Imagínese las siguientes situaciones:

1. Clientes que esperan a ser atendidos en las cajas registradoras de un supermercado.
2. Automóviles que esperan avanzar en una luz de alto.
3. Pacientes que esperan ser atendidos en una clínica.
4. Aviones que esperan para despegar en un aeropuerto.
5. Máquinas descompuestas que esperan ser reparadas por un técnico.
6. Cartas que esperan ser elaboradas por una secretaria.
7. Programas que esperan ser procesados por una computadora digital.

Lo que tienen en común estas situaciones es la espera. Sería más adecuado si se nos pudieran ofrecer estos servicios, y otros similares, sin la "molestia" de tener que esperar. Pero nos guste o no, la espera es parte de nuestra vida diaria y todo lo que esperamos conseguir es reducir su incomodidad a niveles soportables.

El fenómeno de la espera es el resultado directo de la *aleatoriedad* en la operación de instalaciones de servicio. En general, la llegada del cliente y su tiempo de servicio no se conocen con anticipación; pero por otra parte, la operación de la instalación se podría programar en forma tal que eliminaría la espera por completo.

Nuestro objetivo al estudiar la operación de una instalación de servicio en condiciones aleatorias es el de asegurar algunas características que midan el desempeño del sistema sometido a estudio. Por ejemplo, una medida lógica de desempeño es el tiempo que se calcula esperará un cliente antes de ser atendido. Otra medida es el porcentaje de tiempo que no se utiliza en la instalación de servicio. La primera medida vislumbra el sistema desde el punto de vista del cliente, mientras que la segunda evalúa el grado de uso de la instalación. Podemos advertir intuitivamente que cuanto mayor sea el tiempo de espera del cliente, tanto menor es el porcentaje de tiempo que se mantendría ociosa la instalación, y viceversa. Estas medidas de desempeño pueden utilizarse para seleccionar el nivel de servicio (o tasa de servicio) que producirá un equilibrio razonable entre las dos situaciones en conflicto.

En este capítulo se analizan varios modelos de espera o de líneas de espera, l. de e., (colas o filas) que explican una diversidad de operaciones de servicio. El objetivo final de resolver estos modelos consiste en determinar las características que miden el desempeño del sistema. En el capítulo 16 demostramos cómo se puede utilizar esta información en la búsqueda de un diseño "óptimo" para la instalación de servicio.

15.1 ELEMENTOS BASICOS DEL MODELO DE LINEAS DE ESPERA

Desde el punto de vista de un modelo de espera, una situación de línea de espera se genera de la manera siguiente. Cuando el cliente llega a la instalación se forma en una línea de espera (cola o fila). El servidor elige a un cliente de la línea de espera para comenzar a prestar el servicio. Al culminarse un servicio, se repite el proceso de elegir a un nuevo cliente (en espera). Se supone que no se pierde tiempo entre el momento en que un cliente ya atendido sale de la instalación y la admisión de un nuevo cliente de la línea de espera.

Los protagonistas principales en una situación de espera son el **cliente** y el **servidor**. En los modelos de espera, la interacción entre el cliente y el servidor sólo es de interés en tanto que se relacione con el *periodo* que necesita el cliente para completar su servicio. Por lo tanto, desde el punto de vista de las llegadas de clientes, nos interesan los intervalos de tiempo que separan llegadas *sucesivas*. Asimismo, en el caso del servicio, es el tiempo de servicio por cliente el que cuenta en el análisis.

