

*Jorge Eduardo Ortiz Triviño*

---

# ***Modelos estocásticos y simulación***

---

# Índice general

---

Índice de figuras	III
<b>I TEORÍA DE TELETRÁFICO</b>	<b>1</b>
<b>1. SISTEMA DE COLAS SIMPLES</b>	<b>3</b>
1.1. INTRODUCCIÓN	4
1.2. EXPRESIÓN DE KENDALL Y MEDIDAS DE DESEMPEÑO.	4
1.2.1. Notación.	4
1.2.2. Medidas de desempeño.	5
1.3. SISTEMA DE COLAS $(M/M/1)$	6
1.3.1. Procesos binomiales y Procesos de Poisson.	6
1.3.2. Dos propiedades importantes de los procesos de Poisson.	8
1.3.3. Fundamentos del proceso de Poisson.	9
1.3.4. Media y varianza de un proceso de Poisson.	12
1.3.5. Tiempo entre llegadas.	13
1.3.6. Propiedad de pérdida de la memoria.	13
1.3.7. La propiedad de Markov.	15
1.3.8. Tiempos de servicio exponencial.	16
1.3.9. Fundamentos del sistema de colas $(M/M/1)$	16
1.4. EL SISTEMA DE FILAS $(M/M/1)$ EN DETALLE	21
1.4.1. Función de densidad	21
1.4.2. Utilización del servidor	22
1.4.3. Número medio de clientes en el sistema	22
1.4.4. Varianza del Número de clientes en el sistema.	23
1.5. LEY DE LITTLE	24
1.6. TEOREMA DE BURKE	26
1.7. RESUMEN Y PANORAMA HISTÓRICO	26
1.8. EJERCICIOS	27
<b>Bibliografía</b>	<b>29</b>



---

## *Índice de figuras*

---

1.1. Sistema (M/M/1) . . . . .	6
1.2. Bifurcación de un proceso de Poisson . . . . .	8
1.3. Unión de varios procesos de Poisson . . . . .	9
1.4. Diagrama de transición de estados de un proceso de Poisson .	10
1.5. Diagrama de Transición de Estados DTE del sistema (M/M/1)	17
1.6. Balance local en la frontera entre estados . . . . .	20
1.7. Número medio de $\mu_{N_t}$ clientes en un sistema (M/M/1) . . . .	23
1.8. Varianza del número de clientes $\sigma_{N_t}^2$ en un sistema (M/M/1)	24
1.9. Esquema de demostración de la “Ley de Little” . . . . .	25



Parte I

**TEORÍA DE  
TELETRÁFICO**



# 1

## *SISTEMA DE COLAS SIMPLES*

### CONTENIDO

1.1	INTRODUCCIÓN .....	3
1.2	EXPRESIÓN DE KENDALL Y MEDIDAS DE DESEMPEÑO. ....	4
	1.2.1 Notación. ....	4
	1.2.2 Medidas de desempeño. ....	5
1.3	SISTEMA DE COLAS ( $M/M/1$ ) .....	6
	1.3.1 Procesos binomiales y Procesos de Poisson. ....	6
	1.3.2 Dos propiedades importantes de los procesos de Poisson. ....	8
	1.3.3 Fundamentos del proceso de Poisson. ....	9
	1.3.4 Media y varianza de un proceso de Poisson. ....	11
	1.3.5 Tiempo entre llegadas. ....	13
	1.3.6 Propiedad de pérdida de la memoria. ....	13
	1.3.7 La propiedad de Markov. ....	15
	1.3.8 Tiempos de servicio exponencial. ....	15
	1.3.9 Fundamentos del sistema de colas ( $M/M/1$ ) .....	16
1.4	EL SISTEMA DE FILAS ( $M/M/1$ ) EN DETALLE .....	21
	1.4.1 Función de densidad .....	21
	1.4.2 Utilización del servidor .....	22
	1.4.3 Número medio de clientes en el sistema .....	22
	1.4.4 Varianza del Número de clientes en el sistema. ....	23
1.5	LEY DE LITTLE .....	24
1.6	TEOREMA DE BURKE .....	26
1.7	RESUMEN Y PANORAMA HISTÓRICO .....	26
1.8	EJERCICIOS .....	27



---

## 1.1. INTRODUCCIÓN

En este capítulo se presenta el principal modelo de colas simples. En capítulos posteriores se discutirán las redes de colas que son la extensión del modelo simple aquí estudiado. El modelo de colas simple es el ladrillo a partir del cual se construye el edificio de la planeación de redes y el análisis de desempeño en los sistemas de telecomunicaciones. La mayoría de trabajos de modelado que se han hecho para sistemas de colas involucra, necesariamente, al menos una cola simple. La principal razón es que para cualquier investigación en la teoría de colas, el sistema de cola simple es un punto de partida natural. Otra razón es la facilidad de su tratamiento: Es mucho más fácil formalizar los resultados con una cola simple para luego extenderlo a modelos más complejos de red de colas.

En particular, aquí se estudian los fundamentos del sistema de colas básico, llamado  $(M/M/1)$ : Este es un modelo de colas “Markoviano” en el que se distingue un proceso de llegada de Poisson y tiempo de servicio exponencial. Para representar la dinámica de ese sistema se emplean los diagramas de transición de estados.

