del número de personas en la estación j! Este resultado no se cumple en el caso de que no sean exponenciales los tiempos entre llegadas o los tiempos de servicio.

Para determinar L, el número esperado de clientes en el sistema de colas, sume simplemente el número esperado de clientes presentes en cada estación. El tiempo promedio que un cliente pasa en el sistema, W, se encuentra aplicando la fórmula $L = \lambda W$ a todo el sistema. Aquí, $\lambda = r_1 + r_2 + \cdots + r_k$ porque esto representa la cantidad promedio de clientes por unidad de tiempo que llega al sistema. Mediante el ejemplo siguiente se ilustra el análisis de redes abiertas de líneas de espera.

EJEMPLO 14 Ejemplo de redes abiertas de líneas de espera

Considere dos servidores. Un promedio de ocho clientes por hora llega desde fuera al servidor 1 y un promedio de 17 clientes por hora llega desde fuera al servidor 2. Los tiempos entre llegadas son exponenciales. El servidor 1 es capaz de atender a una tasa exponencial de 20 clientes por hora, y el servidor 2 atiende a una tasa exponencial de 30 clientes por hora. Después de completar el servicio en el servidor 1, la mitad de los clientes sale del sistema, y la mitad se dirige al servidor 2. Después de finalizar el servicio en el servidor 2, $\frac{3}{4}$ de los clientes completan sus trámites y $\frac{1}{4}$ regresa al servidor 1.

1 ¿Qué fracción del tiempo el servidor 1 está desocupado?

2 Determine el número esperado de clientes en cada servidor.

3 Encuentre el tiempo promedio que pasa un cliente en el sistema.

4 ¿Qué tanto cambiarían las respuestas de (1) a (3) si el servidor 2 pudiera atender sólo un promedio de 20 clientes por hora?

Solución Ésta es una red abierta de líneas de espera con $r_1 = 8$ clientes/hora y $r_2 = 17$ clientes/hora. Asimismo, $p_{12} = .5$, $p_{21} = .25$ y $p_{11} = p_{22} = 0$. Podemos determinar λ_1 y λ_2 al resolver $\lambda_1 = 8 + .25\lambda_2$ y $\lambda_2 = 17 + .5\lambda_1$. Así se llega a $\lambda_1 = 14$ clientes/hora y $\lambda_2 = 24$ clientes/hora.

1 Se podría tratar al servidor 1 como un sistema $M/M/1/GD/\infty/\infty$ con $\lambda = 14$ clientes/hora y $\mu = 20$ clientes/hora. Entonces, $\pi_0 = 1 - \rho = 1 - .7 = .3$. Por lo tanto, el servidor 1 está desocupado 30% del tiempo.

2 De acuerdo con (26), encontramos que $L = \frac{14}{20-14} = \frac{7}{3}$ en el servidor 1 y $L = \frac{24}{30-24} = 4$ en el servidor 2. Por lo tanto, un promedio de $4 + \frac{7}{3} = \frac{19}{3}$ clientes estará presente en el sistema.

3 $W = \frac{L}{\lambda}$, donde $\lambda = 8 + 17 = 25$ clientes/hora. De aquí que

$$W = \frac{\left(\frac{19}{3}\right)}{25} = \frac{19}{75} \text{ hora}$$

4 En este caso, $s_2\mu_2 = 20 < \lambda_2$, de modo que no existe estado estable.

Modelos de redes de redes de comunicación de datos

Compnetwork.xls

Las redes de líneas de espera se usan por lo regular para modelar redes de comunicación de datos. Los modelos de líneas de espera permiten determinar el retraso característico que sufren los datos transmitidos, y también diseñar la red. El análisis se basa en Tannenbaum (1981). Véase el archivo Compnetwork.xls.

Considere una red de comunicación de datos con cinco nodos (A, B, C, D, E). Suponga que cada paquete de datos transmitido consiste en 800 bits, y que el número de paquetes por segundo que se tiene que transmitir entre cada par de nodos es como se indica en la figura 26.

Por ejemplo, se debe enviar un promedio de cinco paquetes por segundo desde el nodo A hasta el nodo B. Los paquetes no siempre se transmiten por el itinerario más directo. Suponga que las rutas usadas para transmitir cada tipo de mensaje son los que se señalan en la figura 27.

Por ejemplo, todos los mensajes que tienen que ir desde A hasta D se transmiten por medio de la ruta A-B-D. Cada arco o itinerario que une dos nodos tiene una capacidad medida en miles de bits por segundo. Es decir, un arco con 16 000 bits/segundo de capacidad puede "atender" 16 000/800 = 20 paquetes/segundo. Cada capacidad de arco en miles de bits por segundo se presenta en la figura 28. Nuestro interés es, naturalmente, el retraso esperado de un paquete. Asimismo, si la capacidad total de la red es limitada, es importante determinar la capacidad de cada arco que minimizará el retraso esperado de un paquete. La forma común de enfocar este problema es tratar cada arco como si fuera una cola $M/M/1$ independiente, y determinar el tiempo previsto que pasa cada paquete transmitido a través de ese arco mediante la fórmula

	A	B	C	D	E	F	G
23	Packets/second		A	B	C	D	E
24		A	0	5	4	1	7
25		B	5	0	6	3	2
26		C	4	6	0	3	3
27		D	1	3	3	0	3
28		E	7	2	3	3	0

FIGURA 26

	A	B	C	D	E	F	G
30			A	B	C	D	E
31	Route used	A	-	AB	ABC	ABD	AE
32		B	BA	-	BC	BD	BDE
33		C	CBA	CB	-	CD	CDE
34		D	DBA	DB	DC	-	DCE
35		E	EA	EDB	EDC	ECD	-

FIGURA 27

	B	C	D	E	F
4	Line	Packets per second	Capacity (000) bits per second	Service Rate in Packets per second	W in seconds
5	AB	10	20	25	0.066667
6	AE	7	20	25	0.055556
7	BC	10	15	18.75	0.114286
8	BD	6	10	12.5	0.153846
9	CD	9	10	12.5	0.285714
10	CE	3	10	12.5	0.105263
11	DE	5	10	12.5	0.133333
12	BA	10	20	25	0.066667
13	EA	7	20	25	0.055556
14	CB	10	15	18.75	0.114286
15	DB	6	10	12.5	0.153846
16	DC	9	10	12.5	0.285714
17	EC	3	10	12.5	0.105263
18	ED	5	10	12.5	0.133333

FIGURA 28

$$W = \frac{1}{\mu - \lambda}$$

Con el fin de ilustrar lo anterior, considere el arco AB. Los paquetes que se transmiten desde A a B, A a C y A a D usarán este arco. Es un total de $5 + 4 + 1 = 10$ paquetes por segundo. Suponga que el arco AB tiene una capacidad de 20 000 bits por segundo. Entonces, para el arco AB, $\mu = 20\,000/800 = 25$ paquetes por segundo, y $\lambda = 10$ paquetes por segundo. Luego entonces,

$$W = \frac{1}{25 - 10} = .06667 \text{ segundo}$$

En los renglones 5 a 18 de la figura 28, calculamos W por cada arco que hay en esta red de comunicaciones. Observe que se supone que la red es simétrica (es decir, la tasa de llegadas y capacidad AB es igual a la tasa de llegadas y capacidad BA), de modo que los renglones 12 a 18 son copias exactas de los renglones 5 a 11.

La fórmula siguiente se usa para determinar el retraso promedio que sufre un paquete:

$$\text{Retraso promedio por paquete} = \frac{\sum_{\text{todos los arcos}} \begin{array}{c}\text{(Tasa de llegadas del arco)} *\\ \text{(tiempo previsto pasado en el arco)}\end{array}}{\text{Número total de llegadas}}$$

El retraso promedio por paquete se calcula en la celda C20 de la figura 29 con la fórmula:

=SUMPRODUCT(C5:C18,F5:F18)/SUM(C24:G28)

Por lo tanto, el retraso promedio por paquete es .18 s por paquete.

	B	C
19		
20	Mean	0.180416
21	Time in system	
22	seconds	

FIGURA 29

	B	C	D	E	F	G
1		800				
2		bits/packet				
3						
4	Line	Packets per second	Capacity (000) bits per second	Service Rate in Packets per second	W in seconds	Diff
5	AB	10	18.31869	22.89836	0.077529	12.89836
6	AE	7	14.23287	17.79109	0.092669	10.79109
7	BC	10	18.31662	22.89577	0.077545	12.89577
8	BD	6	12.79577	15.99471	0.100053	9.994709
9	CD	9	16.98872	21.2359	0.081727	12.2359
10	CE	3	8.052237	10.0653	0.141537	7.065296
11	DE	5	11.2951	14.11888	0.109663	9.118877
12	BA	10	18.31869	22.89836	0.077529	12.89836
13	EA	7	14.23287	17.79109	0.092669	10.79109
14	CB	10	18.31662	22.89577	0.077545	12.89577
15	DB	6	12.79577	15.99471	0.100053	9.994709
16	DC	9	16.98872	21.2359	0.081727	12.2359
17	EC	3	8.052237	10.0653	0.141537	7.065296
18	ED	5	11.2951	14.11888	0.109663	9.118877
19		Total cap	200			
20	Mean	0.121843				

FIGURA 30

Suponga que tenemos sólo 200 000 bits/segundo de la capacidad total para asignar a la red. ¿Cómo podemos asignar capacidad para minimizar el retraso esperado por paquete? Refiérase a la hoja de *Optimization* (Optimización) en el archivo Compnetwork.xls (figura 30). En G5:G18 calculamos la Tasa de servicio (en paquetes/segundo) – Tasa de llegada (en paquetes/segundo). Limitamos esto a que sea por lo menos .01 de modo que exista un estado estable. Entonces, la ventana de Solver es como la que se muestra en la figura 31.

Escogemos las capacidades D5:D11 (recuerde que D12:D18 son copias exactas de D5:D11) para minimizar el tiempo previsto del sistema (C20). Garantizamos que cada tasa de servicio del arco excede su tasa de llegada (G5:G11 $\geq .01$), cada capacidad es no negativa (D5:D11$\geq$ 0) y la capacidad total es a lo más 200 000 (D19$\leq$200). Encontramos que podemos reducir el tiempo previsto de un paquete en el sistema a .1218 segundos.

Naturalmente, hemos supuesto un itinerario estático, en el cual las tasas de llegadas para cada nodo no varían con el estado de la red. En realidad, se han elaborado muchos esquemas de rutas dinámicos y complejos. Un esquema de rutas sería dinámico si, por ejemplo, al estar el arco *AB* congestionado y el arco *AD* relativamente libre, enviamos mensajes directamente desde *A* hasta *D* por medio del itinerario *A-B-D*.

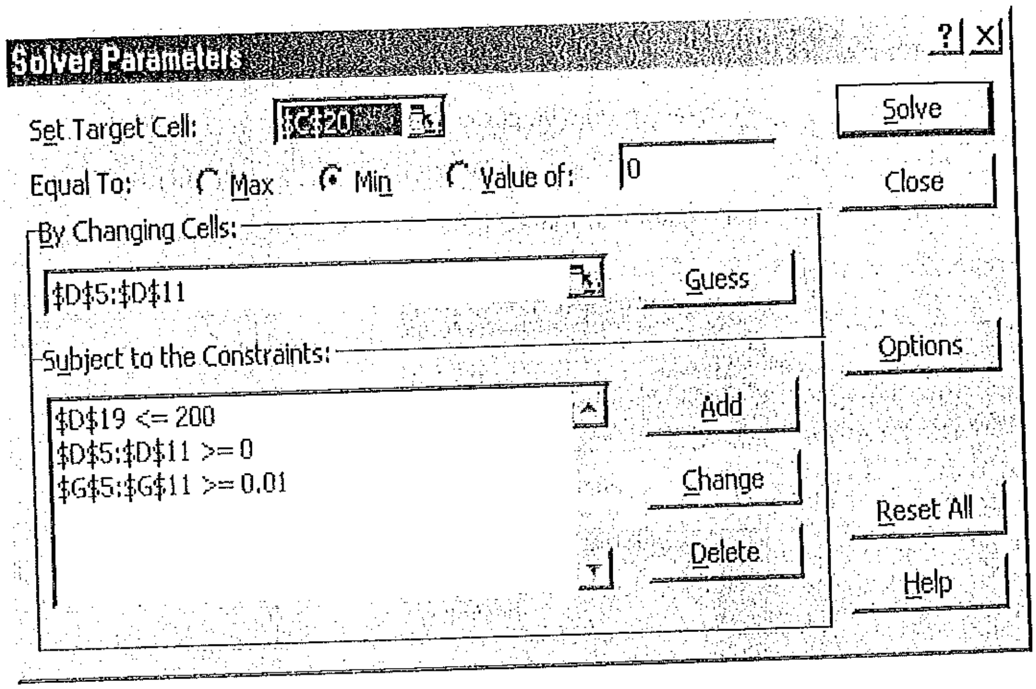


FIGURA 31

PROBLEMAS

Grupo A

1 Un departamento de la Social Security Administration tiene en sus manos las dos opciones siguientes para procesar las solicitudes de Tarjeta de Seguridad Social.

Opción 1 Tres empleados procesan las solicitudes en paralelo a partir de una sola cola. Cada empleado llena el formulario de la solicitud en presencia del solicitante. El tiempo de proceso es exponencial con una media de 15 minutos. Los tiempos entre llegadas son exponenciales.

Opción 2 Cada solicitante llena primero una solicitud sin ayuda del empleado. El tiempo para lograrlo está exponencialmente distribuido con una media de 65 minutos. Luego que el solicitante llenó el formulario, se forma en una sola cola para esperar a que uno de los tres empleados verifique el formulario. El empleado pasa cuatro minutos (distribución exponencial), en promedio, revisando la solicitud.

El tiempo entre llegadas de los solicitantes es exponencial; un promedio de 4.8 solicitantes llega cada hora. ¿Con cuál opción los solicitantes estarán menos tiempo en la oficina?

2 Considere una línea de ensamble de automóviles en la cual cada automóvil se somete a dos tipos de servicio: pintura y, luego, instalación del motor. Un promedio de 22.4 chasis sin pintar llega cada hora a la línea de ensamble. Se requiere un promedio de 2.4 minutos para pintar un automóvil y un promedio de 3.75 minutos para instalar un motor. La línea de ensamble tiene sólo un pintor y dos instaladores de motores. Suponga que los tiempos entre llegadas y los tiempos de servicio son exponenciales.

a ¿Cuántos automóviles pintados, en promedio, sin tener instalado por completo el motor están en la línea de ensamble?

b ¿Cuánto tiempo tendrá que esperar, en promedio, un vehículo pintado antes de que la instalación de su motor empiece?

3 Considere los sistemas de colas siguientes:

Sistema 1 Un promedio de 40 clientes llega cada hora. Los tiempos entre llegadas son exponenciales. Los clientes deben completar dos tipos de servicios antes de salir del sistema. El primer servidor requiere un promedio de 30 segundos (distribuidos exponencialmente) para efectuar el servicio tipo 1. Después de esperar en la cola, los clientes consiguen el servicio tipo 2 (distribución exponencial con media de un minuto) que proporciona un solo servidor. Después de completar el servicio tipo 2 el cliente deja el sistema.

Sistema 2 El proceso de llegada para el sistema 2 es idéntico al proceso entre llegadas del sistema 1. En el sistema 2, un cliente debe completar sólo un tipo de servicio. El tiempo de servicio promedia 1.5 minutos, y está distribuido exponencialmente. Hay dos servidores disponibles.

¿En cuál sistema pasa menos tiempo un cliente representativo?

4 Un promedio de 120 estudiantes llega cada hora (tiempos entre llegadas exponenciales) a la oficina de servicios escolares de la universidad del estado con el objetivo de modificar sus cursos. Para completar este proceso, una persona debe pasar por tres lugares. Cada lugar consta de un solo servidor. Los tiempos de servicio en cada lugar son exponenciales con los siguientes tiempos medios: lugar 1, 20 segundos; lugar 2, 15 segundos; lugar 3, 12 segundos. En promedio, ¿cuántos estudiantes se presentarán en la oficina para hacer ese trámite?

5 En promedio llegan 10 trabajos por hora a un taller. Los tiempos entre llegadas de los trabajos son exponenciales. Se requiere un tiempo promedio de 10/3 minutos (bajo distribución exponencial) para finalizar un trabajo. Un tercio de todos los trabajos terminados requieren, infortunadamente, ser retrabajados. Por lo tanto, hay una probabilidad de 1/3 de que un trabajo terminado debe esperar en la cola para ser trabajado de nuevo. En el estado estable, ¿cuántos trabajos esperaría uno encontrar en el taller? ¿Cuál sería la respuesta si un trabajo se termina, en promedio, en cinco minutos?

6 Considere un sistema de colas que consta de tres estaciones en serie. Cada estación consisten en un solo servidor, el cual puede procesar un promedio de 20 trabajos por hora (los tiempos de proceso en cada estación son exponenciales). Un promedio de 10 trabajos por hora (los tiempos entre llegadas son exponenciales) llega a la estación 1. Cuando un trabajo completa el servicio en la estación 2, hay .1 de probabilidad de que regresará a la estación 1 y .9 de probabilidad de que pasará a la estación 3. Cuando un trabajo completa su servicio en la estación 3 hay .2 de probabilidad de que regresará a la estación 2 y .8 de que dejará el sistema. Todos los trabajos que completan su servicio en la estación 1 pasan inmediatamente a la estación 2.

a Encuentre la fracción de tiempo en que el servidor está ocupado.

b Estime la cantidad esperada de trabajos en el sistema.

c Estime el tiempo promedio que un trabajo pasa en el sistema.