En los modelos de espera, las llegadas y los tiempos de servicio de clientes se resumen en términos de distribuciones de probabilidad que normalmente se conocen como **distribuciones de llegadas** y **de tiempo de servicio.** Estas distribuciones pueden representar situaciones donde llegan clientes y son atendidos *individualmente* (por ejemplo, en bancos o supermercados). En otros casos, los clientes pueden llegar y/o ser atendidos en grupos (por ejemplo, en restaurantes). Este último caso se conoce normalmente como **líneas de espera masivas.**

Aunque los patrones de llegadas y salidas son los factores principales en el análisis de las líneas de espera, también pueden figurar otros factores en forma importante en la elaboración de los modelos. El primer factor es la forma como se elige a los clientes de la línea de espera para dar inicio al servicio. Esta se conoce como la **disciplina de servicio.** La disciplina más común, y en apariencia justa, es la regla FCFS (el primero en llegar es el primero en ser atendido). Las reglas LCFS (el último en llegar es el primero en ser atendido) y SIRO (servicio en orden aleatorio) pueden surgir también en situaciones prácticas. También debemos agregar que en tanto que la disciplina de servicio regula la selección de clientes de una línea de espera, también es posible que los clientes que lleguen a una instalación sean colocados en **líneas de espera con prioridad**, para que aquellas personas con mayor prioridad reciban preferencia de ser atendidos en primer término. No obstante, la selección específica de clientes de cada línea de espera con prioridad puede apegarse a cualquier disciplina de servicio.

El segundo factor tiene que ver con el diseño de la instalación y la ejecución del servicio. La instalación puede incluir más de un servidor, con lo cual es posible atender a tantos clientes en forma simultánea como número de servidores haya (por ejemplo, los cajeros bancarios). En este caso, todos los servidores ofrecen el mismo servicio y se dice que la instalación tiene **servidores paralelos.** Por otra parte, la instalación puede comprender un número de estaciones en serie por las que puede pasar el cliente antes de que se complete el servicio (por ejemplo, el procesamiento de un producto en una serie de máquinas). Las situaciones resultantes se conocen normalmente como **líneas de espera en serie** o **líneas de espera sucesivas.** El diseño más general de una instalación de servicio incluye estaciones de procesamiento en serie y en paralelo. Esto da origen a lo que llamamos **líneas de espera en red.**

El tercer factor tiene que ver con el **tamaño de la línea de espera** admisible. En ciertos casos, sólo se puede admitir a un número limitado de clientes, posiblemente en virtud de la limitación del espacio (por ejemplo, los espacios de estacionamiento que se permiten en un autobanco). Cuando la línea de espera se llena a toda su capacidad, los clientes que llegan no se pueden formar en la línea de espera.

El cuarto factor se relaciona con la naturaleza de la fuente que genera llamadas solicitando servir (llegadas de clientes). **La fuente de llamadas** puede ser capaz de generar un número finito de clientes o (en teoría) infinitamente muchos clientes. Existe una fuente finita cuando una llegada afecta la tasa de llegada de nuevos clientes. En una oficina con *M* máquinas, la fuente de llamadas antes de que se descomponga cualquier máquina consta de *M* clientes en potencia. Cuando se descompone una máquina, se convierte en un cliente y, por lo tanto, no puede generar nuevas llamadas hasta que sea reparada. Se puede hacer una distinción entre la situación de la oficina y otros donde la "causa" para generar llamadas está limitada, no obstante que puede generar un número infinito de llegadas. Por ejemplo, en una oficina de mecanografía, el número de usuarios es finito, no obstante que cada usuario pudiera generar un número ilimitado de llegadas, ya que en términos generales un usuario no necesita esperar a que se termine el material entregado con anterioridad antes de generar otro nuevo.

Los modelos de espera que representan situaciones en las que los seres humanos son clientes y/o servidores, deben estar diseñados para tomar en cuenta el efecto de la **conducta del ser humano.** Un servidor "humano" puede **cambiarse** de una línea de espera a otra, con la esperanza de reducir su tiempo de espera (la próxima vez que el lector esté en un banco o en un supermercado puede "matar" el tiempo de espera observando estos cambios). Algunos clientes "humanos" pueden también **eludir** formarse en una línea de espera en virtud de que *anticipen* una demora apreciable, o bien, pueden **renunciar** *después* de estar un momento en la fila debido a que su espera haya sido demasiado larga. (Nótese que en término de la conducta del ser humano, quizá una espera que es muy larga para una persona no sea tan larga para otra.)