---

## 1.2. EXPRESIÓN DE KENDALL Y MEDIDAS DE DESEMPEÑO.

### 1.2.1. Notación.

Para referirse formalmente a los principales elementos matemáticos que hacen parte de un sistema de líneas de espera (o colas) es ampliamente aceptada la notación de seis (6) parámetros propuesta por Kendall. En dicha convención se expresa un sistema de líneas de espera mediante el siguiente formato de la expresión 1.1.

$$(a/b/c) : (d/e/f) \tag{1.1}$$

Cada uno de los tres parámetros que conforman las dos secciones presentes en la expresión 1.1 tienen un significado asociado. En la primera sección, el parámetro  $a$  representa la estructura (o familia) probabilística que guía la llegada de clientes al sistema; el parámetro  $b$  se emplea para indicar la estructura probabilística del tiempo de atención por parte del servidor o de los servidores. Ellos usualmente se sustituyen por las letras mayúsculas  $M$  que

indica “*Markoviano*” y que implica que el número de clientes corresponde a un proceso de *Poisson*; la letra *D* significa “*determinístico*” o “*degenerado*” queriendo decir que los tiempos de llegada (o atención según el caso) son constantes y no variables aleatorias. La letra *G* se utiliza para indicar que la familia probabilística es “*general*” es decir cualquiera. Por su lado, el último parámetro de la primera sección,  $c \in \mathbb{N}$ , es una constante natural que indica el número de servidores en paralelo que trabajan en el sistema (usualmente  $c = 1$ ).

Los parámetros de la segunda sección de la referida expresión también tienen sus significados específicos. El parámetro  $d$  se emplea para indicar la disciplina de servicio; esto es, el mecanismo que empleará el servidor para seleccionar al siguiente cliente que él va a atender una vez se desocupe del cliente que está actualmente atendiendo. Este parámetro es frecuentemente FIFO (primero en llegar primero en ser atendido); sin embargo, hay otras opciones. Por ejemplo LIFO (último en llegar, primero en ser atendido). Por su parte, el parámetro  $e \in \mathbb{N}$  representa la capacidad del sistema; es decir, el número máximo de clientes (simultáneos) que pueden permanecer dentro del sistema. Finalmente, el tamaño de la población, origen de los clientes, lo representa el parámetro  $f$ .

#### **Ejemplo 1—1: Un sistema de líneas de espera.**

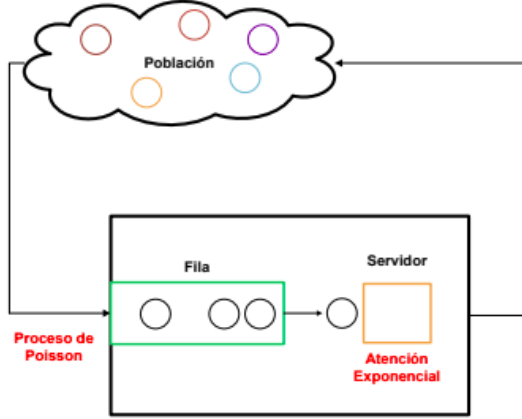
La expresión  $(\text{Cauchy}(\alpha, \beta)/\text{Uniforme}(\eta, \mu)/7):(\text{LIFO}/180/1.000.000)$  describe un sistema de líneas de espera con tiempos entre llegadas de la familia Cauchy, tiempo de atención en los 7 servidores en paralelo de la familia uniforme. Política de atención LIFO, máximo 180 clientes dentro del sistema (capacidad de la fila  $180-7=173$ ) y tamaño de la población de donde provienen los clientes de un millón.

**Observación:** Cuando en esta nomenclatura se suprime la segunda sección se supone: Política FIFO, Capacidad del sistema y tamaño de la población infinitos. Esquemáticamente eso significa:

$$(a/b/c) = (a/b/c) : (\text{FIFO}/+\infty/+\infty) \quad (1.2)$$

#### **1.2.2. Medidas de desempeño.**

Especificar el sistema no es suficiente. A partir de esa especificación formal se deben obtener medidas que describan de alguna manera el comportamiento o desempeño de ese sistema. Aunque esas medidas de desempeño pueden ser de diferentes formas y de distintas propiedades, es frecuente intentar calcular: (1) El número medio de clientes en el sistema, (2) El número de clientes esperado en fila, (3) El tiempo promedio dentro del sistema, (4) El tiempo medio de permanencia en cola (fila) y (5) La utilización, que se define como el porcentaje de tiempo que los servidores permanecen ocupados. Aquí solo nombramos cinco pero la lista puede crecer tanto como se quiera.



**Figura 1.1**  
Sistema (M/M/1)

### 1.3. SISTEMA DE COLAS (M/M/1)

Considérese el sistema de la Figura 1.1. Ese es el sistema de colas simple del cual parte la teoría de colas. El sistema (M/M/1) tiene dos supuestos importantes. El primero es que el proceso de llegadas es un proceso de Poisson, y el otro que cuenta con un único servidor razón por la cual tiene un único tiempo de servicio que corresponde a una variable con distribución exponencial. Estos supuestos llevan a un modelo muy manejable y además razonable, para una gran variedad de situaciones. Se examinarán ahora estos dos supuestos.