7 Antes de finalizar la producción, un bien debe pasar por tres etapas. En promedio, un nuevo producto empieza en la etapa 1 cada seis minutos. El tiempo promedio para procesar el producto en cada etapa es como se indica a continuación: etapa 3, tres minutos; etapa 2, dos minutos; etapa 3, un minuto. Después de finalizar la etapa 3, se inspecciona el producto (suponga que no toma tiempo). De los productos finales, 10% tiene algún defecto y debe regresar a la etapa 1, volver a recorrer el sistema completo. Luego de completar la etapa 3, 20% del producto final tiene algún defecto. Esta cantidad tiene que regresar a la etapa 2 y pasar por 2 y 3, nuevamente. En promedio, ¿cuántos trabajos están en el sistema? Suponga que los tiempos entre llegadas y los tiempos de servicio son exponenciales, y que cada etapa consiste en un solo servidor.

8 Una red de comunicación de datos consta de tres nodos, A, B y C. Cada paquete transmitido contiene 500 bits de información. El número de paquetes por segundo que se tiene que transmitir entre un par de nodos es como sigue:

$$\begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} 0 & 4 & 3 \\ 4 & 0 & 6 \\ 3 & 6 & 0 \end{bmatrix} \end{array}$$

El itinerario usado por cada par de nodos es como sigue:

$$\begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} — & ACB & AC \\ BCA & — & BAC \\ CA & CAB & — \end{bmatrix} \end{array}$$

Suponga que las capacidades (en miles de bits por segundo) por cada arco son

Arco	Capacidad
AB	12
AC	13
BC	15
BA	12
CA	13
CB	15

a Calcule el retraso esperado de un paquete

b Si está disponible un total de 75 000 bits/s de capacidad, ¿cómo se debe asignar?

9[†] Llegan trabajos a un servidor que consiste en una CPU (unidad central de proceso) y dos discos (disco 1 y disco 2). Ahora hay seis clientes, y llega un promedio de tres trabajos por segundo. Cada visita a la CPU toma un promedio de .01 segundo, cada visita al disco 1 requiere un promedio de .02 segundos, y cada visita al disco 2, un promedio de .03 segundos. Cada uno de los trabajos que llega visita primero a la CPU. Después de visitar esta unidad, hay una probabilidad de 7/16 de que el trabajo visite luego al disco 1; una probabilidad de 8/16 de que visite después al disco 2, y una probabilidad de 1/16 de que el trabajo se complete. Después de que el trabajo visita el disco 1 o el 2 regresa a la unidad central.

a En promedio, ¿cuántas veces el trabajo visita la unidad central? ¿Y al disco 1? ¿Y al disco 2?

b ¿Cuánto pasa en promedio un trabajo en la unidad central? ¿Cuánto en el disco 1? ¿Cuánto en el disco 2? ¿Cuánto en el sistema?

10 Suponga que el servidor del problema 9 tiene ya ocho clientes. Conteste las mismas preguntas que en el problema 9.

11 Suponga que se le instala un caché al disco 2. Esto incrementará en 30% el tiempo medio que se requiere para una

[†]Basado en Jain (1991).

visita a la unidad central de proceso, y en 10% el que se requiere para ir al disco 2. Además, la caché para el disco 2 asegura que la mitad de las veces que el trabajo iba a ir al disco 2 ahora permanecerá en la unidad central y será procesado ahí. ¿La caché mejora la operación del sistema?

12 Suponga que eliminamos el disco 2. ¿Qué sucederá con el tiempo de respuesta del sistema? Usted puede suponer, en este problema, que todas las solicitudes que dejan la unidad central van al disco 1.

20.11 Sistema $M/G/s/GD/s/\infty$ (eliminación de clientes rechazados)

En muchos sistemas de líneas de espera, para cuestiones prácticas, el sistema pierde al cliente que encuentra todos los servidores ocupados. Por ejemplo, una persona que llama por teléfono a una aerolínea para solicitar una reservación, y escucha una señal de ocupado llamará probablemente a otra línea de aviación. O bien, suponga que alguien llama a la alarma contra incendios, y no hay bombas disponibles; el fuego se saldrá de control. Por lo tanto, en algún sentido, una solicitud de una bomba contra incendios que se presenta cuando no hay bombas disponibles podría considerarse como una pérdida del sistema. Si las llegadas que encuentran todos los servidores ocupados dejan el sistema, entonces este sistema recibe el nombre de **eliminación de clientes rechazados** (*blocked customers cleared*, BCC). Si se supone que los tiempos entre llegadas son exponenciales, entonces tal sistema se podría modelar como un sistema $M/G/s/GD/s/\infty$.

Para un sistema $M/G/s/GD/s/\infty$, L, W, L_q y W_q tienen interés limitado. Por ejemplo, como nunca se presenta una cola, $L_q = W_q = 0$. Si $\frac{1}{\mu}$ es el tiempo medio de servicio y λ es la tasa de llegadas, entonces $W = W_s = \frac{1}{\mu}$.

En la mayor parte de los sistemas BCC, el interés principal es la fracción de todas las llegadas que no se admiten. Puesto que las llegadas se rechazan sólo cuando s clientes están presentes, se perderá una fracción π_s de todas las llegadas. Por lo tanto, el sistema no atenderá un promedio de $\lambda\pi_s$ llegadas por unidad de tiempo. Como un promedio de $\lambda(1 - \pi_s)$ llegadas por unidad de tiempo realmente entran al sistema, podemos concluir que

$$L = L_s = \frac{\lambda(1 - \pi_s)}{\mu}$$

Por lo que se refiere un sistema $M/G/s/GD/s/\infty$, se puede demostrar que π_s depende de la distribución del tiempo de servicio sólo a través de su media $(\frac{1}{\mu})$. Este hecho se conoce como **fórmula de Erlang de pérdida**. En otras palabras, cualquier sistema $M/G/s/GD/s/\infty$ con una tasa de llegadas λ y tiempo de servicio medio $\frac{1}{\mu}$ tendrá el mismo valor de π_s. Si definimos $\rho = \frac{\lambda}{\mu}$, entonces, para un valor dado de s, el valor de π_s se puede determinar con la figura 32. El valor de ρ se lee simplemente en el eje de las x. Entonces, el valor de y sobre la curva del servidor s que corresponde a ρ será igual a π_s. Mediante el ejemplo siguiente se ilustra el uso de la figura 32.

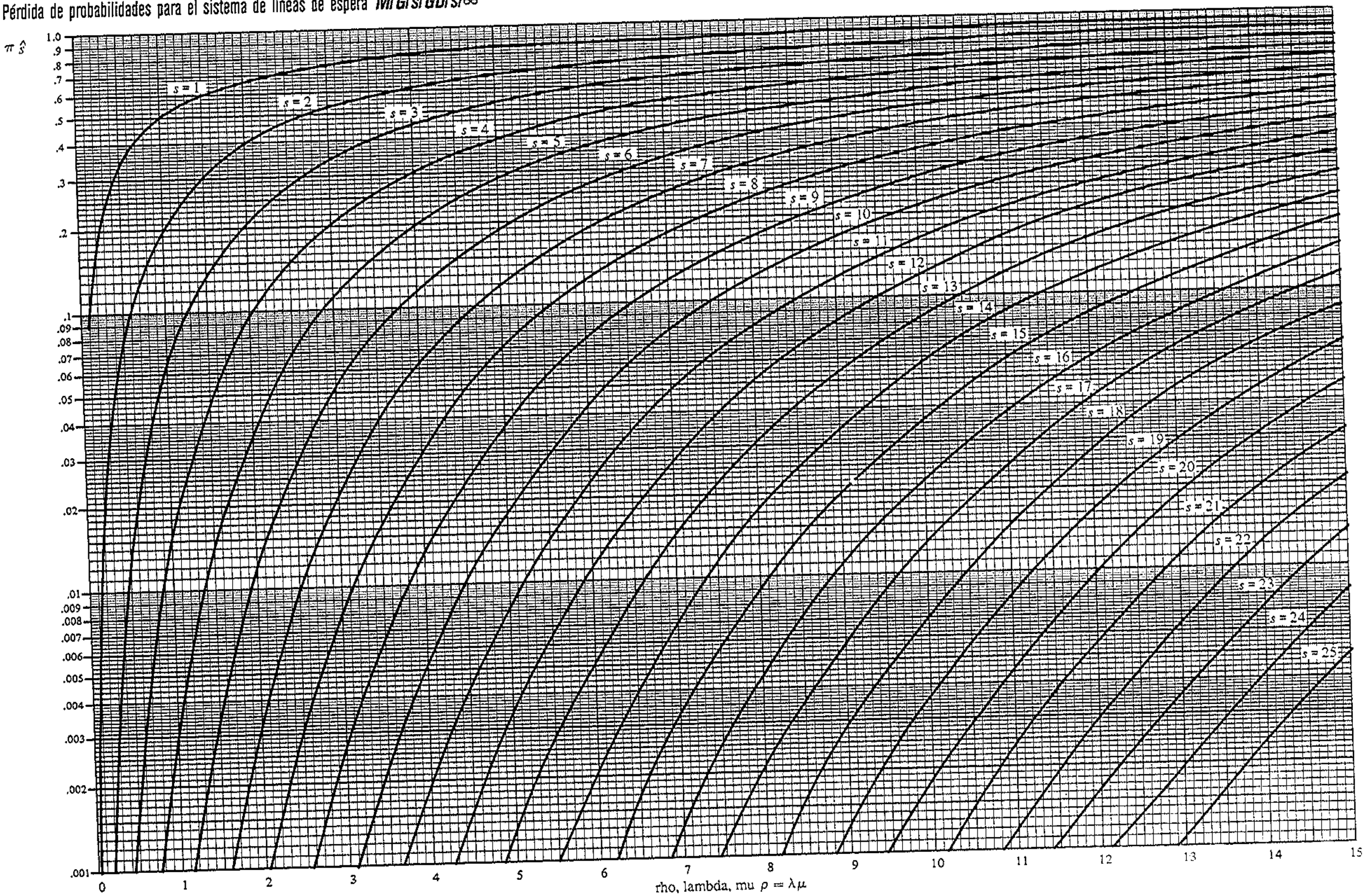
EJEMPLO 15 Llamadas que solicitan una ambulancia

Se recibe un promedio de 20 llamadas por hora que solicitan una ambulancia en el hospital de Gotham City. Una ambulancia tarda 20 minutos en recoger a un paciente y transportarlo al hospital. Entonces la ambulancia ya está lista para recoger a otro paciente. ¿Cuántas ambulancias debe tener el hospital para que asegure que hay cuando mucho 1% de probabilidad de que no sea capaz de responder en forma inmediata a la llamada que solicita una ambulancia? Suponga que los tiempos entre llegadas siguen una distribución exponencial.

Solución Sabemos que $\lambda = 20$ llamadas por hora y que $\frac{1}{\mu} = \frac{1}{3}$ hora. Por lo tanto, $\rho = \frac{\lambda}{\mu} = \frac{20}{3} = 6.67$. Para $\rho = 6.67$, buscamos el valor más pequeño de s para el cual π_s es .01 o más pequeño. De acuerdo con la figura 32, observamos que para $s = 13$, $\pi_s = .011$, y $s = 14$, $\pi_s = .005$. Por lo tanto, el hospital necesita 14 ambulancias para cumplir con sus normas de servicio deseado.

FIGURA 32

Pérdida de probabilidades para el sistema de líneas de espera $M/G/s/GD/s/\infty$



Fuente: Reimpreso con autorización del editor de *Introduction to Queuing Theory* de Robert B. Cooper. p. 316.Copyright©1980 de Elsevier Science Publishing Co.. Inc.

Hoja de cálculo para el MODELO BCC

Bcc.xls

En la figura 33 (archivo Bcc.xls) se proporciona una plantilla de hoja de cálculo para el sistema de líneas de espera $M/G/s/GD/s/\infty$. Escribimos λ en la celda B2; μ en la celda C2, y el número de servidores en la celda D2. En B4 calculamos el número esperado (en el estado estable) de servidores ocupados. El cálculo del valor de π_s tabulado en la figura 32 se presenta en la celda C4. La columna E da las probabilidades de estado estable para este modelo. Se supone que $s \leq 1\,000$. En la figura 33 se presentan los datos de λ, μ y s del ejemplo 15.

Cálculos de BCC con LINGO

La función de LINGO **@PEL**($\lambda/\mu,s$) proporciona π_s. Por lo que toca al ejemplo 15, la función **@PEL** (20/3,13) proporciona .010627, como en la figura 32. La función **@PEL** se puede usar para resolver un problema (tal como el problema 6) donde buscamos el número de servidores que minimizan el costo esperado por unidad de tiempo cuando el costo es la suma del costo de servicio y el costo debido al negocio perdido.

	A	B	C	D	E	F	G
1	BCC MODEL	LAMBDA?	MU?	s?			
2		20	3	14			
3		L OR LS	PI(s)				
4		6.63320534	0.0050192				
5							
6							
7							
8							
9							
10							
11							
12	STATE	LAMBDA(J)	MU(J)	CJ	PROB	#IN QUEUE	COLA*COLE
13	0	20	0	1	0.00127738	0	0
14	1	20	3	6.66666667	0.00851587	0	0.008515872
15	2	20	6	22.2222222	0.02838624	0	0.056772479
16	3	20	9	49.382716	0.06308053	0	0.189241596
17	4	20	12	82.3045267	0.10513422	0	0.420536879
18	5	20	15	109.739369	0.14017896	0	0.700894799
19	6	20	18	121.932632	0.1557544	0	0.934526398
20	7	20	21	116.126316	0.14833752	0	1.038362665
21	8	20	24	96.7719303	0.1236146	0	0.988916824
22	9	20	27	71.6829114	0.09156637	0	0.824097353
23	10	20	30	47.7886076	0.06104425	0	0.610442484
24	11	20	33	28.9627925	0.03699651	0	0.406961656
25	12	20	36	16.0904403	0.02055362	0	0.246643428
26	13	20	39	8.25150783	0.01054032	0	0.137024127
27	14	0	42	3.92928944	0.0050192	0	0.070268783
28	15	0	42	0	0	1	0
29	16	0	42	0	0	2	0
30	17	0	42	0	0	3	0
31	18	0	42	0	0	4	0
32	19	0	42	0	0	5	0

FIGURA 33

PROBLEMAS

Grupo A

1 Suponga que una estación de bomberos recibe cada hora un promedio de 24 solicitudes de bombas contra incendio. Cada solicitud ocasiona que una bomba contra incendio esté ocupada por un promedio de 20 minutos. ¿Cuántas bombas debe tener el departamento de bomberos para tener cuando mucho 1% de probabilidad de que sea incapaz de responder a una solicitud?

2 Una compañía de ventas por teléfono tiene que determinar cuántos operadores se necesitan para contestar los teléfonos durante el turno de 9 AM a 5 PM. La compañía estima que se recibe un promedio de 480 llamadas durante este periodo, y que la llamada promedio dura seis minutos. Si la compañía desea que cuando mucho una de cada 100 personas obtenga señal de ocupado, ¿cuántos operadores debe contratar para el turno mencionado? ¿Qué suposición necesita para la respuesta?

3 Refiérase al ejemplo 15. Suponga que el hospital cuenta con 10 ambulancias. Entonces, ¿cuántas ambulancias, en promedio, estarían en camino a solicitud de una llamada o regresarían del lugar al que acudieron?

4 Se dice que un sistema telefónico recibe 1 Erlang de uso por hora si las personas que hacen llamadas mantienen las líneas ocupadas durante un promedio de 3 600 segundos por hora. Suponga que un sistema telefónico recibe 2 Erlang de uso por hora. Si usted desea que sólo 1% de todas las llamadas sea bloqueado, ¿cuántas líneas telefónicas necesita?[†]

5 (Requiere una hoja de cálculo o LINGO.) En el tiempo de uso máximo, un promedio de 200 personas por hora intenta entrar a Jade Vax. El tiempo promedio que alguien pasa en Vax es 20 minutos. Si los servicios de computación de la universidad de Indiana quieren asegurar que durante el uso máximo sólo 1% de todos los usuarios obtenga un mensaje de "Todos los puestos están ocupados", ¿cuántos puertos debe tener Jade Vax?

6 (Requiere una hoja de cálculo o LINGO.) US Airlines recibe un promedio de 500 llamadas por hora de clientes que desean hacer una reservación (el tiempo entre llamadas sigue una distribución exponencial). Se requiere un promedio de tres minutos para atender una llamada. Cada cliente que compra un boleto contribuye con 100 dólares a las utilidades de US Airlines. Cuesta 15 dólares por hora el personal que atiende una línea telefónica. Cualquier cliente que recibe una señal de ocupado comprará un boleto en otra aerolínea. ¿Cuántas líneas telefónicas debe tener US Airlines?

[†]Basado en Green (1987).