Sin duda alguna, existen otras cualidades y/o características presentes en las situaciones de espera de todos los días. No obstante, desde el punto de vista del *modelo* de espera, estas cualidades y/o características pueden tomarse en cuenta sólo si se pueden cuantificar de manera que haga posible su inclusión matemática en el modelo. Asimismo, los modelos de espera no pueden tomar en cuenta el comportamiento *individual* de los clientes, en el sentido de que se espera que todos los clientes formados en una línea de espera se "comporten" en la misma forma mientras permanecen en la instalación. Por lo tanto, un cliente "parlanchín" se considera un caso raro y su conducta es ignorada en el diseño del sistema. Por otra parte, si sucede que la mayoría de los clientes son excesivamente parlanchines, un diseño *realista* de la instalación de servicio deberá estar basado en el hecho de que este hábito, por pródigo que pueda ser, es una parte integral de la operación. Una manera lógica de incluir el efecto de este hábito consiste en aumentar el tiempo de servicio por cliente.

Ahora vemos que los elementos básicos de un modelo de espera dependen de los siguientes factores:

1. Distribución de llegadas (llegadas individuales o masivas en grupo).
2. Distribución del tiempo de servicio (servicio individual o masivo).
3. Diseño de la instalación de servicio (estaciones en serie, en paralelo o en red).
4. Disciplina de servicio (FCFS, LCFS, SIRO) y prioridad de servicio.
5. Tamaño de la línea de espera (finito o infinito).
6. Fuente de llamadas (finita o infinita).
7. Conducta humana (cambios, elusión y renuncia).

Existen tantos modelos de espera como variaciones de los factores citados. En este capítulo consideramos varios modelos que parecen ser útiles en aplicaciones prácticas. En la sección que sigue se indica que las distribuciones de Poisson y exponencial, desempeñan una función importante en la representación de las llegadas

Una notación que es en particular adecuada para resumir las características principales de las líneas de espera *en paralelo* se ha estandarizado universalmente en el formato siguiente,

$$(a/b/c):(d/e/f)$$

donde los símbolos a, b, c, d, e y f representan elementos básicos del modelo en la forma siguiente (véase la sección 15.1).

$a \equiv$ distribución de llegadas
$b \equiv$ distribución del tiempo de servicio (o de salidas)
$c \equiv$ número de servidores en paralelo ($c = 1, 2, \ldots, \infty$)
$d \equiv$ disciplina de servicio (por ejemplo, FCFS, LCFS, SIRO)
$e \equiv$ número máximo admitido en el *sistema* (en línea de espera + en servicio)
$f \equiv$ tamaño de la fuente de llamadas

La notación estándar reemplaza los símbolos a y b de llegadas y salidas por los códigos siguientes.

$M \equiv$ distribución de llegadas o salidas de Poisson (o markoviana); o, lo que es lo mismo, distribución exponencial entre llegadas o de tiempo de servicio
$D \equiv$ tiempo entre llegadas o de servicio constante o determinista
$E_k \equiv$ distribución de Erlang o gamma de la distribución de tiempo entre llegadas o de servicio con el parámetro k
$GI \equiv$ distribución de llegadas general independiente (o tiempo entre llegadas)
$G \equiv$ distribución de salidas general (o tiempo de servicio)

Para ilustrar la notación, considérese (*DG*, disciplina general en notación de Kendall)

$$(M/D/10):(DG/N/\infty)$$

Aquí tenemos llegadas de Poisson, tiempo de servicio constante y 10 servidores en paralelo en la instalación. La disciplina de servicio es general (*DG*) en el sentido de que pudiera ser FCFS, LCFS, SIRO o cualquier procedimiento que puedan utilizar los servidores para decidir el orden en el que se escogerá a los clientes de la línea de espera para iniciar el servicio. Independientemente de cuántos clientes lleguen a la instalación, el sistema (línea de espera + servicio) puede alojar sólo a un máximo de N clientes; todos los demás deberán buscar ser atendidos en cualquier otra parte. Por último, la fuente que genera los clientes que entran a la instalación tiene una capacidad infinita.