#### 1.3.1. Procesos binomiales y Procesos de Poisson.

El *proceso de Poisson* es el proceso de llegada más sencillo. Una forma de entenderlo es esta. Suponga que el eje temporal se divide en un gran número de pequeños segmentos de longitud  $\Delta t$ . Se supone que la probabilidad de que un único cliente llegue a un segmento es proporcional a la longitud de ese segmento,  $\Delta t$ , multiplicado por una constante de proporcionalidad,  $\lambda$ , que representa la tasa media de llegadas:

$$P[\text{Exactamente 1 llegada en } [t, t + \Delta t]] = \lambda \Delta t \quad (1.3)$$

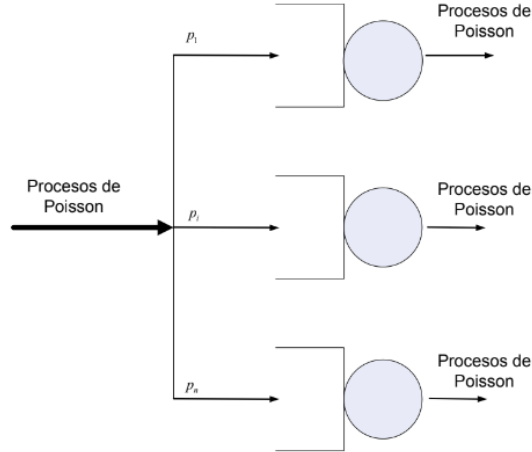
$$P[\text{No llega en } [t, t + \Delta t]] = 1 - \lambda \Delta t \quad (1.4)$$

$$P[\text{Más de 1 llegada en } [t, t + \Delta t]] = 0 \quad (1.5)$$

Aquí (dada su magnitud) se ignoran los términos de orden superior que involucren  $\Delta t$ . Es posible hacer una analogía entre el proceso de llegada y el lanzamiento de una moneda en cada uno de los segmentos. Aquí la probabilidad de llegada es  $\lambda\Delta t$  (es decir, cara) y  $1 - \lambda\Delta t$  es la probabilidad de no llegada (es decir, sello). Los lanzamientos, y en consecuencia los resultados, de cada moneda son, a simple vista, independientes unos con otros; por esta razón, al observar el comportamiento del sistema en un intervalo de tiempo  $[0, t)$  en el cual se pueda escribir  $t = n\Delta t$  con  $n \in \mathbb{N}$  y, en consecuencia,  $t \in \mathbb{R}^+$ , es claro que al definir la variable aleatoria  $N_t$  como el número de clientes que arriban al sistema en  $[0, t)$  su estructura probabilística es  $N_t \sim \text{Binomial}(n, p = \lambda\Delta t)$  o, más específicamente,  $f_{N_t}(k; n, p = \lambda\Delta t) = P[N_t = k]$ , está dada en la ecuación 1.6.

$$f_{N_t}(k; n, p = \lambda\Delta t) = \binom{n}{k} (\lambda\Delta t)^k (1 - \lambda\Delta t)^{n-k} \quad (1.6)$$

De ahí que la estructura probabilística dependa del tiempo de observación del sistema  $t$  y que, por lo tanto, rigurosamente hablando, a este modelo se le conozca con el nombre de *Procesos estocásticos binomiales de nacimiento puro*. Cuando  $\Delta t \rightarrow 0$  se obtiene un proceso de Poisson de tiempo continuo puesto que, como se mostrará más adelante, una familia binomial converge a una familia de Poisson cuando  $n \rightarrow \infty$ ,  $p = \lambda\Delta t \rightarrow 0$  y  $\mu_{N_t} = p \times n = \lambda$  permanece constante. En este escenario, cada lanzamiento de una moneda puede ser una analogía de las llegadas, unas independientes de otras, ya que se puede pensar en un simple resultado positivo de una larga lista de lanzamientos de monedas independientes. Además, se puede ver todos los intervalos de tiempo tienen la misma probabilidad constante de tener una llegada.

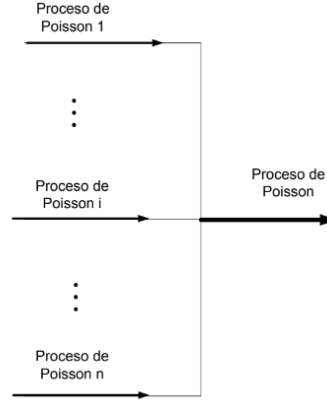
**Figura 1.2**

Bifurcación de un proceso de Poisson

### 1.3.2. Dos propiedades importantes de los procesos de Poisson.

Dos ideas importantes caracterizan a los procesos de Poisson. La primera es el proceso de bifurcación aleatoria de un proceso de Poisson y la segunda es la unión de varios procesos de Poisson. La Figura 1.2 ilustra un sistema de filas con variables ramificación o bifurcaciones. La llegada se divide aleatoriamente entre  $n$  ramas con probabilidades independientes  $p_1, \dots, p_i, \dots, p_n$ . La segunda, que se ilustra en la Figura 1.3, es una situación que envuelve la unión de procesos de Poisson independientes. En este caso, las llegadas de un número de procesos independientes se unen o fusionan en un único proceso agregado. Una de las aplicaciones originales, aunque no la única ni la más importante, de los procesos de Poisson en redes de comunicaciones es el modelo de llegadas de llamadas a una central telefónica.

El uso de cada teléfono, por lo menos en un primer análisis, puede ser modelado por un proceso de Poisson. Así, modelar puede convertirse en algo complicado, pero es indudable la elegancia de formular este tipo de modelos para realizar planeación de redes de computadores y de telecomunicaciones de una manera formal y bella. El punto importante está en que tanto los procesos de ramificaciones aleatorias como las uniones de procesos de Poisson resultan ser también Procesos de Poisson. La analogía del lanzamiento de monedas sirve para demostrar de una manera muy sencilla que estos resultados son verdad.