Grupo B

7 Llegan en promedio 26 socios al año a la biblioteca I.U. a pedir prestado el libro *I Ching* (suponga que los tiempos entre llegadas son exponenciales). Los socios que llegan y se encuentran con que el libro está prestado se van y nunca vuelven. Un socio que se lleva un ejemplar del *I Ching* lo retiene durante un promedio de cuatro semanas.

a Si la biblioteca tiene sólo un ejemplar, ¿cuál es el número esperado de socios que llegará a solicitar el *I Ching* cada año y se encontrará con que no está el libro?

b Suponga que cada persona que llega a pedir el libro *I Ching* y no lo consigue representa un costo para la biblioteca de un dólar de buen nombre. Un ejemplar del *I Ching* dura dos años y cuesta 11 dólares. Un ladrón ha robado el único ejemplar de la bilblioteca. Con el fin de minimizar la suma los costos de compra y de buen nombre en los próximos dos años, ¿cuántos ejemplares del *I Ching* se deben comprar?

8[‡] La bodega de una compañía puede almacenar hasta cuatro unidades de un bien. Se recibe cada mes un promedio de 10 pedidos del producto. Los tiempos entre la recepción de los pedidos sucesivos tienen una distribución exponencial. Luego de usar un producto para cumplir con un pedido, se pide inmediatamente un reemplazo, y tarda un promedio de un mes en llegar el reemplazo. Si no hay productos cuando se recibe un pedido, éste se pierde. ¿Qué fracción de todos los pedidos se perderá debido a los déficits? (*Sugerencia*: considere el espacio de almacenamiento para cada producto como un servidor, y piense lo que significa para un servidor estar ocupado. Luego dé una idea de una definición apropiada de tiempo de "servicio").

[‡]Basado en Karush (1957).

20.12 Cómo saber si los tiempos entre llegadas y los tiempos de servicio son exponenciales[§]

¿De qué modo podemos determinar si los datos reales son consistentes con la suposición de tiempos entre llegadas y tiempos de servicio exponenciales? Suponga que, por ejemplo, se han observado los tiempos entre llegadas $t_1, t_2, \ldots, t_n$. Se puede demostrar que una estimación razonable de la tasa de llegadas λ está dada por

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{i=n} t_i}$$

[§]En esta sección se tratan temas que se podrían omitir sin perder la continuidad.

Por ejemplo, si $t_1 = 20$, $t_2 = 30$, $t_3 = 40$ y $t_4 = 50$, hemos observado cuatro llegadas en 140 unidades de tiempo, es decir, un promedio de una llegada por 35 unidades de tiempo. En este caso, la estimación de la tasa de llegadas $\hat{\lambda}$ está dada por

$$\hat{\lambda} = \frac{4}{20 + 30 + 40 + 50} = \frac{1}{35}$$

clientes por unidad de tiempo. Dada $\hat{\lambda}$, podemos intentar determinar si $t_1, t_2, \ldots, t_n$ son consistentes con la suposición de que los tiempos entre llegadas están regidos por una distribución exponencial con tasa $\hat{\lambda}$ y densidad $\hat{\lambda}e^{-\hat{\lambda}t}$. La manera más fácil de analizar esta conjetura es mediante la prueba de la bondad de ajuste de chi cuadrado con el objeto de determinar si es razonable concluir que $t_1, t_2, \ldots, t_n$ representa una muestra aleatoria de una variable aleatoria con una función de densidad dada $f(t)$. También se podría utilizar una prueba de Kolmogorov-Smirnov (véase Law y Kelton (1990)).

Empecemos por descomponer el conjunto de tiempos entre llegadas posibles en k categorías. Con la suposición de que $f(t)$ sí rige los tiempos entre llegadas, determinamos la cantidad de las t_i que esperaríamos que cayeran en la categoría i. A este número le llamamos e_i. Luego contamos cuántas de las t_i observadas están, en realidad, en la categoría i. A este número lo denominamos o_i. Después, usamos la fórmula siguiente para calcular el valor observado de la variable chi cuadrado, χ^2(obs):

$$\chi^2(\text{obs}) = \sum_{i=1}^{i=k} \frac{(o_i - e_i)^2}{e_i}$$

El valor de χ^2(obs) sigue una distribución chi cuadrado, con $k - 2$ grados de libertad. Los puntos importantes del percentil de la distribución chi cuadrado se dan en la tabla 9.

Si χ^2(obs) es pequeña, es razonable suponer que las t_i son muestras de una variable aleatoria cuya función de densidad es $f(t)$. (Después de todo, un ajuste perfecto tendría $o_i = e_i$ para $i = 1, 2, \ldots, k$, que da por resultado un valor de χ^2 igual a cero.) Si χ^2(obs) es grande, es razonable suponer que las t_i no representan una muestra aleatoria de una variable aleatoria cuya densidad es $f(t)$.

Formalmente, estamos interesados en probar las hipótesis siguientes:

H_0: $t_1, t_2, \ldots, t_n$ es una muestra aleatoria de una variable aleatoria con densidad $f(t)$

H_a: $t_1, t_2, \ldots, t_n$ no es una muestra aleatoria de una variable aleatoria con función de densidad $f(t)$

Dado un valor de α (el error deseado tipo I), aceptamos H_0 si $\chi^2(\text{obs}) \leq \chi^2_{k-r-1}(\alpha)$ y aceptamos H_a si $\chi^2(\text{obs}) > \chi^2_{k-r-1}(\alpha)$. De acuerdo con la tabla 9 obtenemos $\chi^2_{k-r-1}(\alpha)$ que representa el punto en la tabla χ^2_{k-r-1} que tiene un área α a la derecha. Aquí, r es el número de parámetros que se deben estimar para especificar la distribución del tiempo entre llegadas. Para encontrar $\chi^2_{k-r-1}(\alpha)$ en Excel, escribimos simplemente la fórmula CHINV(Alpha, $k-r-1$) (en el paquete en inglés; en español es PRUEBA.CHI.INV). Por lo tanto, si los tiempos entre llegadas son exponenciales, $r = 1$, y si los tiempos entre llegadas siguen una distribución normal o una distribución Erlang, $r = 2$. Cuando se eligen los límites para las categorías k, conviene asegurar que cada e_i es por lo menos 5, $k \leq 30$, y las e_i se deben conservar tan iguales como sea posible. El uso de la prueba de chi cuadrado se ilustra en el ejemplo 16.

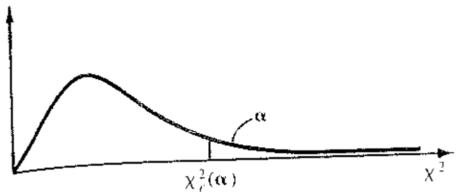
EJEMPLO 16 Tiempos entre llegadas: ¿exponencial o no exponencial?

Se han observado los tiempos entre llegadas siguientes (en minutos): 0.01, 0.07, 0.03, 0.08, 0.04, 0.10, 0.05, 0.10, 0.11, 1.17, 1.50, 0.93, 0.54, 0.19, 0.22, 0.36, 0.27, 0.46, 0.51, 0.11, 0.56, 0.72, 0.29, 0.04, 0.73. ¿Sería razonable concluir que estas observaciones provienen de una distribución exponencial?

Solución Hay 25 observaciones con $\sum_{i=1}^{i=25} t_i = 9.19$. Por lo tanto, $\bar{\lambda} = \frac{25}{9.19} = 2.72$ llegadas por minuto. Enseguida probamos si los datos son consistentes o no con una variable aleatoria

TABLA 9

Distribución de percentil de chi cuadrado



g.l. ν	α .990	.950	.900	.500	.100	.050	.025	.010	.005
1	.0002	.004	.02	.45	2.71	3.84	5.02	6.63	7.88
2	.02	.10	.21	1.39	4.61	5.99	7.38	9.21	10.60
3	.11	.35	.58	2.37	6.25	7.81	9.35	11.34	12.84
4	.30	.71	1.06	3.36	7.78†	9.49	11.14	13.28	14.86
5	.55	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	.87	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	1.24	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.65	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.95
9	2.09	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.56	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	3.05	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.57	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	4.11	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.66	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	5.23	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.81	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	6.41	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	7.01	9.39	10.86	17.34	25.99	28.87	31.53	34.81	37.16
19	7.63	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	8.26	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.90	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	9.54	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	10.20	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	10.86	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	11.52	14.61	16.47	24.34	34.38	37.65	40.65	44.31	46.93
26	12.20	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	12.88	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.64
28	13.56	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	14.26	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	14.95	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	22.16	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	29.71	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	37.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	45.44	51.74	55.33	69.33	85.53	90.53	95.02	100.43	104.21
80	53.54	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	61.75	69.13	73.29	89.33	107.57	113.15	118.14	124.12	128.30
100	70.06	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

Fuente: Richard A. Johnson y Dean W. Wichern, *Applied Multivariate Statistical Analysis*. ©1982, p. 583. Reimpreso con autorización de Prentice Hall, Inc., Englewood Cliffs, Nueva Jersey.

†*Nota*: por ejemplo, $P(\chi_4^2 > 7.78) = .10$.

exponencial (llámela **A**) cuya densidad es $f(t) = 2.72e^{-2.72t}$. Escogemos cinco categorías para asegurar que la probabilidad de que una observación proveniente de **A** se encuentra en cada uno de las cinco categorías es .20. Esto da $e_i = 25(.20) = 5$ para cada categoría. Para establecer los límites de las categorías es necesario determinar la función acumulativa, $F(t)$, para **A**:

$$F(t) = P(\mathbf{A} \leq t) = \int_0^t 2.72e^{-2.72s}\,ds = 1 - e^{-2.72t}$$

Luego escogemos las categorías como se indica enseguida:

Categoría 1 $0 \leq t < m_1$ minutos

Categoría 2 $m_1 \leq t < m_2$ minutos

Categoría 3 $m_2 \leq t < m_3$ minutos

Categoría 4 $m_3 \leq t < m_4$ minutos

Categoría 5 $m_4 \leq t$ minutos

donde $F(m_1) = .20$, $F(m_2) = .40$, $F(m_3) = .60$ y $F(m_4) = .80$.

Como $F(t) = 1 - e^{-2.72t}$, entonces, para cualquier número p, el valor de t que satisface $F(t) = p$ se podría determinar como sigue:

$$1 - e^{-2.72t} = p$$
$$1 - p = e^{-2.72t}$$

Al obtener los logaritmos (base e) de ambos miembros se obtiene

$$t = \frac{\ln(1 - p)}{-2.72}$$

$$m_1 = \frac{\ln .80}{-2.72} = 0.08$$

$$m_2 = \frac{\ln .60}{-2.72} = 0.19$$

$$m_3 = \frac{\ln .40}{-2.72} = 0.34$$

$$m_4 = \frac{\ln .20}{-2.72} = 0.59$$

Por lo tanto, las categorías son las siguientes:

Categoría 1 $0 \leq t < 0.08$ minuto

Categoría 2 $0.08 \leq t < 0.19$ minuto

Categoría 3 $0.19 \leq t < 0.34$ minuto

Categoría 4 $0.34 \leq t < 0.59$ minuto

Categoría 5 $0.59 \leq t$

Después de clasificar los datos en estas categorías, observamos que $o_1 = 6$, $o_2 = 5$, $o_3 = 4$, $o_4 = 5$ y $o_5 = 5$. Mediante la construcción de las categorías, $e_1 = e_2 = e_3 = e_4 = e_5 = .20(25) = 5$. Enseguida calculamos $\chi^2(\text{obs})$:

$$\chi^2(\text{obs}) = \frac{(6-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(4-5)^2}{5} + \frac{(5-5)^2}{5} + \frac{(5-5)^2}{5}$$
$$= .20 + 0 + .20 + 0 + 0 = .40$$

Elegimos arbitrariamente $\alpha = .05$. Como pretendemos ajustar una distribución exponencial a tiempos entre llegadas, $r = 1$. Entonces $\chi^2_3(.05) = 7.81$, y vemos que para $\alpha = .05$, podemos aceptar la hipótesis de que los tiempos entre llegadas provienen de una distribución exponencial con $\lambda = 2.72$ llegadas por minuto.

Otra opción es determinar el punto límite para la prueba de chi cuadrado con la fórmula

=CHINV(.05,3)

Con esta fórmula se obtiene el valor 7.81.

Con el objeto de probar si los tiempos de servicio están distribuidos en forma exponencial, aplicamos simplemente el mismo procedimiento a los tiempos de servicio observados $s_1, s_2, \ldots, s_n$. Empezamos por obtener una estimación (llamémosla $\hat{\mu}$) de la tasa de servicio real μ a partir de

$$\hat{\mu} = \frac{n}{s_1 + s_2 + \cdots + s_n}$$

Luego aplique la prueba chi cuadrado para ver si es razonable o no suponer que los tiempos de servicio observados provienen de una distribución exponencial con densidad $\hat{\mu}e^{-\hat{\mu}t}$.

PROBLEMAS

Grupo A

1 Una agencia de viajes quiere determinar si la duración de las llamadas telefónicas de los clientes se pueden modelar en forma adecuada con una distribución exponencial. La semana pasada la agencia registró la duración de todas las llamadas y obtuvo los resultados siguientes (en segundos): 4, 6, 5, 8, 9, 10, 12, 8, 16, 20, 24, 27, 33, 37, 43, 50, 58, 68, 70, 78, 88, 100, 120, 130. ¿Estos datos indican que la duración de las llamadas telefónicas está regida por una distribución exponencial?

20.13 Redes cerradas de líneas de espera

En cuanto a la producción de unidades que pretende establecer la producción justo a tiempo, tiene sentido mantener un nivel constante de trabajo en proceso. Por lo que se refiere a una red de computadoras ocupada, sería conveniente suponer que tan pronto como un trabajo deja el sistema, otro trabajo llega para reemplazarlo. Estos sistemas de producción y computadoras, donde hay una cantidad constante de trabajos presente, se podría modelar como **redes cerradas de líneas de espera**. Recuerde que, en una red abierta de líneas de espera, las cantidades de trabajos en cada servidor son variables aleatorias independientes. Como la cantidad de trabajos en el sistema siempre es constante, la distribución de trabajos en servidores distintos no puede ser independiente. Enseguida se estudia el **algoritmo de Buzen**, el cual se puede utilizar para determinar probabilidades de estado estable para redes cerradas de líneas de espera.

Sea P_{ij} la probabilidad de que un trabajo irá al servidor j después de completar el servicio en la estación i. Sea P la matriz cuyo dato $(i - j)$-ésimo es P_{ij}. Supongamos que los tiempos de servicio en el servidor j siguen una distribución exponencial con parámetro μ_j. El sistema tiene s servidores y, en todo momento, están presentes exactamente N trabajos. Sea n_i la cantidad de trabajos presentes en el servidor i. Entonces el estado del sistema en cualquier tiempo dado se puede definir mediante un vector n-dimensional $\mathbf{n} = (n_1, n_2, \ldots, n_s)$. El conjunto de estados posibles está dado por $S_N = \{\mathbf{n}$ tal que todos $n_i \geq 0$ y $n_1 + n_2 + \cdots + n_s = N\}$.

Sea λ_j igual a la tasa de llegadas para el servidor j. Como no hay llegadas externas, podríamos hacer todas las $r_j = 0$ y obtener los valores de las λ_j a partir de la ecuación usada en el caso de las redes abiertas. Es decir,

$$\lambda_j = \sum_{i=1}^{i=s} \lambda_i P_{ij} \qquad j = 1, 2, \ldots, s \tag{59}$$

Como los trabajos nunca dejan el sistema, por cada i, $\sum_{j=1}^{j=s} P_{ij} = 1$. Este hecho ocasiona que la ecuación (59) no tenga solución única. Por fortuna, resulta que podemos usar cualquier solución en (59) para ayudarnos a obtener las probabilidades de estado estable. Si definimos

$$\rho_i = \frac{\lambda_i}{\mu_i}$$

entonces determinamos, para cualquier estado n, su probabilidad de estado estable $\Pi_N(\mathbf{n})$ a partir de la ecuación siguiente:

$$\Pi_N(\mathbf{n}) = \frac{\rho_1^{n_1}\,\rho_2^{n_2}\cdots\rho_n^{n_s}}{G(N)} \tag{60}$$

Aquí, $G(N) = \sum_{\mathbf{n}\in S_N} \rho_1^{n_1}\,\rho_2^{n_2}\cdots\rho_s^{n_s}$.

El algoritmo de Buzen proporciona una manera eficaz para determinar (en una hoja de cálculo) $G(N)$. Una vez que tenemos la distribución de probabilidades de estado estable, calculamos con facilidad otras medidas de efectividad, tal como la longitud esperada de la cola en cada servidor, y el tiempo previsto que un trabajo permanece durante cada visita a un servidor, la fracción de tiempo que un servidor está ocupado y el rendimiento de cada servidor (trabajos por segundo procesados por cada servidor).