La notación estándar descrita fue ideada originalmente por D. G. Kendall [1953] en la forma $(a/b/c)$ y se le conoce en la literatura como **notación de Kendall.** Después, A. M. Lee [1966] agregó los símbolos d y e a la notación de Kendall. En este libro advertimos que conviene aumentar la notación de Kendall-Lee mediante el uso del símbolo f, que representa la capacidad de la fuente de llamadas.

El objetivo final de analizar situaciones de espera consiste en generar medidas de desempeño para evaluar los sistemas reales. No obstante, como cualquier sistema de espera opera como función del tiempo, debemos decidir con anticipación si nos interesa analizar el sistema en condiciones **transitorias** o **de estado estable.** Las condiciones transitorias prevalecen cuando el comportamiento del sistema sigue dependiendo del tiempo. Por lo tanto, los procedimientos de nacimiento y muerte pura (sección 15.3) siempre operan en condiciones transitorias. Por otra parte, las líneas de espera con llegadas y salidas combinadas inician en condiciones transitorias y llegan gradualmente al estado estable después de haber transcurrido un tiempo lo *suficientemente grande*, siempre que los parámetros del sistema permitan se alcance el estado estable (por ejemplo, una línea de espera con tasa de llegadas λ mayor que su tasa de salidas μ nunca alcanzará el estado estable, sin importar el tiempo transcurrido, ya que el tamaño de la línea de espera aumentará con el tiempo). Advertimos que el análisis de estado transitorio es bastante complejo, desde el punto de vista matemático, por lo que no se considerará en esta presentación.

El resto de esta sección desarrolla primero un modelo de estado estable para la línea de espera generalizada de Poisson con c servidores en paralelo. El resultado principal que se obtiene de este modelo es la determinación de las probabilidades de estado estable de tener n clientes en el sistema. Estas probabilidades se usan para desarrollar las medidas de desempeño del modelo generalizado de líneas de espera.

15.5 LINEAS DE ESPERA ESPECIALIZADAS DE POISSON

4

En esta sección aprovechamos los resultados del modelo generalizado dado en la sección 15.4.1 para estudiar las colas especializadas de Poisson. Cada modelo se describe en términos de la notación extendida de Kendall, presentada en la sección 15.4. Como la deducción de p_n en la sección 15.4.1 es completamente independiente de la disciplina de la línea de espera, es apropiado usar el símbolo *DG* (disciplina general) en la notación de Kendall.

15.5.1 (*M*/*M*/1) : (*DG*/∞/∞)

Este es un modelo de servidor único sin límite en la capacidad del sistema o de la fuente de llamadas. Se supone que las tasas de llegadas son independientes del número en el sistema; o sea, $\lambda_n = \lambda$ para toda n. Similarmente, se supone que el servidor único en el sistema completa el servicio a una tasa constante; o sea, $\mu_n = \mu$ para toda n. En efecto, el presente modelo tiene llegadas y salidas de Poisson con tasas medias λ y μ, respectivamente. De la sección 15.4.2 se infiere que $\lambda_{ef} = \lambda$.