**Figura 1.3**

Unión de varios procesos de Poisson

**1.3.3. Fundamentos del proceso de Poisson.**

En este apartado se deduce una ecuación diferencial que se obtiene del proceso de Poisson a partir del modelo discreto resulta de partir de pequeños intervalos  $\Delta t$  de tiempo para, finalmente, tomar el límite cuando  $\Delta t \rightarrow 0$ . En primer lugar

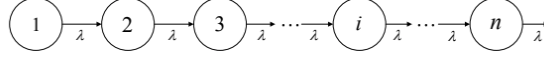
$$P_n(t) = P(\# \text{ llegan} = n \text{ en tiempo } t) \quad (1.7)$$

Además, sea  $P_{ij}(\Delta t)$  la probabilidad de pasar de  $i$  llegadas (nacimientos) a  $j$  en  $\Delta t$  segundos. El proceso de deducción que aquí se presenta está inspirado en el libro [1]. En este modelo, el número de llegadas (o nacimientos) es el “estado” del sistema. Este dato contiene toda la información necesaria para describir el sistema de una manera adecuada. Dicho esto, es claro, entonces que para esta situación se cumple la ecuación 1.8

$$P_n(t + \Delta t) = P_n(t) P_{n,n}(\Delta t) + P_{n-1}(t) P_{n-1,n}(\Delta t) \quad (1.8)$$

Nuevamente aquí, no se tiene en cuenta (por ser de magnitudes muy pequeñas) los términos de orden superior en  $\Delta t$ . Lo que dice la ecuación 1.8 es que el sistema puede llegar al estado  $n$ , esto tener clientes o nacimiento en el tiempo en el intervalo  $[0, t + \Delta t)$  ya sea porque en el intervalo  $[0, t)$  el sistema ya se encontraba en el estado  $n$  y que durante el intervalo  $[t, t + \Delta t)$  no haya habido ninguna llegada, o bien, que en el intervalo  $[0, t)$  el sistema estaba en el estado  $n - 1$  y que durante el intervalo  $[t, t + \Delta t)$  haya habido una llegada. Nótese que

se puede asumir que  $\Delta t$  es suficientemente pequeño de modo que es razonable suponer que a lo sumo un cliente puede llegar durante este intervalo.



**Figura 1.4**

Diagrama de transición de estados de un proceso de Poisson

Para este sistema, el diagrama de transición de estados se presenta en la Figura 1.4 1—4, en él los círculos representan los estados del sistema (número de llegadas o nacimientos) y  $\lambda$  representa la razón (clientes por unidad de tiempo) asociada con cada transición. Adicionalmente, se necesita la condición de frontera, es decir la probabilidad para el estado  $n = 0$ , cuya expresión se muestra en la ecuación 1.9

$$P_0(t + \Delta t) = P_0(t) P_{0,0}(\Delta t) \quad (1.9)$$

Las expresiones 1.8 y 1.9 son suficientes para determinar completamente  $P_n(t)$ . Al sustituir las probabilidades de las expresiones 1.3, 1.4 y 1.5 en las ecuaciones 1.8 y 1.9 se obtiene 1.10.

$$P_n(t + \Delta t) = P_n(t) (1 - \lambda \Delta t) + P_{n-1}(t) (\lambda \Delta t) \quad (1.10)$$

A partir de 1.10 y aplicando algunas operaciones elementales y reorganizando se llega a 1.11.

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = -\lambda P_n(t) + \lambda P_{n-1}(t) \quad (1.11)$$

Y de forma similar a partir de 1.9 se llega a 1.12

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda P_0(t) \quad (1.12)$$

Al tomar el límite cuando, en 1.11 y 1.12,  $\Delta t \rightarrow 0$ , este conjunto de ecuaciones se obtiene 1.13. En ella  $n \geq 1$ , y se desea, naturalmente, encontrar la solución a estas ecuaciones diferenciales,  $P_n(t)$ .

$$\frac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t) \quad (1.13)$$

y

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) \quad (1.14)$$

De la ecuación 1.14, es trivial, observar que su solución es 1.15

$$P_0(t) = e^{-\lambda t} \quad (1.15)$$

Para  $n = 1$  se parte de la ecuación 1.13 en la cual se reemplaza el resultado 1.15 con lo cual se llega a la ecuación diferencial 1.16.

$$\frac{dP_1(t)}{dt} = -\lambda P_1(t) + \lambda e^{-\lambda t} \quad (1.16)$$

La solución de la ecuación diferencial 1.16 esta dada en 1.17.

$$P_1(t) = \lambda t e^{-\lambda t} \quad (1.17)$$

De forma similar la siguiente ecuación diferencial, con  $n = 2$ , es 1.18.

$$\frac{dP_2(t)}{dt} = -\lambda P_2(t) + \lambda^2 t e^{-\lambda t} \quad (1.18)$$

Por su parte, la solución de 1.18 es 1.19.

$$P_2(t) = \frac{\lambda^2 t^2}{2} e^{-\lambda t} \quad (1.19)$$

Continuando con un proceso idéntico, se puede encontrar por inducción matemática que la solución general de la ecuación diferencial dada en 1.13 es la función 1.20.

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (1.20)$$

Esta es, como puede notarse por simple inspección, que 1.20 es la *distribución de Poisson*. En este caso modela la probabilidad de  $n$  llegadas (nacimientos) en un intervalo de  $n$  segundos para un proceso de Poisson de parámetro  $\lambda$ .