Para obtener $G(N)$ calculamos en forma recursiva las cantidades $C_i(k)$, para $i = 1, 2, \ldots, s$ y $k = 0, 1, \ldots, N$. Iniciamos la recursión con $C_1(k) = \rho_1^k$, $k = 0, 1, \ldots, N$ y $C_i(0) = 1$, $i = 1, 2, \ldots, s$. Para otros valores de k e i, generamos los valores de $C_i(k)$ en forma recursiva por medio de la relación siguiente:

$$C_i(k) = C_{i-1}(k) + \rho_i C_i(k-1)$$

Luego se puede demostrar que $G(N) = C_s(N)$. El uso del algoritmo de Buzen se ilustra mediante el ejemplo siguiente.†

EJEMPLO 17 Sistema de producción flexible

Considere un sistema de producción flexible en el cual 10 partes están siempre en proceso. Cada parte requiere dos operaciones. Cada una empieza por la operación 1 que se efectúa en la máquina 1. Luego, con probabilidad de .75 la pieza pasa a la operación 2 en la máquina 2, y con probabilidad .25, la pieza pasa a la máquina 3 para la operación 2. Una vez que la parte completa la operación 2, sale del sistema y se le reemplaza en forma inmediata con otra pieza. Se nos proporcionan las tasas siguientes de las máquinas (el tiempo por cada operación está distribuido en forma exponencial): $\mu_1 = .25$ minutos, $\mu_2 = .48$ minutos y $\mu_3 = .08$ minutos.

a Determine la distribución de probabilidad del número de partes en cada máquina.

b Encuentre el número esperado de partes presente en cada máquina.

c ¿Qué fracción del tiempo está ocupada cada máquina?

d ¿Cuántas partes por minuto se terminan en cada máquina?

Buzen.xls **Solución** Este problema está en el archivo Buzen.xls. Empezamos con el cálculo de una solución para las ecuaciones (59) que definen λ_1, λ_2 y λ_3. Debemos resolver

$$\lambda_1 = \lambda_2 + \lambda_3$$
$$\lambda_2 = .75\lambda_1$$
$$\lambda_3 = .25\lambda_1$$

†Tomado de Kao (1996).

Hay un número infinito de soluciones para este sistema. Elegimos en forma arbitraria $\lambda_1 = 1$, lo cual genera la solución $\lambda_2 = .75$ y $\lambda_3 = .25$. En las celdas G8:I8 calculamos $\rho_i = \frac{\lambda_i}{\mu_i}$. En G10:G20 calculamos $C_1(k) = \rho_1^k, k = 0, 1, \ldots, 10$, y en G10:I10 escribimos $C_i(0) = 1, i = 1, 2, 3$. Al copiar desde H11 a H11:I20 la fórmula

=G11+H$8*H10

se desarrolla la recursión $C_i(k) = C_{i-1}(k) + \rho_i C_i(k-1)$. Entonces podemos ya determinar G(10) = 7 231 883 a partir del valor de $C_3(10)$ en la celda H20. Véase la figura 34.

Ahora ya podemos generar todos los estados del sistema posibles en forma efectiva iniciando con $n_1 = 0$ y listarlos según el orden creciente del valor de n_2. Luego incrementamos el valor de n_1 a 1 y listamos todos los estados según los valores crecientes de n_2, etc. Una vez que tenemos $n_1 = 10$ habremos listado todos los estados. (Véase figura 35.) Para generar en forma eficaz todos los estados posibles, copiamos de C25 la fórmula (todas las funciones en este problema están como en el paquete Excel en inglés)

=IF(D25=0,B25+1,B25)

Esta fórmula incrementa n_1 en una unidad si $n_3 = 0$ (lo cual es lo mismo que tener $n_2 = 10 - n_1$). En otras circunstancias, la fórmula mantiene n_1 constante.

Luego copiamos de D25 la fórmula

=IF(B25-B24=1,0,C24+1)

Esta fórmula hace $n_2 = 0$ si ya incrementamos el valor de n_1; si no es así, la fórmula incrementa el valor de n_2 una unidad.

Por último, según E25, copiamos la fórmula

=10-B24-C24

Esto asegura que $n_3 = 10 - n_1 - n_2$.

En E24:E89 usamos (60) para calcular la probabilidad de estado estable para cada estado copiando la fórmula siguiente desde E24 a E25:E89

=($G$8^B24)*($H$8^C24)*($I$8^D24)/$I$20

Parte (a) Damos luego respuesta al inciso (a) determinando la distribución de probabilidades del número de partes en cada máquina. Utilizamos la función SUMIF (SUMAR.SI en el paquete en español) y una tabla de datos para lograr este objetivo. Para empezar calculamos la probabilidad de 0 partes en la máquina 1 en H24 con la fórmula

=SUMIF($B$24:$B$89,I23,E24:E89)

Esta fórmula suma los números de la columna D (la cual contiene las probabilidades del estado) para los renglones en los cuales la columna B (que es parte en la máquina 1) tiene una entrada 0. Véase figura 36.

	F	G	H	I
7	Mui	0.25	0.48	0.08
8	phoi	4	1.5625	3.125
9		1	2	3
10	0	1	1	1
11	1	4	5.5625	8.6875
12	2	16	24.6914063	51.83984
13	3	64	102.580322	264.5798
14	4	256	416.281754	1243.094
15	5	1024	1674.44024	5559.108
16	6	4096	6712.31287	24084.53
17	7	16384	26871.9889	102136.1
18	8	65536	107523.483	426698.9
19	9	262144	430149.442	1763583
20	10	1048576	1720684.5	7231883

FIGURA 34

	B	C	D	E
23	Parts at 1	Parts at 2	Parts at 3	Probability
24	0	0	10	0.01228143
25	0	1	9	0.00614071
26	0	2	8	0.00307036
27	0	3	7	0.00153518
28	0	4	6	0.00076759
29	0	5	5	0.00038379
30	0	6	4	0.0001919
31	0	7	3	9.5949E-05
32	0	8	2	4.7974E-05
33	0	9	1	2.3987E-05
34	0	10	0	1.1994E-05
35	1	0	9	0.01572023
36	1	1	8	0.00786011
37	1	2	7	0.00393006
38	1	3	6	0.00196503
39	1	4	5	0.00098251
40	1	5	4	0.00049126
41	1	6	3	0.00024563
42	1	7	2	0.00012281
43	1	8	1	6.1407E-05
44	1	9	0	3.0704E-05
45	2	0	8	0.02012189
46	2	1	7	0.01006094
47	2	2	6	0.00503047
48	2	3	5	0.00251524
49	2	4	4	0.00125762
50	2	5	3	0.00062881
51	2	6	2	0.0003144
52	2	7	1	0.0001572
53	2	8	0	7.8601E-05
54	3	0	7	0.02575602
55	3	1	6	0.01287801
56	3	2	5	0.006439
57	3	3	4	0.0032195
58	3	4	3	0.00160975
59	3	5	2	0.00080488
60	3	6	1	0.00040244
61	3	7	0	0.00020122
62	4	0	6	0.0329677
63	4	1	5	0.01648385
64	4	2	4	0.00824193
65	4	3	3	0.00412096
66	4	4	2	0.00206048
67	4	5	1	0.00103024
68	4	6	0	0.00051512
69	5	0	5	0.04219866
70	5	1	4	0.02109933
71	5	2	3	0.01054967
72	5	3	2	0.00527483
73	5	4	1	0.00263742
74	5	5	0	0.00131871
75	6	0	4	0.05401429
76	6	1	3	0.02700714
77	6	2	2	0.01350357
78	6	3	1	0.00675179
79	6	4	0	0.00337589
80	7	0	3	0.06913829
81	7	1	2	0.03456914
82	7	2	1	0.01728457
83	7	3	0	0.00864229
84	8	0	2	0.08849701
85	8	1	1	0.0442485
86	8	2	0	0.02212425
87	9	0	1	0.11327617
88	9	1	0	0.05663809
89	10	0	0	0.1449935

FIGURA 35

Al seleccionar el intervalo de la tabla G24:H35 y la celda I23 de las entradas de la columna nos permite hacer un bucle completo y calcular las probabilidades de estado estable para cada número de partes en la máquina 1. De modo similar obtenemos las distribuciones de probabilidad de estado estable para las máquinas 2 y 3. Véase la figura 37.

Parte (b) El número medio de partes presente en la máquina 1 se podría calcular como $\sum_{i=0}^{i=10} i*$ (probabilidad de que i partes estén en la máquina 1). En la celda K31 calculamos el número medio de partes en la máquina 1 con la fórmula

=SUMPRODUCT(G25:G35,H25:H35)

De manera similar calculamos el número medio de partes en las máquinas 2 y 3 en las celdas K32 y K33. Véase figura 38. Obsérvese que la máquina 1 es evidentemente el cuello de botella.

Parte (c) Para calcular la probabilidad de que cada máquina esté ocupada, sólo efectuamos la diferencia entre 1 y la probabilidad de cada máquina tiene 0 partes. Estos cálculos se realizan en L31:L33. Así llegamos a saber que la máquina 1 está ocupada 97% del tiempo, la máquina 2, 38% del tiempo y la máquina 3, 76% del tiempo.

	G	H	I
21			
22			Parts
23		Prob	0
24	Machine 1 parts	0.02455086	
25	0	0.02455086	
26	1	0.03140975	
27	2	0.04016518	
28	3	0.05131082	
29	4	0.06542029	
30	5	0.08307862	
31	6	0.10465268	
32	7	0.12963429	
33	8	0.15486976	
34	9	0.16991426	
35	10	0.1449935	

FIGURA 36

	G	H
37	Machine 2 Parts	0.61896518
38	0	0.61896518
39	1	0.23698584
40	2	0.09017388
41	3	0.03402481
42	4	0.01269126
43	5	0.00465769
44	6	0.00166949
45	7	0.00057718
46	8	0.00018798
47	9	5.4691E-05
48	10	1.1994E-05

	G	H
50	Machine 3 Parts	0.23793036
51	0	0.23793036
52	1	0.18587372
53	2	0.14519511
54	3	0.1133962
55	4	0.08851582
56	5	0.06900306
57	6	0.0536088
58	7	0.0412822
59	8	0.03105236
60	9	0.02186094
61	10	0.01228143

FIGURA 37

	J	K	L	M
29				
30	Mean Number	Mean Number	Prob busy	Completions per second
31	Machine 1	6.696224299	0.97544914	0.243862285
32	Machine 2	0.609634749	0.38103482	0.182896714
33	Machine 3	2.694140952	0.76206964	0.060965571

FIGURA 38

Parte (d) Para calcular el número medio de servicios completados por minuto en cada máquina, simplemente multiplicamos la probabilidad de que una máquina esté ocupada por la tasa de servicio de la máquina. Estos cálculos se efectúan en M31:M33. Encontramos que la máquina 1 completa, en promedio, .24 partes/minuto, la máquina 2, .18 partes/minuto, y la máquina 3, .06 partes/minuto.

PROBLEMAS

Grupo A

1 Los trabajos llegan a un servidor que consta de una unidad central de proceso (CPU) y dos discos (disco 1 y disco 2). La probabilidad de que un trabajo vaya desde CPU al disco 1 es 13/20, y de que vaya al disco 2 es 6/20. La probabilidad de que un trabajo esté terminado después de la operación en la CPU y que sea reemplazado inmediatamente por otro trabajo es 1/20. Hay tres trabajos en el sistema. El tiempo medio para completar la operación en la CPU es .039 s. El tiempo medio para completar la operación en el disco 1 es .18 s y el tiempo medio para terminar la operación en el disco 2 es .26 s.

a Determine la distribución de estado estable del número de trabajos en cada parte del sistema.

b ¿Cuál es el número promedio de trabajos en CPU? ¿Y en el disco 1? ¿Y en el disco 2?

c ¿Cuál es la probabilidad de que la CPU esté ocupada? ¿Y el disco 1? ¿Y el disco 2?

d ¿Cuál es el número promedio de trabajos que completa por segundo la CPU? ¿Y el disco 1? ¿Y el disco 2?

2 Un proceso de manufactura siempre tiene 8 partes en proceso. Una parte debe completar con éxito dos pasos (paso 1 y paso 2). Una sola máquina ejecuta el paso 1 y puede procesar un promedio de 8 partes por minuto. Una sola máquina hace el paso 2 y tiene la aptitud de procesar 11 partes por minuto. Infortunadamente, el paso 2 no es del todo confiable. (El paso 1 sí es por completo confiable.) Cada vez que una pieza es enviada por el paso 2 hay un 10% de probabilidades de que el paso 2 tenga que ser repetido.

a Encuentre la distribución de estado estable de las partes en cada máquina.

b Estime el número promedio de partes en cada máquina.

c Determine la probabilidad de que todas las máquinas estén ocupadas.

d Dé el número de partes por minuto que completan con éxito su servicio en cada máquina.

20.14 Una aproximación para el sistema de líneas de espera *G/G/m*

ggm.xls

En la mayoría de las situaciones, los tiempos entre llegadas siguen una variable aleatoria exponencial. (Véase una explicación de este hecho en Denardo (1982).) Pero los tiempos de servicio no se apegan, a menudo, a una distribución exponencial. Cuando los tiempos entre llegadas y los tiempos de servicio no se apegan a una variable aleatoria no exponencial, el sistema de líneas de espera se llama *G/G/m*. La primera *G* indica que los tiempos entre llegadas siempre siguen la misma variable aleatoria (pero no necesariamente exponencial), y la segunda *G* significa que los tiempos de servicio siempre siguen la misma variable aleatoria (pero no necesariamente exponencial). Para estas situaciones, no funcionan las plantillas estudiadas en las secciones anteriores de este capítulo. Por fortuna, la aproximación de Allen-Cunneen (véase Tanner (1995)) proporciona con frecuencia un buen valor de L, W, L_q y W_q para los sistemas *G/G/m*. En el archivo ggm.xls se pone en práctica la aproximación de Allen-Cunneen. El usuario requiere escribir sólo la información siguiente:

- El número promedio de llegadas por unidad de tiempo (λ) en la celda B3.
- La tasa promedio a la cual se atiende a los clientes (μ) en la celda B4.
- El número de servidores (s) en la celda B5.
- El coeficiente cuadrado de la variación –(varianza de tiempos entre llegadas)/(tiempo entre llegadas promedio)^2– de los tiempos entre llegadas en la celda B6.
- El coeficiente cuadrado de la variación –(varianza de tiempos de servicio)/tiempo promedio de servicio)^2– de los tiempos de servicio en la celda B7.

El coeficiente cuadrado de la variación para los tiempos entre llegadas o de servicio se puede estimar con facilidad mediante las funciones de Excel =AVERAGE y =VARP. (PROMEDIO y VARP, en el paquete de Excel en español) Recuerde que la variable aleatoria exponencial tiene la propiedad de que varianza = media^2. Por lo tanto, el coeficiente cuadrático de la variación para tiempos entre llegadas exponenciales o tiempos de servicio será igual a 1, y la diferencia respecto a la unidad del coeficiente cuadrático de la variación para tiempos de entre llegadas o tiempos de servicio indica el grado de desviación de la exponencialidad. La aproximación de Allen-Cunneen es exacta si los tiempos entre llegadas y los tiempos de servicio son exponenciales. Las extensas pruebas de Tanner indican que, en una gran diversidad de situaciones, los valores de L, W, L_q y W_q obtenidos mediante la aproximación están a 10% de sus valores verdaderos. Sigue una ilustración de la aproximación de Allen-Cunneen.

EJEMPLO 18 Banco NBD

La sucursal del banco NBD en Bloomington, Indiana tiene seis cajeros. En las horas de afluencia máxima, llega un promedio de 4.8 clientes por minuto al banco. Un cajero requiere un promedio de un minuto atender a un cliente. El coeficiente cuadrático de variación tanto para los tiempos de entre llegadas y tiempos de servicio es .5. Estime el tiempo promedio que un cliente tendrá que esperar antes de ver a un cajero. ¿Cuántos clientes estarán presentes en promedio?

Solución Después de escribir la información pertinente en las celdas B3 a B7 (véase figura 39), encontramos que un cliente espera, en promedio, .216 minutos antes de ver al cajero. En promedio, 5.83 clientes estarán presentes en el banco. Al parecer, la afluencia está bajo control. Este resultado favorable se debe en gran medida al bajo coeficiente cuadrático de variación tanto para el tiempo entre llegadas como para el tiempo de servicio. Por ejemplo, si ambos coeficientes cuadrados de variación fueran 4, entonces W_q sería 1.73 minutos, lo que significa un 800% de aumento.

	A	B	C	D
1		G/G/m Template		
2		Allen-Cunneen Approximation		
3	Lambda	4.8		
4	Mu	1		
5	s	6		
6	CV arrive	0.5		
7	CV service	0.5		
8	u	4.8		
9	ro	0.8		
10	R(s,mu)	0.82322		
11	E_c(s,mu)	0.517772		
12	W_q	0.215738		
13	L_q	1.035544		
14	W	1.215738		
15	L	5.835544		

FIGURA 39

PROBLEMAS

Grupo A

Los problemas 1 a 4 se refieren al ejemplo 18.

1 NBD opina que el nivel de afluencia es satisfactorio si el número promedio de clientes en la cola es igual al número de servidores. De acuerdo con la información proporcionada en el ejemplo, ¿cuál es la tasa de llegadas máxima que pueden manejar en forma satisfactoria los seis servidores?

2 Demuestre qué tanto depende del número de servidores el tiempo promedio que un cliente debe esperar un cajero.

3 Determine, mediante el uso de tablas de datos, cómo afectan los cambios en el coeficiente cuadrático de variación para tiempos entre llegadas y de servicio el número promedio de clientes en la sucursal de NBD.

4 Suponga que un cajero cuesta 30 dólares por hora. Suponga que el banco evalúa el tiempo de un cajero en NBD en c dólares por hora. Muestre cómo las variaciones en c afectan el número de cajeros que NBD debe utilizar.