Definiendo $\rho = \lambda/\mu$, la expresión para p_n para n en el modelo generalizado se reduce a

$$p_n = \rho^n p_0, \qquad n = 0, 1, 2, \ldots$$

Determinamos p_0 considerando que la suma de todas las p_n para $n = 0, 1, 2, \ldots$ es igual a 1; tenemos

$$p_0(1 + \rho + \rho^2 + \cdots) = 1$$

Si suponemos que $\rho < 1$, entonces

$$p_0\left(\frac{1}{1-\rho}\right) = 1$$

o bien

$$p_0 = 1 - \rho$$

Obtenemos así la siguiente fórmula general:

$$p_n = (1 - \rho)\rho^n, \qquad n = 0, 1, 2, \ldots \qquad (M/M/1):(DG/\infty/\infty)$$

que es una **distribución geométrica.**

El requisito matemático de que $\rho < 1$, necesario para garantizar la convergencia de la serie geométrica $(1 + \rho + \rho^2 + \cdots)$ conduce a un argumento intuitivo. Fundamentalmente, $\rho < 1$ significa que $\lambda < \mu$ lo que establece que la tasa de llegadas debe ser estrictamente menor que la tasa de servicio en la instalación, para que el sistema alcance estabilidad (condiciones de estado estable). Esto tiene sentido porque bajo otras condiciones, el tamaño de la línea de espera crecería indefinidamente.

La medida L_s se puede deducir de la siguiente manera:

$$\begin{aligned} L_s &= \sum_{n=0}^{\infty} np_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n \\ &= (1-\rho)\rho \frac{d}{d\rho} \sum_{n=0}^{\infty} \rho^n \\ &= (1-\rho)\rho \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) \\ &= \frac{\rho}{1-\rho} \end{aligned}$$

Obsérvese que la convergencia de $\sum_{n=0}^{\infty} \rho^n$ se garantiza porque $\rho < 1$. Mediante el uso de las relaciones anteriores, obtenemos todas las medidas básicas de desempeño como

$$L_s = E\{n\} = \frac{\rho}{1-\rho}$$

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

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu(1-\rho)}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\rho}{\mu(1-\rho)}$$

Ejemplo 15.5-1

En una instalación de servicio de lavado de autos, la información que se tiene indica que los autos llegan para ser atendidos, según una distribución de Poisson, con media de 4 por hora. El tiempo para lavar y asear cada automóvil varía, pero se advierte que sigue una distribución exponencial con media de 10 minutos por automóvil. La instalación no puede atender a más de un auto a la vez.

Para analizar esta situación utilizando los resultados del $(M/M/1):(DG/\infty/\infty)$, debemos suponer que la fuente de llamadas es tan grande que se puede considerar infinita. Además, existe suficiente espacio de estacionamiento para dar cabida a todos los autos que lleguen.

En la situación dada tenemos $\lambda = 4$ autos por hora y $\mu = 60/10 = 6$ autos por hora. Como $\rho = \lambda/\mu = 4/6$ es menor que 1, el sistema puede operar en condiciones de estado estable. La salida del modelo se muestra en la figura 15-5. El valor de $L_q = 1.33$ autos nos da una idea acerca de cuántos autos están esperando, en promedio, cuando llega un cliente. Sin embargo, observamos que L_q representa un valor esperado, de modo que el número de autos en espera, en cualquier momento, puede ser mayor o menor que 1.333 autos. Podría entonces interesarnos determinar el número de lugares de estacionamiento, de manera que se asocie una probabilidad razonable de que un auto que llegue encuentre lugar. Por ejemplo, suponga que necesitamos proporcionar suficientes lugares de estacionamiento, de modo que un auto que llegue encuentre un lugar por lo menos el 90% de las veces.