**Ejemplo 1—2: Proceso de Poisson.**

Una central telefónica recibe en promedio 100 llamadas por minuto, de acuerdo con un proceso de Poisson. ¿Cuál es la probabilidad de que no entren llamadas en un intervalo de 5 segundos?

La solución es simple.  $P_0\left(\frac{1}{12}\right) = e^{-\lambda t} = e^{100 * \frac{1}{12}} = 0,00024$



### 1.3.4. Media y varianza de un proceso de Poisson.

Ahora se hará la deducción de la media. Ahora se hará la deducción de la media  $\mu_t$  y la varianza  $\sigma_t^2$  de una distribución de *Poisson*. Sea  $\mu_t$  el número medio de llegadas en un intervalo de longitud  $t$ . De esta manera los cálculos presentados en 1.21 son los primeros pasos para obtener el resultado. La primera igualdad corresponde a la definición, en la segunda se emplea la expresión 1.20 mientras que en la tercera y última serie de igualdades se factorizan aquellos elementos que no dependen del índice de la sumatoria.

$$\mu_t = E[N_t] = \sum_{n=0}^{\infty} n \left[ \frac{(\lambda t)^n}{n!} e^{-\lambda t} \right] = e^{-\lambda t} \sum_{n=1}^{\infty} n \frac{(\lambda t)^n}{(n-1)!} \quad (1.21)$$

La deducción continúa en 1.22 con un cambio de variable muy simple,  $m = n - 1$ , y luego, nuevamente, factorizando elementos que no son funcionalmente dependientes del índice de la sumatoria y, finalmente, recordando la expansión en series de Taylor de la función exponencial.

$$\mu_t = e^{-\lambda t} \sum_{m=0}^{\infty} (m+1) \frac{(\lambda t)^{m+1}}{(m+1)!} = e^{-\lambda t} \lambda t \sum_{m=0}^{\infty} \frac{(\lambda t)^m}{m!} = e^{-\lambda t} \lambda t (e^{\lambda t}) = \lambda t \quad (1.22)$$

La ecuación  $\mu_t = \lambda t$  es un resultado muy intuitivo y, por la misma razón, muy bella. Ella indica que el número medio  $\mu_t$  es proporcional a la longitud del intervalo  $t$  y a la razón de llegadas de clientes por unidad de tiempo  $\lambda$ . Por su parte, para la varianza  $\sigma_t^2$ , el tratamiento adoptado en este texto es similar al presentado en [1]. Sin embargo, antes de entrar de lleno en su deducción, es útil calcular la siguiente esperanza matemática que aplica para un proceso de Poisson.

$$E[N_t(N_t - 1)] = \sum_{n=0}^{\infty} n(n-1) \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \sum_{n=2}^{\infty} \frac{(\lambda t)^n}{(n-2)!} e^{-\lambda t} \quad (1.23)$$

Ahora, haciendo el cambio de variable  $m = n - 2$  y luego recordando la expansión en series de Taylor para la función exponencial se obtiene 1.24.

$$E[N_t(N_t - 1)] = e^{-\lambda t} (\lambda t)^2 \sum_{m=0}^{\infty} \frac{(\lambda t)^m}{m!} = (\lambda t)^2 \quad (1.24)$$

Con este resultado es fácil deducir la variabilidad en un proceso de Poisson. En primer lugar, la identidad de la varianza 1.25 brinda un camino.

$$\sigma_t^2 = E[N_t^2] - \mu_t^2 \quad (1.25)$$

Obsérvese que basta con calcular  $E[N_t^2]$  puesto que  $\mu_t^2 = (\lambda t)^2$  como consecuencia de 1.22. Ahora bien, este dato faltante puede encontrarse como se indica en 1.26. Ahí se aplican los resultados obtenidos en 1.22 y 1.24.

$$E[N_t^2] = E[N_t] + E[N_t(N_t - 1)] = (\lambda t)^2 + \lambda t \quad (1.26)$$

Sustituyendo 1.26 y 1.22 en 1.25 se concluye 1.27, es decir la varianza del proceso.

$$\sigma_n^2 = (\lambda t)^2 + \lambda t - (\lambda t)^2 = \lambda t \quad (1.27)$$

En síntesis, el número medio de clientes y su varianza en un proceso de Poisson de nacimientos puros son, respectivamente,  $\mu_t = \lambda t$  y  $\sigma_n^2 = \lambda t$ .

### 1.3.5. Tiempo entre llegadas.

El tiempo entre eventos sucesivos en un proceso de llegada (o nacimiento puro) se denomina *tiempo inter-llegadas* o *tiempo entre llegadas*. Para un proceso de Poisson (de nacimiento puro), esos tiempos son variables aleatorias independientes, distribuidas exponencialmente. Para ver que esto es verdad, sea  $T$  el tiempo que transcurre entre dos llegadas sucesivas en un proceso de Poisson, la distribución de  $T$ , esto es  $F_r(t)$ , luego de unos breves pasos, es la expresión dada en 1.28.

$$F_r(t) = P[T \leq t] = 1 - P[T > t] = 1 - P_0(t) = 1 - e^{-\lambda t} \quad (1.28)$$

Al derivar 1.28 se encuentra  $f_r(t)$ , la función de densidad de  $T$ , ésta función se muestra en 1.29.

$$f_T(t) = \lambda e^{-\lambda t} \quad (1.29)$$

La expresión 1.29 es la estructura probabilística del tiempo entre llegadas  $T$  y, claramente, pertenece a la *familia paramétrica exponencial*.