5 Un promedio de 230 clientes por hora llegan a un controlador de boletos de Southbest Airlines, en donde trabajan ocho agentes. Cada agente puede atender en promedio a 30 clientes por hora. El coeficiente cuadrático de variación para los tiempos entre llegadas es 1.5 y 2 para los tiempos de servicio.

a ¿Cuántos clientes, en promedio, estarán presentes en el controlador de boletos?

b ¿Cuánto tendrá que esperar, en promedio, un cliente a un agente?

20.15 Modelos de líneas de espera con preferencias†

Hay muchas situaciones en las cuales los clientes no son atendidos conforme van llegando (FCFS). En la sección 20.1 se estudia también el servicio según las disciplinas de líneas de espera en orden aleatorio, SIRO, y según LCFS. Sean $\mathbf{W}_{FCFS}$, $\mathbf{W}_{SIRO}$ y $\mathbf{W}_{LCFS}$ las variables aleatorias que representan el tiempo de espera de un cliente en los sistemas de líneas de espera según las disciplinas FCFS, SIRO y LCFS, respectivamente. Se puede demostrar que

$$E(\mathbf{W}_{FCFS}) = E(\mathbf{W}_{SIRO}) = E(\mathbf{W}_{LCFS})$$

Por lo tanto, el tiempo promedio (estado estable) que un cliente pasa en el sistema no depende de cuál de las tres disciplinas se escoja. También se puede demostrar que

$$\text{var } \mathbf{W}_{FCFS} < \text{var } \mathbf{W}_{SIRO} < \text{var } \mathbf{W}_{LCFS} \tag{61}$$

Como una varianza grande se relaciona por lo regular con una variable aleatoria que tiene una probabilidad relativamente grande de asumir valores extremos, la ecuación (61) indica que lo más probable es que ocurran tiempos de espera relativamente grandes con una disciplina LCFS y menos probable que se presenten con una FCFS. Es razonable porque en un sistema LCFS, un cliente puede tener suerte e ingresar en forma inmediata al servicio, pero también puede ser lanzado al final de la cola. No obstante, el cliente no puede ser enviado al final de una larga cola en FCFS, de modo que una espera muy larga es relativamente improbable.

En muchas compañías, el orden en el cual se atiende a los clientes depende del "tipo" de cliente. Por ejemplo, en las salas de urgencias de los hospitales se atiende a los pacientes muy graves antes que a los otros pacientes. Asimismo, en muchos sistemas computarizados, los trabajos más grandes no entran al servicio antes de que todos los trabajos más pequeños que están en la cola hayan sido completados. Los modelos en los cuales un tipo de cliente determina el orden en el cual los clientes pasan al servicio se denominan **modelos de líneas de espera con preferencias.**

El siguiente escenario abarca diversos modelos de líneas de espera que se rigen según prioridades (sin olvidar todos los modelos que se estudian en esta sección. Suponga que hay n tipos de clientes (tipo 1, tipo 2, . . . , tipo n). Los tiempos entre llegadas de los clientes tipo i están distribuidos en forma exponencial con una tasa λ_i. Se supone que los tiempos entre llegadas de tipos distintos de clientes son independientes. El tiempo de servicio de un cliente tipo i se representa mediante una variable aleatoria $\mathbf{S}_i$ (no necesariamente exponencial). Se supone que los tipos de clientes de número más bajo tienen prioridad sobre los tipos de clientes de número superior.

†Esta sección trata temas que se podrían omitir sin perder la continuidad.

Modelos con preferencias sin prioridades

Al inicio se tratan los modelos con preferencias sin prioridades (*Nonpreemptive Priority Models*). En un modelo sin prioridades, no se puede interrumpir el servicio de un cliente. Después de que cada servicio se completa, el siguiente cliente en entrar al servicio se elige dando prioridad a los tipos de clientes de número más bajo (sin tomar en cuenta FCFS). Por ejemplo, si $n = 3$ y están presentes tres clientes tipo 2 y cuatro tipo 3, el siguiente cliente en pasar al servicio sería el cliente tipo 2 que, entre los de su tipo, fue el primero en llegar.

Según la notación de Kendall-Lee, un modelo con preferencias sin prioridades se representa poniendo en la cuarta característica NPRP. Para indicar los diversos tipos de clientes, se escribe un subíndice i en las primeras dos características. Por lo tanto, $M_i/G_i/\cdots$ representaría una situación en la cual los tiempos entre llegadas para el cliente tipo i-ésimo son exponenciales y los tiempos de servicio para el cliente tipo i-ésimo tienen una distribución general. En lo que sigue, sean

W_{qk} = Tiempo de espera en estado estable que pasa un cliente tipo k en la cola.
W_k = Tiempo esperado en estado estable en el sistema que pasa un cliente tipo k en la cola.
L_{qk} = Número esperado en estado estable de clientes tipo k que esperan en la cola.
L_k = Número esperado en estado estable de clientes tipo k en el sistema.

Modelo $M_i/G_i/1/\text{NPRP}/\infty/\infty$

El primer resultado se relaciona con el sistema de un solo servidor, $M_i/G_i/1/\text{NPRP}/\infty/\infty$ sin prioridades. Definamos $\rho_i = \frac{\lambda_i}{\mu_i}$, $a_0 = 0$ y $a_k = \sum_{i=1}^{i=k} \rho_i$. Suponemos† que

$$\sum_{i=1}^{i=n} \frac{\lambda_i}{\mu_i} < 1$$

Entonces,

$$\begin{aligned} W_{qk} &= \frac{\sum_{k=1}^{k=n} \lambda_k E(S_k^2)/2}{(1 - a_{k-1})(1 - a_k)} \\ L_{qk} &= \lambda_k W_{qk} \\ W_k &= W_{qk} + \frac{1}{\mu_k} \\ L_k &= \lambda_k W_k \end{aligned} \tag{62}$$

En el ejemplo siguiente se ilustra el uso de la ecuación (62).

EJEMPLO 19 Preferencia en el servicio de copiado

Un centro de copiado da preferencia a los trabajos pequeños y no a los grandes. Los tiempos entre llegadas para cada tipo de trabajo son exponenciales, y llega cada hora un promedio de 12 trabajos pequeños y seis trabajos grandes. Sea trabajo tipo 1 = trabajo pequeño y trabajo tipo 2 = trabajo grande. Luego nos indican que

$$E(S_1) = 2 \text{ minutos} \qquad E(S_1^2) = 6 \text{ minutos}^2 = \frac{1}{600} \text{ hora}^2$$

$$E(S_2) = 4 \text{ minutos} \qquad E(S_2^2) = 18 \text{ minutos}^2 = \frac{1}{200} \text{ hora}^2$$

Determine el tiempo promedio que cada tipo de trabajo pasa en el centro de copiado.

†Si esta condición no se cumple, entonces para uno o más tipos de clientes, no existe tiempo de espera en estado estable.

Solución Sabemos que $\lambda_1 = 12$ trabajos por hora, $\lambda_2 = 6$ trabajos por hora, $\mu_1 = 30$ trabajos por hora y $\mu_2 = 15$ trabajos por hora. Luego $\rho_1 = \frac{12}{30} = .4$ y $\rho_2 = \frac{6}{15} = .4$. Como $\rho_1 + \rho_2 < 1$, existe un estado estable. Ahora, $a_0 = 0$, $a_1 = .4$ y $a_2 = .4 + .4 = .8$. Entonces, la ecuación (62) genera,

$$W_{q1} = \frac{\dfrac{12\left(\dfrac{1}{600}\right)}{2} + \dfrac{6\left(\dfrac{1}{200}\right)}{2}}{(1-0)(1-.4)} = \frac{\dfrac{30}{1\,200}}{.6} = 0.042 \text{ hora}$$

$$W_{q2} = \frac{\dfrac{12\left(\dfrac{1}{600}\right)}{2} + \dfrac{6\left(\dfrac{1}{200}\right)}{2}}{(1-.4)(1-.8)} = \frac{\dfrac{30}{1\,200}}{.12} = 0.208 \text{ hora}$$

Asimismo,

$$W_1 = W_{q1} + \frac{1}{\mu_1} = 0.042 + 0.033 = 0.075 \text{ hora}$$

$$W_2 = W_{q2} + \frac{1}{\mu_2} = 0.208 + 0.067 = 0.275 \text{ hora}$$

Por lo tanto, como se esperaba, los trabajos grandes pasan más tiempo en el centro de copiado que los trabajos pequeños.

Modelo $M_i/G_i/1/NPRP/\infty/\infty$ con costos de espera que dependen del cliente

Considere un sistema con preferencias, sin prioridades y con un solo servidor, en el cual se genera un costo c_k por cada unidad de tiempo que un cliente tipo k pasa en el sistema. Si deseamos minimizar el costo esperado que se genera por unidad de tiempo (en el estado estable), ¿qué orden de preferencias debe haber en los tipos de clientes? Suponga que n clientes están numerados como se indica

$$c_1\mu_1 \geq c_2\mu_2 \geq \cdots \geq c_n\mu_n \tag{63}$$

Entonces, el costo esperado se minimiza dando la preferencia más alta a los clientes tipo 1, la preferencia siguiente a los clientes tipo 2, y así sucesivamente, y la preferencia menor a los clientes tipo n. Para ver por qué este orden de preferencia es razonable, observe que cuando un cliente tipo k está recibiendo atención, el costo de salida del sistema a una tasa $c_k\mu_k$. Por lo tanto, el costo se puede minimizar dando la preferencia mayor a los tipos de clientes con los valores más grandes de $c_k\mu_k$.

Como un caso especial de este resultado, suponga que deseamos minimizar L, el número esperado de trabajos en el sistema. Sea $c_1 = c_2 = \cdots = c_n = 1$. Entonces, en cualquier momento, el costo por unidad de tiempo es igual al número de clientes en el sistema. Por lo tanto, el costo esperado por unidad de tiempo es igual a L. Ahora (63) se vuelve

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \qquad \text{o bien,} \qquad \frac{1}{\mu_1} \leq \frac{1}{\mu_2} \leq \cdots \leq \frac{1}{\mu_n}$$

Por lo tanto, podemos concluir que el número esperado de trabajos en el sistema se minimiza si se da la preferencia más alta a los tipos de clientes con el tiempo de servicio medio más pequeño. Esta disciplina de preferencia se conoce como disciplina de tiempo de proceso más corto (*shortest processing time*, SPT).

Modelo $M_i/M/s$/NPRP/∞/∞

Si deseamos obtener resultados analíticos manejables en los sistemas de preferencia con varios servidores, debemos suponer que los tiempos de servicio de cada tipo de cliente se apegan a una distribución exponencial con media $\frac{1}{\mu}$, y que los tiempos entre llegadas de los clientes tipo i siguen una distribución exponencial con tasa λ_i. Dicho sistema con s servidores se representa con la notación $M_i/M/s$/NPRP/∞/∞. Para este modelo,

$$W_{qk} = \frac{P(j \geq s)}{s\mu(1 - a_{k-1})(1 - a_k)} \tag{64}$$

En (64)

$$a_k = \sum_{i=1}^{i=k} \frac{\lambda_i}{s\mu} \qquad (k \geq 1)$$

$a_0 = 0$ y $P(j \geq s)$ se obtiene de la tabla 6 para un sistema de s servidores con

$$\rho = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_n}{s\mu}$$

El uso de la ecuación (64) se ilustra en el ejemplo 20.

EJEMPLO 20 Respuesta de la policía

La municipalidad de Gotham tiene cinco patrullas. El departamento de policía recibe dos tipos de llamadas: urgentes (tipo 1) y no urgentes (tipo 2). Los tiempos entre llegadas para cada tipo de llamada están distribuidos en forma exponencial, con un promedio de 10 llamadas urgentes y 20 no urgentes que se reciben cada hora. Cada tipo de llamada tiene un tiempo de servicio exponencial, con una media de 8 minutos (suponga que, en promedio, 6 de los 8 minutos es el recorrido desde la estación de policía hasta la llamada y regreso a la estación). Las llamadas urgentes tienen preferencia respecto a las llamadas que no lo son. ¿Cuánto tiempo, en promedio, transcurre entre el momento que se hace la llamada no urgente y la llegada de una patrulla?

Solución Tenemos que $s = 5$, $\lambda_1 = 10$ llamadas por hora, $\lambda_2 = 20$ llamadas por hora, $\mu = 7.5$ llamadas por hora, $\rho = \frac{10+20}{5(7.5)} = .80$, $a_0 = 0$, $a_1 = \frac{10}{37.5} = .267$ y $a_2 = \frac{10+20}{37.5} = .80$. De acuerdo con la tabla 6, con $s = 5$ y $\rho = .80$, $P(j \geq 5) = .55$. Entonces, con (64) se obtiene

$$W_{q2} = \frac{.55}{5(7.5)(1 - .267)(1 - .80)} = \frac{.55}{5.50} = 0.10 \text{ hora} = 6 \text{ minutos}$$

El tiempo promedio entre el momento en que se hace la llamada no urgente y la llegada de la patrulla es $W_{q2} + (\frac{1}{2})$ (tiempo total del viaje por llamada) $= 6 + 3 = 9$ minutos.

Preferencias con prioridades

Concluimos el estudio de los sistemas de líneas de espera con preferencias con el análisis del **sistema de líneas de espera con prioridades** y un solo servidor. En un sistema de líneas de espera con prioridades, un cliente con preferencia baja (digamos un cliente tipo i) puede ser sacado del servicio siempre que llegue un cliente con preferencia superior. Cuando ya no haya presentes clientes con preferencia alta, el cliente tipo i desplazado reingresa al servicio. En un **modelo con prioridades y reanudación**, el servicio de un cliente continúa desde el punto donde fue interrumpido. En el caso de un **modelo con prioridades y repetición**, el cliente inicia de nueva cuenta el servicio cada vez que reingresa al servicio. Naturalmente, si los tiempos de servicio están distribuidos exponencialmente, las disciplinas de reanudación y de repetición son idénticas. (¿Por qué?) En la notación de Kendall-Lee, un sistema de líneas de espera con prioridades se denota poniendo en la cuarta característica las siglas PRP. Ahora consideramos un sistema $M_i/M/1/\text{PRP}/\infty/\infty$ con un solo servidor en el cual el tiempo de servicio de cada cliente es exponencial con media $\frac{1}{\mu}$ y los tiempos entre llegadas para el tipo i-ésimo del cliente siguen una distribución exponencial con tasa λ_i. Entonces,

$$W_k = \frac{\frac{1}{\mu}}{(1 - a_{k-1})(1 - a_k)} \tag{65}$$

donde $a_0 = 0$ y

$$a_k = \sum_{i=1}^{i=k} \frac{\lambda_i}{\mu}$$

Por razones obvias, las disciplinas con prioridades se usan raras veces si los clientes son personas. Algunas veces se usan con "clientes" como trabajos con computadora. Mediante el ejemplo siguiente se ilustra el uso de la ecuación (65).

EJEMPLO 21 Sistema de computadoras de la universidad

En el sistema de computadoras de la Universidad Podunk, los trabajos de los maestros (tipo 1) siempre tienen prioridad ante los trabajos de los estudiantes (tipo 2). El tiempo que requiere cada tipo de trabajo sigue una distribución exponencial, con media de 30 segundos. Cada hora llega un promedio de 10 trabajos de los maestros y 50 de los estudiantes. ¿Cuál es el tiempo promedio entre el momento en que un estudiante envía su trabajo y el momento en que se termina su trabajo? Suponga que los tiempos entre llegadas son exponenciales.

Solución Sabemos que $\mu = 2$ trabajos por minuto, $\lambda_1 = \frac{1}{6}$ de trabajo por minuto y $\lambda_2 = \frac{5}{6}$ de trabajo por minuto. Entonces,

$$a_0 = 0, \qquad a_1 = \frac{\frac{1}{6}}{2} = \frac{1}{12}, \qquad a_2 = \frac{1}{12} + \frac{\frac{5}{6}}{2} = \frac{1}{2}$$

La ecuación (65) genera

$$W_2 = \frac{\frac{1}{2}}{\left(1 - \frac{1}{12}\right)\left(1 - \frac{1}{2}\right)} = \frac{12}{11} \text{ minutos} = 1.09 \text{ minutos}$$

Transcurrirá un promedio de 1.09 minutos entre el tiempo en el que el estudiante envía el trabajo y el momento en que lo termina.

PROBLEMAS

Grupo A

1 El profesor de inglés, Jacob Bright tiene una mecanógrafa, que escribe durante ocho horas al día. Él le entrega tres tipos de trabajos: exámenes, trabajos de investigación y notas para las clases. Se cuenta con la información de la tabla 10. El profesor Bright ha indicado a la mecanógrafa que los exámenes tienen preferencia sobre los trabajos de investigación, y que los trabajos de investigación tienen preferencia sobre las notas de clase. Si se supone que el sistema es sin prioridades, determine el tiempo previsto que el profesor Bright tendrá que esperar antes de que cada tipo de trabajo esté terminado.