Problem title: Example 15.5-1
[Título del problema: Ejemplo 15.5-1]
Scenario 1 — (M/M/1):(GD/*/*)
[Situación 1 — (M/M/1):(DG/*/*)]

Lambda =	4.00000	Lambda eff = Lambda ef =	4.00000
Mu =	6.00000	Rho = Ro =	0.66667
Ls =	2.00000	Lq =	1.33333
Ws =	0.50000	Wq =	0.33333

Values of p(n) for n=0 to 25, else p(n) < .00001
[Valores de p(n) para n = 0 hasta 25, si no p(n) < 0.00001]

0 0.33333	1 0.22222	2 0.14815	3 0.09877	4 0.06584
5 0.04390	6 0.02926	7 0.01951	8 0.01301	9 0.00867
10 0.00578	11 0.00385	12 0.00257	13 0.00171	14 0.00114
15 0.00076	16 0.00051	17 0.00034	18 0.00023	19 0.00015
20 0.00010	21 0.00007	22 0.00004	23 0.00003	24 0.00002
25 0.00001				

Cumulative values of p(n) for n=0 to 25
[Valores acumulados de p(n) para n = 0 hasta 25]

0 0.33333	1 0.55556	2 0.70370	3 0.80247	4 0.86831
5 0.91221	6 0.94147	7 0.96098	8 0.97399	9 0.98266
10 0.98844	11 0.99229	12 0.99486	13 0.99657	14 0.99772
15 0.99848	16 0.99898	17 0.99932	18 0.99955	19 0.99970
20 0.99980	21 0.99987	22 0.99991	23 0.99994	24 0.99996
25 0.99997				

15.5.2 $(M/M/1):(DG/N/\infty)$ 5

La única diferencia entre este modelo y el $(M/M/1):(DG/\infty/\infty)$ es que el número máximo de clientes permitido en el sistema es N (longitud máxima de la línea de espera $= N-1$). Esto significa que cuando haya N clientes en el sistema, se impiden todas las nuevas llegadas o no se les permite unirse a la línea de espera. En términos del modelo generalizado esta situación se traduce en

$$\lambda_n = \begin{cases} \lambda, & n = 0, 1, 2, \ldots, N-1 \\ 0, & n = N, N+1, \ldots \end{cases}$$

$$\mu_n = \mu \qquad \text{para toda } n = 0, 1, 2, \ldots$$

Haciendo $\rho = \lambda/\mu$, obtenemos

$$p_n = \begin{cases} \rho^n p_0, & n \leq N \\ 0, & n > N \end{cases}$$

El valor de p_0 se determina con la ecuación

$$\sum_{n=0}^{N} p_n = 1, \quad \text{que da} \quad p_0(1 + \rho + \rho^2 + \cdots + \rho^N) = 1$$

o bien

$$p_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{N+1}}, & \rho \neq 1 \\ \dfrac{1}{N+1}, & \rho = 1 \end{cases}$$

Las fórmulas para p_n pueden resumirse como

$$p_n = \begin{cases} \dfrac{1-\rho}{1-\rho^{N+1}}\rho^n, & \rho \neq 1 \\ \dfrac{1}{N+1}, & \rho = 1 \end{cases} \qquad n = 0, 1, 2, \ldots, N \qquad (M/M/1):(DG/N/\infty)$$

~~Obsérvese que~~ $\rho = \lambda/\mu$ *no* necesita ser menor que 1 como en el caso del modelo $(M/M/1):(DG/\infty/\infty)$. En forma intuitiva advertimos este resultado porque el número que se admite en el sistema está controlado por la longitud de la línea de espera $(= N-1)$, y no por las tasas relativas de llegadas y salidas, λ y μ.

Usando el valor anterior de p_n, encontraremos que el número esperado en el sistema se calcula como sigue:

$$
\begin{aligned}
L_s = E\{n\} &= \sum_{n=0}^{N} np_n \\
&= \frac{1-\rho}{1-\rho^{N+1}} \sum_{n=0}^{N} n\rho^n \\
&= \frac{1-\rho}{1-\rho^{N+1}}\, \rho \frac{d}{d\rho}\left(\frac{1-\rho^{N+1}}{1-\rho}\right) \\
&= \frac{\rho\{1-(N+1)\rho^N + N\rho^{N+1}\}}{(1-\rho)(1-\rho^{N+1})}
\end{aligned}
$$

o bien

$$
\boxed{L_s = \begin{cases} \dfrac{\rho\{1-(N+1)\rho^N + N\rho^{N+1}\}}{(1-\rho)(1-\rho^{N+1})}, & \rho \neq 1 \\[2ex] \dfrac{N}{2}, & \rho = 1 \end{cases}}
$$