### 1.3.6. Propiedad de pérdida de la memoria.

La familia exponencial, es la única variable aleatoria continua que tiene la *propiedad de pérdida de memoria*. El concepto se puede entender con un ejemplo sencillo. Suponga que un tren llega a una estación de acuerdo con un proceso de Poisson con tiempo medio entre trenes de 20 minutos.

Supóngase ahora que un pasajero llega a la estación y que alguien que

está esperando en la plataforma le informa que el último tren llegó hace 19 minutos. ¿Cuál es el tiempo medio que el pasajero que recién llega a la estación debe esperar antes de que el próximo tren llegue?

El sentido común (el cual no siempre es un buen consejero) diría que la respuesta natural es un minuto, y esto sería verdad en el caso en el cual la llegada del tren fuera un evento determinístico separado exactamente por veinte minutos. Sin embargo, como se trata de un proceso de llegadas de Poisson, ¡La respuesta es 20 minutos!

Intuitivamente, la explicación de ello es sencilla. Para ver que esto es verdad, es necesario volver atrás con la explicación de lanzamiento de monedas del proceso de Poisson. Una llegada es un evento con resultado positivo (“éxito”) de un “lanzamiento de monedas” en un intervalo infinitamente pequeño. A medida que el tiempo avanza es como un gran número de lanzamientos de monedas. Cuando se pone el proceso de Poisson en este contexto es fácil ver que la distribución hasta el próximo evento positivo, esto es “éxito”, no depende de cuando tiempo ha pasado desde el último evento positivo.

Sin embargo, el razonamiento intuitivo no es suficiente. La deducción formal o matemática, es decir, ya no intuitiva, es importante y se encuentra de la siguiente manera. Supóngase que una llegada ocurre en el tiempo  $t=0$ , se sabe, además de 1.29, que la estructura probabilística del tiempo antes de la siguiente llegada,  $T$ , es exponencial. También supóngase ahora que no han ocurrido llegadas antes del tiempo  $t_0$ . Dada esta información, se desea responder la pregunta ¿cuál es la distribución del tiempo antes de la siguiente llegada? La respuesta se deduce aquí de forma similar a la expuesta en [1] con la definición de una variable aleatoria ficticia  $T^*$  y, con ella, deducir su función de distribución  $F_{T^*}(t)$ .

Por definición de función de distribución  $F_{T^*}(t) = P[T^* \leq t]$  e incorporando la información disponible  $F_{T^*}(t) = P[T^* \leq t_0 + t | T^* > t_0]$ .

Ahora, teniendo en cuenta la definición de probabilidad condicional

$F_{T^*}(t) = \frac{P[T^* \leq t_0 + t, T^* > t_0]}{P[T^* > t_0]} = \frac{P[t_0 < T^* \leq t_0 + t]}{P[T^* > t_0]}$ . De acuerdo con las propiedades de la función de distribución y teniendo en cuenta la expresión 1.28

$F_{T^*}(t) = \frac{F_{T^*}(t+t_0) - F_{T^*}(t_0)}{1 - F_{T^*}(t_0)} = \frac{(1 - e^{-\lambda(t+t_0)}) - (1 - e^{-\lambda t_0})}{1 - (1 - e^{-\lambda t_0})}$  de donde se deduce 1.30.

$$F_{T^*}(t) = 1 - e^{-\lambda t} \quad (1.30)$$

Lo que, en palabras sencillas, significa que la estructura probabilística de  $T^*$ , el tiempo que se debe esperar hasta la siguiente llegada dado que se sabe que ese tiempo  $T^*$  es más grande que  $t_0$ , es una variable aleatoria perteneciente a la familia exponencial. Ello implica que  $\mu_{T^*} = \frac{1}{\lambda}$  y que  $\sigma_{T^*}^2 = \frac{1}{\lambda^2}$ .

Se termina este apartado señalando que la familia paramétrica *discreta* con la propiedad de pérdida de memoria es la distribución *geométrica* razón por la cual se dice que la familia exponencial es al mundo continuo lo que la familia geométrica al mundo discreto. Ambas familias juegan un papel protagónico en los modelos de tráfico en redes de telecomunicaciones.

### 1.3.7. La propiedad de Markov.

La propiedad de pérdida de memoria se define con mayor precisión como la *propiedad de Markov*. Esta propiedad establece que es posible predecir un estado futuro de un sistema basado (únicamente) en el estado sin necesidad de conocer completamente su historial pasado. En términos de una variable aleatoria real no negativa  $T \in \mathbb{R}^+$ , la propiedad de Markov se formaliza, de acuerdo con [1], en la expresión 1.31, para un  $t_0 > 0$  dado y específico y para cada  $t > 0$ .

$$P[T > t_0 + t | T > t_0] = P[T > t] \quad (1.31)$$

En términos de una variable aleatoria discreta  $X(t_n)$ , la misma propiedad se puede escribir, de acuerdo con [1], como en la expresión 1.32. Ahí, los  $t_i$  crecen monótonamente y los valores  $x_n$  pertenecen a un espacio de estados discretos.

$$P[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n, \dots, X(t_1) = x_1] = P[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n] \quad (1.32)$$

Los sistemas Markovianos son sistemas con pérdida de la memoria. Esto hace que sea “muy simple” su análisis teórico puesto que resulta relativamente fácil describir los cambios de estado del sistema, uno necesita tener en cuenta el tiempo desde la última llegada, o el tiempo que el servicio ha estado en curso desde el último cliente. Cuando el servicio no es Poisson, lo que significa que no se cuenta con la propiedad de pérdida de la memoria, o el tiempo de servicio no es exponencial entonces no se cuenta con las premisas necesarias para aprovechar ésta virtud. En este caso, el análisis del sistema, incluso en una cola simple, no es un trabajo trivial.