TABLA 10

Tipo de trabajo	Frecuencia (número por día)	$E(S_i)$ (horas)	$E(S_i^2)$ (horas)^2
Exámenes	2	1	2
Trabajo de investigación	0.5	4	20
Notas de clase	5	0.5	0.50

2 Suponga que un supermercado aplica un sistema en el cual todos los clientes forman una sola cola en la cual esperan al primer cajero desocupado. Suponga, además, que el tiempo de servicio para un cliente que compra k productos está distribuido en forma exponencial, con una media de k segundos. Asimismo, un cliente que compra k productos siente que el costo de espera en la cola es de un dólar/k por minuto. Si los clientes pueden ser asignados a preferencias, ¿qué asignación de preferencias minimizaría el costo de espera previsto por los clientes del supermercado? ¿Por qué sería una función decreciente de k el costo por minuto de la espera de los clientes?

3 Cuatro médicos trabajan en una sala de urgencias de un hospital, que atiende a tres tipos de pacientes. El tiempo que un médico gasta con cada tipo de paciente está distribuido en forma exponencial, con media de 15 minutos. Los tiempos entre llegadas de cada tipo de cliente son exponenciales, y el número promedio de llegadas por hora de cada tipo de paciente es como sigue: tipo 1, tres pacientes; tipo 2, cinco pacientes; tipo 3, tres pacientes. Suponga que los pacientes del tipo 1 tienen la mayor preferencia, y que los pacientes del tipo 3, la menor preferencia (no se permite la prioridad). ¿Cuál es el promedio de tiempo que cada tipo de paciente debe esperar antes de ver a un médico?

4 Considere un sistema de computadoras al cual se envían dos tipos de trabajos. El tiempo medio para ejecutar cada tipo de trabajo es $\frac{1}{\mu}$. Los tiempos entre llegadas por cada tipo de trabajo son exponenciales, con un promedio de λ_i trabajos tipo i que llegan cada hora. Considere las tres situaciones siguientes:

a Los trabajos tipo 1 tienen preferencia sobre los trabajos tipo 2, y se permite la prioridad.

b Los trabajos tipo 1 tienen preferencia sobre los trabajos tipo 2, y no se permite la prioridad.

c Todos los trabajos se atienden según la disciplina FCFS.

¿En cuál sistema los trabajos tipo 1 son los que están en mejores condiciones? ¿Y en las peores? Conteste las mismas preguntas para los trabajos tipo 2.

20.16 Comportamiento momentáneo de los sistemas de líneas de espera

A lo largo de todo este capítulo se ha supuesto que la tasa de llegadas, la tasa de servicio y el número de servidores permanecen constantes respecto al tiempo. Lo anterior permite hablar razonablemente acerca de la existencia de un estado estable. En muchas situaciones, podrían variar la tasa de llegadas, la tasa de servicio y el número de servidores. He aquí algunos ejemplos.

- Un restaurante que sirve bocadillos es el idóneo para experimentar una tasa de llegadas más grande durante el tiempo que va del mediodía a la 1:30 PM que durante otras horas del día. Asimismo, la cantidad de servidores (en un restaurante con servidores en paralelo) varía durante el día: habrá más servidores durante los periodos de mayor afluencia de clientela.
- Como la mayor parte de ataques al corazón se presenta durante la mañana, una unidad de atención coronaria experimentará más llegadas durante la mañana.
- La mayor parte de los electores votan antes o después de ir a trabajar, por lo que las casillas para votar estarán menos ocupadas durante el mediodía.

Cuando los parámetros que definen el sistema de líneas de espera varían respecto al tiempo, decimos que el sistema es no estacionario. Considere, por ejemplo, un restaurante de bocadillos que abre a las 10 AM y cierra a las 6 PM. Nos interesa conocer la distribución de probabilidades del número de clientes presentes en todo el tiempo entre las 10 AM y la hora de cierre. Estas distribuciones de probabilidad se denominan **probabilidades transitorias**. Por ejemplo, si queremos determinar la probabilidad de que por lo menos 10 clientes estén presentes, esta probabilidad será mayor a las 12:30 PM que a las 3 PM.

Ahora suponga que en el tiempo t, los tiempos entre llegadas son exponenciales con tasa $\lambda(t)$. Asimismo, $s(t)$ servidores están disponibles en el tiempo t, con tiempos de servicio exponenciales y tasa $\mu(t)$. Suponga, además, que el número máximo de clientes presentes en cualquier momento está dado por N. Entonces, para determinar las probabilidades momentáneas, escoja un segmento pequeño de tiempo Δt y suponga que cuando mucho puede ocurrir un evento (una llegada o la finalización de un servicio) durante un intervalo de duración Δt. Suponga que k clientes están presentes ahora en el tiempo t, y que

- La probabilidad de una llegada durante un intervalo Δt es $\lambda(t)^*(\Delta t)$.
- La probabilidad de más de una llegada durante un intervalo Δt es $o(\Delta t)$.
- Las llegadas durante intervalos distintos son independientes.
- La probabilidad de que se complete un servicio durante un intervalo Δt está dada por $(s(t), k)^* \mu t \Delta t$.
- La probabilidad de que más de un servicio se complete durante un intervalo Δt es $o(\Delta t)$.

Cuando las primeras tres suposiciones rigen las llegadas, se dice que las llegadas siguen un **proceso Poisson no homogéneo**. La suposición implica que, dadas la tasa de llegadas y la tasa de servicio, el número esperado de llegadas, la finalización del servicio, o ambas situaciones durante el Δt siguiente corresponderá con lo que esperamos. La fuente del error en nuestra aproximación es el hecho de que por lo menos dos eventos pueden ocurrir durante un tiempo Δt. La probabilidad de que esto suceda es $o(\Delta t)$, de modo que si hacemos Δt lo suficiente pequeño, la aproximación no debe ocasionar grandes errores al calcular las probabilidades momentáneas.

Ahora definamos $P_i(t)$ como la probabilidad de que i clientes estén presentes en el tiempo t. Supondremos (aunque no es necesario), que el sistema inicialmente está vacío, de modo que $P_0(0) = 1$ y para $i > 0$, $P_0(i) = 0$. Entonces, dado que conocemos $P_i(t)$, podemos estimar $P_i(t + \Delta t)$ como sigue

$$
\begin{aligned}
P_0(t + \Delta t) &= (1 - \lambda(t)\Delta t)P_0(t) + \mu(t)\Delta t P_1(t) \\
P_i(t + \Delta t) &= \lambda(t)\Delta t P_{i-1}(t) + (1 - \lambda(t)\Delta t - \min(s(t), i)\mu(t)\Delta t)P_i(t) + \min(s(t), \\
&\quad i + 1)\mu(t)\Delta t P_{i+1}(t), N - 1 \geq i \geq 1 \\
P_N(t + \Delta t) &= \lambda(t)\Delta t P_{N-1}(t) + (1 - \min(s(t), N)\mu(t))\Delta t P_N(t)
\end{aligned}
$$

Como se estableció previamente, estas ecuaciones se basan en la suposición de que si el estado es i en el momento t, entonces durante el siguiente Δt, la probabilidad de una llegada es $\lambda(t)\ \Delta t$, y la probabilidad de que se complete un servicio es min $(s(t), i)\mu(t)\Delta t$. La primera ecuación se infiere después de observar que estar en el estado 0 en el tiempo $t + \Delta t$ sólo puede suceder si estuviéramos en el estado 1 en el tiempo t y se completara un servicio durante el Δt siguiente, o bien, si estuviéramos en el estado 0 en el tiempo t y no ocurriera ninguna llegada durante el siguiente Δt. La segunda ecuación se deduce después de observar que para $N - 1 \geq i \geq 1$, sólo podemos estar en el estado i en el tiempo $t + \Delta t$ si ocurre una de las situaciones siguientes:

- Si estamos en el estado $i - 1$ en el tiempo t y hubiera una llegada durante el Δt siguiente.
- Si estamos en el estado $i - 1$ en el tiempo t y se completa un servicio durante el Δt siguiente.
- Si estamos en el estado i en el tiempo t, y no ocurre ninguna llegada ni se completa un servicio durante el Δt siguiente.

La ecuación final se deduce después de observar que para estar en el estado N en el tiempo $t + \Delta t$, debe ocurrir una de las dos situaciones siguientes:

- Estamos en el estado $N - 1$ en el tiempo t y ocurre una llegada durante el Δt siguiente.
- Estamos en el estado N en el tiempo t, y no se completa ningún servicio durante el siguiente Δt.

Mediante el ejemplo siguiente se ilustra cómo podemos usar las aproximaciones para determinar las probabilidades momentáneas en el caso de un sistema de líneas de espera no estacionario.

EJEMPLO 22 Afluencia de comensales a la hora del almuerzo

Un pequeño restaurante de bocadillos pretende modelar el ajetreo que ocurre a la hora del almuerzo. El restaurante abre a las 11 AM y todos los clientes esperan en una sola cola

TABLA 11

Tiempo	Tasa de llegadas por hora
11–11.30 AM	30
11:30 AM–mediodía	40
Mediodía–12:30 PM	50
12:30 PM–1 PM	60
1 PM–1:30 PM	35
1:30 PM–2 PM	25

para hacer su pedido. La tasa de llegadas por hora en los diferentes momentos es como se indica en la tabla 11. Las llegadas siguen un proceso Poisson. El restaurante tiene la capacidad de atender a un promedio de 50 personas por hora. Los tiempos de servicio son exponenciales. La gerencia quiere modelar la distribución de probabilidades de clientes desde las 11 AM a las 2 PM.

a Estime el número promedio de personas en la cola o en servicio a las 12:30 PM.

b Estime el número promedio de personas en la cola o en servicio a las 11:30 PM.

restaurant.xls

Solución El problema está en el archivo restaurant.xls (las funciones de Excel están como en el paquete en inglés). (Véase figura 40.) Usamos incrementos de tiempo de 5 segundos y procedemos como se indica:

Paso 1 Calculamos, en E4, la probabilidad de que se complete un servicio en 5 s al multiplicar la tasa de servicio por hora por $\Delta t = 1/720$.

Paso 2 Usamos el comando de Excel DATA FILL en la columna A para generar tiempos, con incrementos de 5 segundos, que van desde 0 hasta 10 800 (2 PM).

Paso 3 Al copiar la fórmula siguiente desde B11 hasta B11:B2171,

A11/3600

convertimos los segundos en horas.

Paso 4 Al copiar la fórmula siguiente desde C11 hasta C12:C2171

=VLOOKUP(B11,$G$2:$H$7,2)/720

buscamos la tasa de llegadas por hora para el tiempo actual y la convertimos a tasa de llegadas por 5 segundos al multiplicar la tasa de llegadas por hora por $\Delta t = 1/720$. Observe que la tasa de llegadas es en alto grado no estacionaria. Este hecho afectará en gran medida el nivel de congestión del sistema.

Paso 5 Suponemos que estará presente un máximo de $N = 30$ clientes. Por lo tanto, necesitamos 31 columnas para calcular la probabilidad de que 0, 1, . . . 30 personas estén presentes en cada momento. En el tiempo 0, suponemos que el restaurante está vacío, de modo que la probabilidad de que 0 personas estén presentes es igual 1. Para i por lo menos 1, hay una probabilidad de 0 de que i personas estén presentes. Esta probabilidades se ingresan en el renglón 11 de las columnas D-AH.

Paso 6 La probabilidad de que nadie esté en el sistema en el tiempo 5 segundos se calcula en la celda D12 con ayuda de la fórmula

=(1-C11)*D11+sprob*E11

Esta fórmula establece la primera de las ecuaciones de aproximación.

Paso 7 Al copiar la fórmula siguiente desde la celda E12 hasta E12:AG12

=$C11*D11+(1-$C11-sprob)*E11+sprob*F11

calculamos la probabilidad de que 1, 2, . . . , 29 personas estén presentes después de 5 segundos. Esta fórmula establece la segunda ecuación de aproximación.

FIGURA 40

	A	B	C	D	E	F	G	H
1								
2							0	30
3				srate	50		0.5	40
4				sprob	0.069444		1	50
5							1.5	60
6							2	35
7							2.5	25
8								
9				0	1	2	3	4
10	Time	Hour	Arrival Prob	Prob 0	Prob 1	Prob 2	Prob 3	Prob 4
11	0	0	0.04166667	1	0	0	0	0
12	5	0.001389	0.04166667	0.958333	0.041667	0	0	0
13	10	0.002778	0.04166667	0.921296	0.076968	0.001736	0	0
14	15	0.004167	0.04166667	0.888254	0.106924	0.00475	7.23E-05	0
15	20	0.005556	0.04166667	0.858669	0.132384	0.008683	0.000262	3.01E-06
16	25	0.006944	0.04166667	0.832084	0.154055	0.013252	0.000595	1.36E-05
17	30	0.008333	0.04166667	0.808112	0.172528	0.01824	0.001082	3.69E-05
18	35	0.009722	0.04166667	0.786422	0.188297	0.023477	0.001724	7.79E-05
19	40	0.011111	0.04166667	0.766731	0.201773	0.028834	0.002516	0.000141
20	45	0.0125	0.04166667	0.748795	0.213303	0.034212	0.003448	0.000231

Paso 8 La probabilidad de que 30 personas estén presentes después de 5 segundos se calcula en la celda AH12 con la fórmula

=(1-sprob)*AH11+C11*AG11

Así se plantea la tercera ecuación de aproximación.

Paso 9 Seleccione el intervalo de celdas D12:AH12 y sitúe el cursor sobre el cruce de la esquina inferior derecha de la celda AH12. Ahora, con un doble clic con el botón izquierdo del ratón copiará las fórmulas en D12:AH12 hacia abajo para que corresponda con el número de renglones de la columna C. Por lo tanto, ya terminamos los cálculos de la distribución de probabilidades de los clientes desde las 11 AM hasta las 2 PM. Véase la figura 41.

Parte (a) El número esperado de clientes presentes a las 12:30 PM se calcula en la celda K5 (observe que el renglón 1091 tiene tiempo 1.5 h, es decir, 5400 s) con la fórmula

=SUMPRODUCT($D$9:$AH$9,D1091:AH1091)

Encontramos que estará presente un promedio de 6.25 clientes a las 12:30 PM.

FIGURA 41

	D	E	F	G	H	I	J	K	L	M	AH
2				0	30						
3	srate	50		0.5	40		minutes				
4	sprob	0.069444		1	50						
5				1.5	60		Mean #	6.253338	Mean#	1.430961	
6				2	35		12:30		11:30		
7				2.5	25						
8											
9	0	1	2	3	4	5	6	7	8	9	30
10	Prob 0	Prob 1	Prob 2	Prob 3	Prob 4	Prob 5	Prob 6	Prob 7	Prob 8	Prob 9	Prob 30
11	1	0	0	0	0	0	0	0	0	0	0
12	0.958333	0.041667	0	0	0	0	0	0	0	0	0
13	0.921296	0.076968	0.001736	0	0	0	0	0	0	0	0
14	0.888254	0.106924	0.00475	7.23E-05	0	0	0	0	0	0	0

Parte (b) Observe que el renglón 371 es el tiempo 11:30 AM. En la celda M5, la fórmula

=SUMPRODUCT(D9:AH9,D371:AH371)

muestra que se espera que sólo un promedio de 1.43 clientes estén presentes a las 11:30.

PROBLEMAS

Grupo A

1[†] Se usa una sola máquina entre las 8 AM y las 4 PM para tomar electrocardiogramas. Hay tres lugares para esperar, y cualquier paciente que llegue, y no encuentre espacio para él, es una pérdida para el sistema. La tasa de llegadas por hora en el tiempo t(t = 0 es 8 AM y t = 8 es 4 PM) está dada por

$$\lambda(t) = 9.24 - 1.584 \cos\left(\frac{\pi t}{1.51}\right) + 7.897 \operatorname{sen}\left(\frac{\pi t}{3.02}\right) - 10.434 \cos\left(\frac{\pi t}{4.53}\right) + 4.293 \cos\left(\frac{\pi t}{6.04}\right)$$

Suponga que los tiempos de servicio son exponenciales y que se ejecuta un promedio de 7 electrocardiogramas por hora. Suponga, asimismo, que las llegadas se apegan a un proceso Poisson no homogéneo. Determine cómo varía durante el día la probabilidad de que el sistema pierda a un paciente que llegue y no encuentre lugar disponible.

[†]Basado en Kao (1996).

2 Las casillas para votar están abiertas en Gotham City desde las 11 AM hasta las 6 PM. La ciudad cuenta con tres máquinas para votar. Se requiere un promedio de 1.5 minutos (distribución exponencial) para que un elector termine de votar. La tasa de llegadas de los electores en todo el día es como se indica en la tabla 12. ¿Cuál es la probabilidad de que toda la votación se complete a las 6:30 PM?

TABLA 12

Tiempo	Tasa de llegadas por hora
11 AM–mediodía	80
Mediodía–1 PM	125
1 PM–2 PM	110
2 PM–3 PM	90
3 PM–4 PM	80
4 PM–5 PM	70
5 PM–6 PM	100

RESUMEN

Distribución exponencial

Una variable aleatoria **X** sigue una distribución exponencial con parámetro λ si la densidad de **X** se obtiene con

$$f(t) = \lambda e^{-\lambda t} \qquad (t \geq 0)$$

Entonces,

$$E(\mathbf{X}) = \frac{1}{\lambda} \qquad \text{y} \qquad \operatorname{var} \mathbf{X} = \frac{1}{\lambda^2}$$

La distribución exponencial tiene la propiedad de carencia de memoria. Esto quiere decir, por ejemplo, que si los tiempos entre llegadas están distribuidos en forma exponencial con tasa o parámetro λ, entonces, sin importar cuánto tiempo pasó desde la última llegada, hay una probabilidad $\lambda\Delta t$ de que ocurrirá una llegada durante las siguientes unidades de tiempo Δt.