Las medidas L_q, W_s y W_q se pueden deducir a partir de L_s, una vez que se determina la tasa efectiva de llegada λ_{ef}. De la sección 15.4.2 tenemos

$$\lambda_{ef} = \lambda(p_0 + p_1 + \cdots + p_{N-1}) + 0p_N$$

Obtenemos así

$$\boxed{\lambda_{ef} = \lambda(1 - p_N)}$$

La fórmula para λ_{ef} tiene sentido porque la probabilidad de que un cliente no sea capaz de unirse al sistema es p_N. Así, la probabilidad de clientes que pueden unirse al sistema es $(1 - p_N)$, lo que conduce directamente a la fórmula para λ_{ef}. Usando L_s y λ_{ef}, obtenemos

$$
\boxed{
\begin{aligned}
L_q &= L_s - \frac{\lambda_{ef}}{\mu} = L_s - \frac{\lambda(1-p_N)}{\mu} \\
W_q &= \frac{L_q}{\lambda_{ef}} = \frac{L_q}{\lambda(1-p_N)} \\
W_s &= W_q + \frac{1}{\mu} = \frac{L_s}{\lambda(1-p_N)}
\end{aligned}
}
$$

Se puede demostrar que

$$\boxed{\lambda_{ef} = \mu(L_s - L_q) = \lambda(1 - p_N)}$$

(véase el problema 15-58).

Ejemplo 15.5-2

Considere la instalación de lavado de autos del ejemplo 15.5-1. Suponga que la instalación tiene un total de 4 lugares de estacionamiento. Si el estacionamiento está lleno, los carros que lleguen buscarán servicio en otra parte.

Observando que $N = 4 + 1 = 5$, encontramos que la salida de TORA del modelo es como el que se da en la figura 15.6. Una información que puede interesar al propietario de la instalación es saber cuántos clientes se pierden debido al limitado espacio de estacionamiento. Esto equivale a determinar el valor de λ_{pN} o bien, equivalentemente, $\lambda - \lambda_{ef}$. De la figura 15-6,

$$\lambda - \lambda_{ef} = 4 - 3.8075 = 0.1925 \text{ autos/hora}$$

o bien, con base en un día de ocho horas, la instalación pierde aproximadamente 2 ($\cong 8 \times 0.1925$) autos por día en promedio o sea, $(2/4 \times 8) \times 100 = 6.25\%$ de todos los carros que llegan por día. Una decisión respecto a incrementar el tamaño del estacionamiento a más de 4 lugares debe basarse en el "valor" de clientes perdidos.

```
Problem title: Example 15.5-2
[Título del problema: Ejemplo 15.5-2]
Scenario 1 — (M/M/1):(GD/5/*)
[Situación 1 — (M/M/1):(DG/5/*)]
--------------------------------------------------------------
Lambda =     4.00000     Lambda eff =    3.80752   ≃ 11 min.
                         Lambda ef =
Mu =         6.00000     Rho =           0.66667
                         Ro =
Ls =         1.42256     Lq =            0.78797
Ws =         0.37362     Wq =            0.20695
--------------------------------------------------------------
Values of p(n) for n=0 to 5, else p(n) < .00001
[Valores de p(n) para n = 0 hasta 5, si no p(n) < 0.00001]

 0 0.36541    1 0.24361    2 0.16241    3 0.10827    4 0.07218
 5 0.04812

Cumulative values of p(n) for n=0 to 5
[Valores acumulados de p(n) para n = 0 hasta 5]

 0 0.36541    1 0.60902    2 0.77143    3 0.87970    4 0.95188
 5 1.00000
```