Un concepto estrechamente relacionado con la propiedad markoviana es el de *cadena de Markov*. Una cadena de Markov es un proceso (estocástico) de Markov, con un espacio de estados discreto. La Figura 1—4, muestra un diagrama de transición de estados, *DTE*, del sistema de líneas de espera ( $M/M/1$ ), en él se ilustra la sencillez que presume garantizar la propiedad de pérdida de la memoria para modelar su comportamiento.

### 1.3.8. Tiempos de servicio exponencial.

Una conclusión derivada de lo discutido en la sección 1.3.7 es que en un sistema  $(M/M/1)$  el proceso de llegada, el proceso de llegada, representado por la primera  $M$ , es “markoviano” lo que es equivalente a asegurar que (1) el número de clientes que llegan al sistema en un intervalo de tiempo de longitud  $t$  fijo es una variable aleatoria de Poisson, o bien que (2) el tiempo que transcurre entre la llegada del  $(i-1)$ -ésimo y es el  $i$ -ésimo clientes es una variable aleatoria exponencial  $T_i \sim \text{Exp}(\lambda)$  con independencia entre los  $T_i$ 's. En consecuencia, se ha reiterado bastante, la estructura probabilística de los tiempos entre llegadas es  $f_T(t) = \lambda e^{-\lambda t}$  y, por ello, el tiempo medio entre llegadas es  $\mu_T = \frac{1}{\lambda}$  mientras que la varianza del tiempo entre llegadas es  $\sigma_T^2 = \frac{1}{\lambda^2}$ .

Por otro lado, la segunda  $M$  en un sistema  $(M/M/1)$  indica que el tiempo  $S_i$  que el servidor emplea para atender al  $i$ -ésimo cliente también exhibe la propiedad markoviana; con todo lo que ello implica. Por ejemplo, implica la propiedad de pérdida de la memoria y la independencia entre los tiempos que el servidor “gasta” atendiendo al cliente  $i$ . También implica que los  $S_i \sim \text{Exp}(\lambda)$ . En específico  $f_S(s) = \mu e^{-\mu s}$  y, por ello, el tiempo medio de atención es  $\mu_T = \frac{1}{\mu}$  mientras que la varianza del tiempo de atención es  $\sigma_S^2 = \frac{1}{\mu^2}$ . En el caso de los tiempos de servicio como variables aleatorias exponenciales  $\mu$  es la tasa de atención (clientes por unidad de tiempo) mientras que  $\frac{1}{\mu}$  es la media del tiempo servicio (unidades de tiempo).

¿Qué tan realista es un tiempo de servicio de naturaleza exponencial? Se ha visto que es razonable como modelo de duración en conversaciones telefónicas. Pero se ha discutido también que tiene algún uso en situaciones donde la base física para su uso es más débil. Por ejemplo, el tiempo que transcurre entre paquetes de información de longitud fija a través de un canal en una red de computadores. A veces este tipo de hipótesis es hecha por conveniencia práctica y elegancia teórica, aunque en gran medida el realismo sea sacrificado. Con todas sus virtudes y debilidades, la teoría clásica de Teletráfico emplea la familia exponencial porque, de la mano con los procesos de Poisson, los dos conducen a los sistemas de tipo Markovianos que imprimen una elegante y muy loable forma matemática para modelar sistemas complejos de telecomunicaciones.

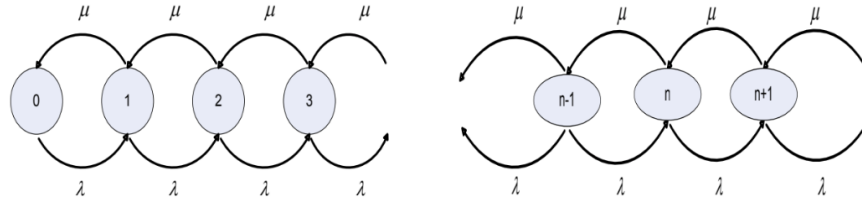
### 1.3.9. Fundamentos del sistema de colas $(M/M/1)$

El fundamento probabilístico del modelo del sistema de líneas de espera  $(M/M/1)$  se denomina *proceso de nacimiento y muerte*. En él, el estado del sistema  $N$  representa en número de clientes dentro del sistema en el instante  $t$ . Puesto que la dinámica del sistema admite “nacimientos”, es decir llegadas de clientes al sistema y “muertes” es decir salidas de clientes del sistema, el estado no es creciente como sí ocurre tanto en los procesos de “nacimientos

puros” o de “muertes puras” conforme aumenta  $t$  sino que algunas veces aumentará mientras que en otras disminuirá conforme el tiempo avanza.

En el área de computación, redes de computadores y telecomunicaciones hay varios ejemplos de sistemas cuyas dinámicas son, por naturaleza, procesos de nacimiento y muerte. La ejecución de procesos en un sistema multitarea, un enrutador en una red de computadores al cual llegan paquetes provenientes desde un computador fuente en la búsqueda de su destino final, es decir, otro computador en el sistema de red. El número de nuevas llamadas en una central telefónica, o el número de usuarios conectados en un instante de tiempo particular a un computador con el rol de servidor central. Estos, entre muchos otros campos de aplicación en computación, justifican su estudio matemático para aplicarlo en el diseño y la planeación de sistemas de cómputo, de redes de computadores y de sistemas de telecomunicaciones.