Los tiempos entre llegadas son exponenciales con parámetro λ si y sólo si el número de llegadas que ocurran en un intervalo de duración t sigue una distribución Poisson con parámetro λt. La función de masa para una distribución Poisson con parámetro λ se obtiene con

$$P(\mathbf{N} = n) = \frac{e^{-\lambda}\lambda^n}{n!} \qquad (n = 0, 1, 2, \ldots)$$

Distribución Erlang

Si los tiempos entre llegadas o de servicio no son exponenciales, entonces se puede usar, con frecuencia, una variable aleatoria Erlang para modelarlos. Si T es una variable aleatoria Erlang con parámetro de proporcionalidad R y parámetro de forma k, la densidad de T se estima con

$$f(t) = \frac{R(Rt)^{k-1}e^{-Rt}}{(k-1)!} \qquad (t \geq 0)$$

y

$$E(\mathrm{T}) = \frac{k}{R} \qquad \text{y} \qquad \operatorname{var} \mathrm{T} = \frac{k}{R^2}$$

Procesos de nacimiento-muerte

Por lo que se refiere a estos procesos, la probabilidad de estado estable (π_j) o fracción del tiempo que el proceso pasa en el estado j se puede determinar a partir de las ecuaciones de balance de flujo siguientes

$$\begin{array}{rr} (j = 0) & \pi_0\lambda_0 = \pi_1\mu_1 \\ (j = 1) & (\lambda_1 + \mu_1)\pi_1 = \lambda_0\pi_0 + \mu_2\pi_2 \\ (j = 2) & (\lambda_2 + \mu_2)\pi_2 = \lambda_1\pi_1 + \mu_3\pi_3 \\ \vdots & \\ (j\text{-ésima ecuación}) & (\lambda_j + \mu_j)\pi_j = \lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1} \end{array}$$

La j-ésima ecuación de balance de flujo establece que el número esperado de transiciones por unidad de tiempo fuera del estado j = (número esperado de transiciones por unidad de tiempo en el estado j). La solución de las ecuaciones de balance se determina mediante

$$\pi_j = \pi_0 \frac{\lambda_0\lambda_1 \cdots \lambda_{j-1}}{\mu_1\mu_2 \cdots \mu_j} \qquad (j = 1, 2, \ldots)$$

y el hecho de que $\pi_0 + \pi_1 + \cdots = 1$.

Notación para las características de los sistemas de líneas de espera

π_j = probabilidad de estado estable de que j clientes estén en el sistema

L = número esperado de clientes en el sistema

L_q = cantidad esperada de clientes formados en la cola

L_s = cantidad esperada de clientes en servicio

W = tiempo esperado que un cliente pasa en el sistema

W_q = tiempo esperado que un cliente pasa en la fila

W_s = tiempo esperado que un cliente pasa en el servicio

λ = cantidad promedio de clientes por unidad de tiempo

μ = cantidad promedio de servicios terminados por unidad de tiempo (tasa de servicio)

$\rho = \dfrac{\lambda}{s\mu}$ = intensidad de tránsito o factor de utilización

Modelo $M/M/1/GD/\infty/\infty$

Si $\rho \geq 1$, no existe estado estable. Para $\rho < 1$,

$$
\begin{aligned}
\pi_j &= \rho^j(1-\rho) \qquad (j = 0, 1, 2, \ldots)\\
L &= \frac{\lambda}{\mu - \lambda}\\
L_q &= \frac{\lambda^2}{\mu(\mu - \lambda)}\\
L_s &= \rho\\
W &= \frac{1}{\mu - \lambda}\\
W_q &= \frac{\lambda}{\mu(\mu - \lambda)}\\
W_s &= \frac{1}{\mu}
\end{aligned}
$$

(Las últimas tres fórmulas se obtienen de las fórmulas de L, L_q y L_s por medio de la relación $L = \lambda W$.)

Modelo $M/M/1/GD/c/\infty$

Si $\lambda \neq \mu$,

$$
\begin{aligned}
\pi_0 &= \frac{1-\rho}{1-\rho^{c+1}}\\
\pi_j &= \rho^j \pi_0 \qquad (j = 1, 2, \ldots, c)\\
\pi_j &= 0 \qquad (j = c+1, c+2, \ldots)\\
L &= \frac{\rho[1-(c+1)\rho^c + c\rho^{c+1}]}{(1-\rho^{c+1})(1-\rho)}
\end{aligned}
$$

Si $\lambda = \mu$,

$$
\begin{aligned}
\pi_j &= \frac{1}{c+1} \qquad (j = 0, 1, \ldots, c)\\
L &= \frac{c}{2}
\end{aligned}
$$

Para todos los valores de λ y μ,

$$
\begin{aligned}
L_s &= 1 - \pi_0\\
L_q &= L - L_s\\
W &= \frac{L}{\lambda(1-\pi_c)}\\
W_q &= \frac{L_q}{\lambda(1-\pi_c)}\\
W_s &= \frac{1}{\mu}
\end{aligned}
$$

Modelo $M/M/s/GD/\infty/\infty$

Si $\rho \geq 1$, no existe estado estable. Para $\rho < 1$,

$$\pi_0 = \frac{1}{\displaystyle\sum_{i=0}^{i=s-1} \frac{(s\rho)^i}{i!} + \frac{(s\rho)^s}{s!(1-\rho)}}$$

$$\pi_j = \frac{(s\rho)^j\,\pi_0}{j!} \qquad (j = 1, 2, \ldots, s)$$

$$\pi_j = \frac{(s\rho)^j\,\pi_0}{s!s^{j-s}} \qquad (j = s, s+1, s+2, \ldots)$$

$$P(j \geq s) = \frac{(s\rho)^s\,\pi_0}{s!(1-\rho)} \qquad \text{(calculado en tabla 6)}$$

$$L_q = \frac{P(j \geq s)\rho}{1-\rho}$$

$$W_q = \frac{P(j \geq s)}{s\mu - \lambda}$$

$$L_s = \frac{\lambda}{\mu}$$

$$W_s = \frac{1}{\mu}$$

$$L = L_q + \frac{\lambda}{\mu}$$

$$W = \frac{L}{\lambda}$$

Modelo $M/G/\infty/GD/\infty/\infty$

$$L = L_s = \frac{\lambda}{\mu}$$

$$W = W_s = \frac{1}{\mu}$$

$$W_q = L_q = 0$$

Modelo $M/G/1/GD/\infty/\infty$

σ^2 = varianza de la distribución del tiempo de servicio

$$L_q = \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)}$$

$$L = L_q + \rho$$

$$L_s = \lambda\left(\frac{1}{\mu}\right)$$

$$W_q = \frac{L_q}{\lambda}$$

$$W = W_q + \frac{1}{\mu}$$

$$W_s = \frac{1}{\mu}$$

$$\pi_0 = 1 - \rho$$

Modelo de reparación de máquinas (*M/M/R/GD/K/K*)

$\rho = \dfrac{\lambda}{\mu}$

L = número esperado de máquinas descompuestas

L_q = cantidad esperada de máquinas que esperan servicio

W = tiempo promedio que una máquina pasa descompuesta

W_q = tiempo promedio que una máquina pasa en espera de servicio

π_j = probabilidad de estado estable de que j máquinas estén descompuestas

λ = tasa a la cual se descomponen las máquinas

μ = tasa a la cual se reparan las máquinas

Asimismo,

$$\pi_j = \binom{K}{j}\rho^j\pi_0 \qquad (j = 0, 1, \ldots, R)$$

$$= \frac{\binom{K}{j}\rho^j j!\,\pi_0}{R!R^{j-R}} \qquad (j = R+1, R+2, \ldots, K)$$

$$L = \sum_{j=0}^{j=K} j\pi_j$$

$$L_q = \sum_{j=R}^{j=K} (j - R)\pi_j$$

$$\bar{\lambda} = \sum_{j=0}^{j=K} \pi_j\lambda_j = \sum_{j=0}^{j=K} \lambda(K - j)\pi_j = \lambda(K - L)$$

$$W = \frac{L}{\bar{\lambda}}$$

$$W_q = \frac{L_q}{\bar{\lambda}}$$

Líneas de espera exponenciales en serie

Si existe un estado estable y si (1) los tiempos entre llegadas para un sistema de líneas de espera en serie son exponenciales con tasa λ; (2) los tiempos de servicio para cada servidor en la etapa i son exponenciales, y (3) cada etapa tiene una sala de espera de capacidad infinita, entonces los tiempos entre llegada de los clientes a cada etapa del sistema de la línea de espera son exponenciales con tasa λ.

Modelo $M/G/s/GD/s/\infty$

El sistema pierde una fracción π_s de todos los clientes, y π_s depende sólo de la tasa de llegadas λ y de la media $\frac{1}{\mu}$ del tiempo de servicio. Con la figura 21 se puede determinar π_s.

Qué hacer si los tiempos entre llegadas o los de servicio no son exponenciales

Se puede aplicar una prueba chi cuadrada para determinar si los datos reales indican que los tiempos entre llegadas o de servicio son exponenciales. Si el tiempo entre llegadas, o el tiempo de servicio, o ambos no son exponenciales, entonces se podría llegar a una aproximación de L, L_q, W y W_q mediante la fórmula de Allen-Cunneen.

No hay fórmula o tabla que se pueda usar para calcular las características de operación del sistema en muchos sistemas de líneas de espera. En estos casos hay que recurrir a la simulación (véanse capítulos 21 y 22).

Redes cerradas de líneas de espera

Los sistemas de manufactura de computadoras en los cuales hay un número constante de trabajos presentes se modelan, algunas veces, como **redes cerradas de líneas de espera**.

Sea P_{ij} la probabilidad de que un trabajo irá al servidor j después de completar su servicio en la estación i. Sea P la matriz cuya entrada $(i-j)$ es P_{ij}. Suponemos que los tiempos de servicio en el servidor j están distribuidos en forma exponencial con parámetro μ_j. El sistema tiene s servidores, y están presentes, en todo momento, exactamente N trabajos. Sea n_i el número de trabajos presentes en el servidor i. Entonces, el estado del sistema, en cualquier momento dado, se puede definir mediante un vector n-dimensional $\mathbf{n} = (n_1, n_2, \ldots, n_s)$. El conjunto de estados posibles se obtiene mediante $S_N = \{\mathbf{n}$ tal que todas $n_i \geq 0$ y $n_1 + n_2 + \cdots + n_s = N\}$.

Sea λ_j igual a la tasa de llegadas al servidor j. Como no hay llegadas externas, podríamos hacer que todas las $r_j = 0$ y obtener los valores de las λ_j mediante la ecuación utilizada en la situación de redes abiertas. Es decir,

$$\lambda_j = \sum_{i=1}^{i=s} \lambda_i P_{ij} \qquad j = 1, 2, \ldots, s$$

Como los trabajos nunca dejan el sistema, para cada i, $\sum_{j=1}^{j=s} P_{ij} = 1$. Este hecho da lugar a que la ecuación anterior no tenga solución única. Resulta que, por fortuna, podemos usar cualquier solución para ayudarnos a obtener las probabilidades de estado estable. Si definimos

$$\rho_i = \frac{\lambda_i}{\mu_i}$$

entonces, determinamos, para cualquier estado n, sus probabilidad de estado estable $\Pi_N(\mathbf{n})$ a partir de la ecuación siguiente:

$$\Pi_N(\mathbf{n}) = \frac{\rho_1^{n_1} \rho_2^{n_2} \cdots \rho_n^{n_s}}{G(N)}$$

Aquí, $G(N) = \sum_{\mathbf{n} \in S_N} \rho_1^{n_1} \rho_2^{n_2} \cdots \rho_s^{n_s}$.

El algoritmo de Buzen proporciona una manera eficaz de estimar $G(N)$ (en una hoja de cálculo). Una vez que tenemos la distribución de probabilidades de estado estable, podemos determinar con facilidad otras medidas de efectividad, tal como la longitud de la cola esperada en cada servidor y el tiempo previsto que un trabajo tarda en cada visita a un servidor, fracción de tiempo que un servidor está ocupado y la producción de cada servidor (trabajos por segundo procesados por cada servidor).

Con el objeto de obtener $G(N)$, calculamos en forma recursiva las cantidades $C_i(k)$ para $i = 1, 2, \ldots, s$ y $k = 0, 1, \ldots, N$. La recursión inicia con $C_1(k) = \rho_1^k$, $k = 0, 1, \ldots, N$ y $C_i(0) = 1$, $i = 1, 2, \ldots, s$. Para otros valores de k e i, generamos los valores de $C_i(k)$ en forma recursiva por medio de la relación siguiente:

$$C_i(k) = C_{i-1}(k) + \rho_i C_i(k-1)$$

Luego se puede demostrar que $G(N) = C_s(N)$.

Una aproximación para el sistema de líneas de espera *G/G/m*

Los tiempos entre llegadas se ajustan, la mayoría de las veces, a una variable aleatoria exponencial. Los tiempos de servicio, en cambio, no se apegan, con frecuencia, a una distribución exponencial. Cuando los tiempos entre llegadas y los tiempos de servicio se ajustan cada uno a una variable aleatoria no exponencial, el sistema de líneas de espera se denomina sistema *G/G/m*. En el caso de estas situaciones, no funcionan las plantillas analizadas en las secciones anteriores de este capítulo. La aproximación de Allen-Cunneen proporciona, por fortuna en muchas ocasiones, una buena aproximación de L, W, L_q y W_q para sistemas *G/G/m*. El archivo ggm.xls pone en práctica una hoja de cálculo con la aproximación de Allen-Cunneen. El usuario sólo tiene que escribir la información siguiente:

ggm.xls

- El número promedio de llegadas por unidad de tiempo (λ) en la celda B3.
- La tasa promedio a la cual se atiende a los clientes (μ) en la celda B4.
- El número de servidores (s) en la celda B5.
- El coeficiente cuadrático de la variación –(varianza de tiempos entre llegadas)/tiempo promedio entre llegadas)^2– de los tiempos entre llegadas en la celda B6.
- El coeficiente cuadrático de la variación –(varianza de tiempos de servicio)/(tiempo promedio de servicio)^2– de los tiempos de servicio en la celda B7.

La aproximación de Allen-Cunneen es exacta si los tiempos entre llegadas y los tiempos de servicio son exponenciales. Las extensas pruebas de Tanner indican que, en una gran diversidad de situaciones, los valores de L, W, L_q y W_q obtenidos mediante la aproximación están a 10% de sus valores verdaderos.

Comportamiento transitorio de los sistemas de líneas de espera

Definimos $P_i(t)$ como la probabilidad de que i clientes estén presentes en el tiempo t. Luego suponemos (aunque no es necesario), que el sistema está al principio vacío, de modo que $P_0(0) = 1$ y para $i > 0$, $P_0(i) = 0$. Entonces, dado que conocemos $P_i(t)$, podemos estimar $P_i(t + \Delta t)$ como sigue

$$\begin{aligned}
P_0(t+\Delta t) &= (1-\lambda(t)\Delta t)P_0(t) + \mu(t)\Delta t P_1(t)\\
P_i(t+\Delta t) &= \lambda(t)\Delta t P_{i-1}(t) + (1-\lambda(t)\Delta t - \min(s(t), i)\mu(t)\Delta t)P_i(t) + \min(s(t),\\
&\quad i+1)\mu(t)\Delta t P_{i+1}(t),\ N-1 \geq i \geq 1\\
P_N(t+\Delta t) &= \lambda(t)\Delta t P_{N-1}(t) + (1-\min(s(t), N)\mu(t))\Delta t P_N(t)
\end{aligned}$$

Estas ecuaciones se basan en el supuesto de que, si el estado es i en el tiempo t, entonces durante el Δt siguiente, la probabilidad de una llegada es $\lambda(t)\ \Delta t$, y la probabilidad de que termine un servicio es $(s(t), i)\mu(t)\Delta t$.

PROBLEMAS

Grupo A

1 Los autobuses llegan a la parada del centro del pueblo y salen de la parada del bulevar. Las experiencias anteriores indican que 20% de las veces, el intervalo entre autobuses es de 20 minutos; 40% de las veces, el intervalo es 40 minutos, y 40% de las veces, el intervalo es de 2 horas. Si apenas llegué a la parada del centro de la ciudad, ¿cuánto tiempo en promedio debo prever que voy a esperar el autobús?

2 Las inscripciones en la universidad del estado proceden como sigue: los estudiantes, al entrar al vestíbulo de inscripciones, los estudiantes primero se forman para inscribirse a las clases. Un solo empleado maneja la inscripción a las clases, y se tarda un promedio de dos minutos atender la inscripción de un estudiante. Luego, el estudiante debe esperar en otra fila para pagar la inscripción. Un solo empleado atiende los pagos de la inscripción. Este empleado se tarda un promedio de dos minutos en procesar el pago de un estudiante. Entonces, el estudiante abandona el edificio de inscripciones. Un promedio de 15 estudiantes por hora llega al lugar de las inscripciones.

a Si los tiempos entre llegadas y de servicio son exponenciales, ¿cuál es el tiempo previsto que un estudiante pasa en el lugar de las inscripciones?

b ¿Cuál es la probabilidad de que en los cinco minutos próximos entren exactamente dos estudiantes al vestíbulo de inscripciones?

c Sin mayor información, ¿cuál es la probabilidad de que durante los tres minutos próximos ningún estudiante llegue al escritorio del empleado que recibe los pagos de inscripción?

d Suponga que el sistema de inscripciones cambió, de modo que ahora el estudiante se inscribe en sus clases y paga en el mismo lugar. Si el tiempo de servicio en este único lugar se apega una distribución Erlang con parámetro de proporcionalidad de 1.5 por minuto y parámetro de forma igual a 2, ¿cuál es el tiempo previsto que un estudiante pasa en la cola?

3 En la oficina de correos de una pequeña ciudad, los usuarios esperan en una sola cola a la primera ventanilla que abra. Un promedio de 100 usuarios por hora llega a la oficina de correos, y cada ventanilla puede atender a un promedio de 45 personas por hora. La oficina estima que hay un costo de 10 centavos por cada minuto que un usuario pasa en la cola, y opina que cuesta 20 dólares por hora mantener una ventanilla abierta. Los tiempos entre llegadas y los tiempos de servicio son exponenciales.

a Para minimizar el costo total esperado por hora, ¿cuántas ventanillas deben estar abiertas?

b Si el objetivo de la oficina de correos es asegurar que cuando mucho 5% de todos los usuarios pasarán más de cinco minutos en la cola, ¿cuántas ventanillas deben estar abiertas?

4 Un promedio de 500 personas aprueba, cada año, el examen de la barra de abogados del estado de Nueva York, e ingresa a la profesión de las leyes. Un abogado se dedica a la práctica profesional de las leyes en el estado de Nueva York un promedio de 35 años. Veinte años después, a partir de hoy, ¿cuántos abogados espera que haya en el estado de Nueva York?

5 Hay cinco estudiantes y un barril de cerveza en una fiesta desenfrenada y extravagante. El tiempo para llenar un vaso de cerveza del barril está distribuido en forma exponencial, con un tiempo promedio de dos minutos. El tiempo para beber una cerveza también está distribuido en forma exponencial, con una media de 18 minutos. Después de terminar una cerveza, cada estudiante regresa inmediatamente por otra.

a ¿Cuánto tiempo espera en promedio un estudiante en la cola para conseguir una cerveza?

b ¿Qué fracción del tiempo no está en uso el barril?

c Si el barril contiene 500 vasos de cerveza, ¿cuánto tiempo, en promedio, tardará en vaciarse el barril?

6 La gerente de un grupo grande de empleados debe decidir si requiere otra fotocopiadora. El costo de una fotocopiadora es de 40 dólares por ocho horas al día aunque la máquina no esté en uso. Un promedio de cuatro personas por hora necesita usar la fotocopiadora. Cada persona usa la copiadora por un promedio de 10 minutos. Los tiempos entre llegadas y los tiempos de copiado están distribuidos en forma exponencial. Los empleados reciben ocho dólares como pago por hora, y suponemos que se genera un costo por la espera cuando un trabajador hace cola o está usando la máquina. ¿Cuántas fotocopiadoras se deben rentar?

7 Un taller de lavado automático de automóviles asea un vehículo en 10 minutos. Los clientes llegan cada 15 minutos, en promedio, (distribución exponencial).

a En promedio, ¿cuántos vehículos están formados en espera de que los laven?

b Si el lavado de automóviles fuera más rápido, ¿qué tiempo de lavado reduciría el tiempo de espera promedio a cinco minutos?

8 La compañía Newcoat Painting ha experimentado durante algún tiempo una alta demanda en su servicio de repintado para automóviles. Como ha tenido que rechazar negocios, a la administración le preocupa que el espacio limitado disponible para guardar los automóviles que esperan que los pinten incurra en un costo por el ingreso perdido. Un pequeño terreno vacío, que está adyacente a la compañía de pintura, se puede rentar, desde hace poco, durante largo tiempo, a un costo de 10 dólares por día. La administración opina que cada cliente perdido cuesta 20 dólares en ganancias. Se estima que la demanda actual es de 21 automóviles por día con tiempos entre llegadas exponenciales (incluidos los rechazados), y la compañía puede dar servicio a una tasa exponencial de 24 automóviles por día. Los vehículos se procesan de acuerdo con la disciplina FCFS. El espacio de espera está limitado ahora a nueve automóviles, pero puede aumentar a 20 vehículos si se renta el terreno de al lado. Newcoats quiere determinar si debe rentar el terreno. La administración también desea conocer la ganancia esperada perdida por día debido a los clientes que rechace si llega a rentar el terreno. Sólo se puede pintar un vehículo a la vez.

9 En un restaurante exclusivo hay sólo una mesa y espacio de espera nada más para otro grupo. Las personas que lleguen cuando la zona de espera está llena, se les negará la entrada. La tasa de llegadas sigue una distribución exponencial a razón de un grupo por hora. Un grupo promedio requiere una hora (distribución exponencial) desde el momento en que se le sirve hasta que consumen sus alimentos. ¿Cuál es el tiempo promedio que tarda un grupo esperando una mesa?

10 El dueño de un restaurante exclusivo tiene dos mesas y sólo un mesero. Si la segunda mesa está ocupada, el dueño atiende esa mesa. Los tiempos de servicio están distribuidos en forma exponencial, con media de una hora, y el tiempo entre llegadas está distribuido en forma exponencial con media de 1.5 horas. Cuando el restaurante está lleno, las personas deben esperar afuera haciendo cola.

a ¿Qué porcentaje del tiempo el dueño atiende una mesa?

b Si el dueño quiere pasar cuando mucho 10% de su tiempo atendiendo mesas, ¿cuál es la tasa de llegadas máxima que se puede tolerar?

11 Los barcos llegan a un puerto a una tasa promedio de 2 barcos cada tres días. En promedio, un solo grupo de trabajadores requiere un día para descargar el barco. Suponga que los tiempos entre llegadas y de servicio son exponenciales. La compañía de embarque es dueña de las instalaciones del puerto, así como de los barcos que las usan. Se estima que a la compañía le cuesta 1 000 dólares cada día que un barco pasa en el puerto. El personal que atiende a los barcos consiste en 100 trabajadores, a quienes se les paga un promedio de 30 dólares por día. Un asesor recomienda que la compañía de embarque contrate 40 trabajadores adicionales y divida al personal en cuadrillas de 70 personas cada una. Esto daría a cada cuadrilla un tiempo promedio de carga y descarga de 3/2 días. ¿Qué arreglo de personal recomendaría usted a la compañía?

12 Un promedio de 40 trabajos por día llega a una fábrica. El tiempo entre las llegadas de los trabajos está distribuido en forma exponencial. La fábrica es capaz de procesar un promedio de 42 trabajos por día, y el tiempo para procesar un trabajo sigue una distribución exponencial.

a ¿Cuál es la probabilidad de que lleguen exactamente 180 trabajos a la fábrica durante un periodo de cinco días?

b En promedio, ¿cuánto tiempo se requiere antes de que un trabajo esté terminado (medido desde el momento en que el trabajo llega a la fábrica)?

c ¿Qué fracción del tiempo está ociosa la fábrica?

d ¿Cuál es la probabilidad de que inicien las actividades sobre un trabajo luego de dos días de que éste llegó a la fábrica?

13 Un taller de impresión recibe en promedio un pedido por día. El tiempo en promedio que se requiere para terminar el pedido es .5 día. En cualquier momento, el taller puede trabajar en más de un pedido.

a ¿Cuántos trabajos están presentes en promedio en el taller?

b ¿Cuánto tiempo tiene que esperar, en promedio, una persona que hace un pedido hasta que está terminado?

c ¿Cuál es la probabilidad de que un pedido se terminará dentro de dos días después de su llegada?

Grupo B

14 Un compañía de ventas por correo L.L. Pea recibe un promedio de 200 llamadas por hora (los tiempos entre las llamadas siguen una distribución exponencial). Los operadores de L.L. Pea necesitan un promedio de tres minutos para atender una llamada. Si el que llama escucha una señal de ocupado, L.L. Pea supone que el posible cliente llamará a Seas Beginning (una compañía competidora), y L.L. Pea perderá un promedio de 30 dólares en ganancias. El costo de conservar una línea telefónica abierta es de nueve dólares por hora. ¿Cuántos operadores debe tener L.L. Pea en funciones?

15 Cada hora llega un promedio de tres clientes tipo 1 y tres clientes tipo 2 a una estación que cuenta con un solo servidor. Los tiempos entre llegadas para cada tipo de cliente son exponenciales e independientes. El tiempo de servicio promedio para un cliente tipo 1 es de seis minutos, y el tiempo de servicio promedio para un cliente tipo 2 es tres minutos (todos los tiempos de servicio están distribuidos en forma exponencial). Considere los tres arreglos de servicio siguientes:

Arreglo 1 Todos los clientes esperan en una sola cola y se atienden según la disciplina FCFS.

Arreglo 2 Los clientes tipo 1 tienen preferencia sin prioridad sobre los clientes tipo 2.

Arreglo 3 Los clientes tipo 2 tienen preferencia sin prioridad sobre los clientes tipo 1.

¿Qué arreglo dará como resultado el tiempo de espera esperado mínimo para el cliente? ¿Qué arreglo dará el tiempo de espera esperado mayor por cliente?

16 El departamento de investigación de operaciones de la universidad de Podunk tiene dos líneas telefónicas. Un promedio de 30 personas por hora intenta llamar al departamento, y la duración promedio de una llamada telefónica es de un minuto. Si una persona llama cuando las dos líneas están ocupadas, cuelga, y el sistema la pierde. Suponga que el tiempo entre los intentos de las personas para llamar y los tiempos de servicio es exponencial.

a ¿Qué fracción del tiempo ambas líneas están libres? ¿Qué fracción del tiempo exactamente una línea estará libre?

b En promedio, ¿cuántas líneas estarán ocupadas?

c En promedio, ¿cuántas personas que llaman colgarán cada hora?

17† Una ciudad pequeña tiene dos ambulancias. La ambulancia 1 está en la universidad local y la ambulancia 2 está en el centro de la ciudad. Si se solicita la ambulancia desde la universidad, la ambulancia que está en la institución se envía, si está disponible. Si no es así, se envía la ambulancia que está en el centro de la ciudad, si está disponible. Si ninguna ambulancia está disponible, se supone que el sistema perdió la llamada. Si la solicitud de ambulancia proviene de cualquier otro lado de la ciudad, se envía, si está disponible, la ambulancia que está en el centro de la ciudad. Si no está disponible, se envía la ambulancia que está en la universidad. Si ninguna ambulancia está disponible, se considera que el sistema perdió la llamada. El tiempo entre las llamadas sigue una distribución exponencial. Se recibe un promedio de tres llamadas por hora desde la universidad, y un promedio de cuatro llamadas por hora desde el resto de la ciudad. El tiempo promedio (distribución exponencial) que se tarda una ambulancia en responder a una llamada y en estar lista para contestar otra se proporciona en la tabla 13.

a ¿Qué fracción del tiempo está ocupada la ambulancia del centro de la ciudad?

b ¿Qué fracción del tiempo está ocupada la ambulancia de la universidad?

c ¿Qué fracción de llamadas pierde el sistema?

TABLA 13

La ambulancia viene de	Ambulancia va a Universidad	Ambulancia va a No a la universidad
Universidad	4 minutos	7 minutos
Centro de la ciudad	5 minutos	4 minutos

†Basado en Carter (1972).

d ¿Quién espera más, en promedio, una ambulancia? ¿Un estudiante de la universidad o una persona de la ciudad?

18 Un promedio de 10 personas por hora llega (tiempo entre llegadas exponenciales) con la pretensión de nadar en la YMCA. Cada una pretende nadar 30 minutos. La YMCA tiene tres carriles abiertos para nadar en ellos. Si un nadador está en un carril, nada hacia arriba y hacia abajo por el lado derecho del carril. Si dos nadadores están en un carril cada uno nada hacia arriba y hacia abajo por un lado del carril. Los nadadores siempre se forman en el carril donde hay en menor número de nadadores. Si los tres carriles están ocupados por dos nadadores, un posible nadador se disgusta y se va.

a ¿Qué fracción del tiempo tres personas estarán nadando en los carriles?

b En promedio, ¿cuántas personas están nadando en los carriles de la piscina?

c ¿Cuántos carriles necesita la YMCA asignar a las personas que nadan en carriles para tener la seguridad de que cuando mucho 5% de todos los que desean nadar en carriles se disgusten y se vayan?

19† (Requiere una hoja de cálculo). Un promedio de 140 personas por año solicita una vivienda pública en Boston. Está disponible un promedio de 20 unidades habitacionales por año. Durante un año dado, hay 10% de probabilidades de que una familia en la lista de espera encuentre un alojamiento privado y se borrará de la lista. Suponga que todas las variables aleatorias pertinentes siguen una distribución exponencial.

a ¿Cuántas familias, en promedio, estarán en la lista de espera?

b En promedio, ¿cuánto tiempo pasará una familia en la lista antes de obtener alojamiento (ya sea público o privado)? Para la última pregunta, ¡recuerde que $L = \lambda W$!

†Basado en Kaplan (1986).

BIBLIOGRAFÍA

Las obras siguientes contienen análisis excelentes acerca de la teoría de colas para un nivel intermedio:

Cooper, R. *Introduction to Queuing Theory*, 2a ed. Nueva York: North-Holland, 1981.

Gross, D. y C. Harris. *Fundamentals of Queuing Theory*, 3a ed. Nueva York: Wiley, 1997.

Karlin, S. y H. Taylor. *A First Course in Stochastic Processes*, 2a ed. Orlando, Fla.: Academic Press, 1975.

Lee, A. *Applied Queuing Theory*. Nueva York: St. Martin's Press, 1966.

Ross, S. *Applied Probability Models with Optimization Applications*. San Francisco, Calif.: Holden-Day, 1970.

Saaty, T. *Elements of Queuing Theory with Applications*. Nueva York: Dover, 1983.

En los tres libros siguientes encontrará estudios excelentes de la teoría de colas para un nivel avanzado:

Heyman, D. y M. Sobel. *Stochastic Models in Operations Research*, vol. I. Nueva York: McGraw-Hill, 1984.

Kao, E. *An Introduction to Stochastic Processes*. Belmont, Cal.: Duxbury, 1997.

Kleinrock, L. *Queuing Systems*, vol. 1 y 2. Nueva York: Wiley, 1975.

Se recomienda por ser estudios orientados a las aplicaciones de la teoría de colas:

Hall, R. *Queuing Methods for Service and Manufacturing*. Englewood Cliffs, N.J.: Prentice Hall, 1991.

Tanner, M. *Practical Queuing Analysis*. Nueva York: McGraw-Hill. 1995.

Brigham, G. "On a Congestion Problem in an Aircraft Factory", *Operations Research* 3(1955):412-428.

Buzen, J. P. "Computational Algorithms for Closed Queuing Networks with Exponential Servers", *Communications of the ACM* 16:9(1973):527-531.

Carter, G., J. Chaiken y E. Ignall. "Response Areas for Two Emergency Units", *Operations Research* 20(1972): 571-594.

Denardo, E. *Dynamic Programming: Theory and Applications*. Englewood Cliffs, N.J.: Prentice Hall, 1982. Explica por qué los tiempos entre llegadas son con frecuencia exponenciales.

Erickson, W "Management Science and the Gas Shortage", *Interfaces* 4(1973):47-51.

Feller, W. *An Introduction to Probability Theory and Its Applications*, 2a ed., vol. I. Nueva York: Wiley, 1957. Demuestra que la distribución exponencial es la única variable aleatoria continua con la propiedad de carencia de memoria.

Gilliam, R. "An Application of Queuing Theory to Airport Passenger Security Screening", *Interfaces* 9(1979): 117-123.

Green, J. "Managing a Telephone System Demands Skill", *The Office* (Noviembre 1987):144-145.

Hillier, F. y O. Yu. *Queuing Tables and Graphs*. Nueva York: North-Holland, 1981.

Jackson, J. "Networks of Waiting Lines", *Operations Research* 5(1957):518-521.

Jain, R. *The Art of Computer System Performance Analysis*. Nueva York: Wiley, 1991.

Kaplan, E. "Tenant Assignment Models", *Operations Research* 34(no. 6, 1986):833-843.

Karmarker, U. "Lot-Sizing and Lead-Time Performance in a Manufacturing Cell", *Interfaces* 15(no. 2, 1985):1-9.

Karush, W "A Queuing Model for an Inventory Problem", *Operations Research* 5(1957):69:3-703.

Kendall, *D.* "Some Problems in the Theory of Queues", *Journal of the Royal Statistical Society*, Serie B, 13(1951): 151-185.

Kolesar, P. "A Quick and Dirty Response to the Quick and Dirty Crowd: Particularly to Jack Byrd's 'The Value of Queuing Theory'", *Interfaces* 9(1979):77-82.

Law, A. y W. Kelton. *Simulation Modeling and Analysis*. Nueva York: McGraw-Hill, 1990. Estudia la simulación de los sistemas de colas y el ajuste de variables aleatorias a datos de tiempos entre llegadas y tiempos de servicio reales.

Tannenbaum, A. *Computer Networks*. Englewood Cliffs, N.J.: Prentice Hall, 1981.

Vogel, M. "Queuing Theory Applied to Machine Manning", *Interfaces* 9(1979):1-8.