El tratamiento formal que se expone a continuación sigue un esquema similar al presentado en [1]. Al ser el proceso de llegadas y el proceso de atención modelos de tipo Markovianos, el Diagrama de Transición de estados, *DTE*, de un proceso de nacimiento y muerte se presenta en la Figura 1—5 1.5.



**Figura 1.5**

Diagrama de Transición de Estados DTE del sistema (M/M/1)

De la misma manera como se estudió en la sección 1.3.1, en donde se trataron los sistemas de “nacimiento puro”, se empleará la misma notación para el proceso de nacimiento involucrado aquí. Las expresiones 1.33, 1.34 y 1.35 son el fundamento del proceso de nacimiento incluido en el proceso general de nacimiento y muerte.

$$P[\text{Exactamente llegada en } [t, t + \Delta t]] = \lambda \Delta t \quad (1.33)$$

$$P[\text{Ninguna llega en } [t, t + \Delta t]] = 1 - \lambda \Delta t \quad (1.34)$$

$$P[\text{Mas de 1 llegada en } [t, t + \Delta t]] = 0 \quad (1.35)$$

También es necesario incorporar el proceso (de muerte) de acuerdo con el

análisis de la sección 1.3.8. Las expresiones 1.36, 1.37 y 1.38 son el fundamento del proceso de muerte incluido en el proceso general de nacimiento y muerte.

$$P[\text{Exactamente 1 salida en } [t, t + \Delta t]] = \mu \Delta t \quad (1.36)$$

$$P[\text{Ninguna salida en } [t, t + \Delta t]] = 1 - \mu \Delta t \quad (1.37)$$

$$P[\text{Mas de 1 salida en } [t, t + \Delta t]] = 0 \quad (1.38)$$

El objetivo es encontrar  $P_n(t)$ , la función de densidad de la variable aleatoria  $N_t$ , es decir el estado del sistema, que denota el número de clientes dentro del sistema en el instante  $t$  cuando la dinámica del sistema obedece a un proceso de nacimiento y muerte. Formalmente, la expresión 1.39 define este concepto.

$$f_{n_t}(n) = P_n(t) \equiv P[N_t = n \text{ en el intervalo } [0, t]] \quad (1.39)$$

Y sea  $p_{ij}(\Delta t)$  la probabilidad de pasar del estado (número de clientes dentro del sistema)  $i$  al estado (número de clientes dentro del sistema)  $j$  en un intervalo de tiempo de  $\Delta t$  segundos.

A partir de las expresiones numeradas desde la 1.33 a la 1.39 se deduce la expresión 1.40

$$P_n(t + \Delta t) = P_n(t) p_{n,n}(\Delta t) + P_{n-1}(t) p_{n-1,n}(\Delta t) + P_{n+1}(t) p_{n+1,n}(\Delta t) \quad (1.40)$$

Términos de orden superior funcionalmente dependientes de  $\Delta t$  son ignorados puesto que son, en magnitud, muy pequeños. La ecuación 1.40 dice que se puede llegar a tener  $n$  clientes en el tiempo  $t + \Delta t$  ya sea porque el estado era  $n - 1$ ,  $n$  ó  $n + 1$  clientes en el tiempo  $t$ . Debido a que se está frente a un proceso de *nacimiento y muerte*, estas son las únicas posibilidades. Para llegar al estado 0, se tendrá la condición inicial y, es evidente que, esa es una situación especial y que se cumple la ecuación 1.41.

$$P_0(t + \Delta t) = P_0(t) p_{0,0}(\Delta t) + P_1(t) p_{1,0}(\Delta t) \quad (1.41)$$

Ahora sustituyendo en la expresión 1.41 las probabilidades  $p_{ij}(\Delta t)$  de acuerdo con las definiciones de nacimientos de las ecuaciones 1.33, 1.34 y 1.35 así como las definiciones para el proceso de muertes dadas en las ecuaciones 1.36, 1.37 y 1.38 para los eventos que ocurren en un intervalo  $[t, t + \Delta t]$ , la expresión 1.40 se convierte en la ecuación

$$P_n(t + \Delta t) = P_n(t)(1 - \Delta t)(1 - \mu \Delta t) + P_{n-1}(t)(\lambda \Delta t) + P_{n+1}(t)(\mu \Delta t) \quad (1.42)$$

Mientras que, bajo esas mismas consideraciones, la expresión 1.41 se transforma en 1.43.

$$P_0(t + \Delta t) = P_0(t)(1 - \lambda\Delta t) + P_1(t)(\mu\Delta t) \quad (1.43)$$

Realizando unas sencillas operaciones algebraicas, reorganizando y tomando el límite cuando  $\Delta t \rightarrow 0$  sobre las ecuaciones 1.42 y 1.43 se llega a la expresión 1.44

$$\frac{dP_n(t)}{dt} = -(\lambda + \mu)P_n(t) + \lambda P_{n-1}(t) + \mu P_{n+1}(t) \quad (1.44)$$

Y a la ecuación 1.45.

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t) \quad (1.45)$$

Para  $n \geq 1$ .

Las ecuaciones diferenciales 1.44 y 1.45, describen la evolución de las probabilidades  $P_n(t)$  de los estados del sistema de líneas de espera ( $M/M/1$ ) a través del tiempo. Pero ¿qué significan realmente estas ecuaciones? La respuesta a esta pregunta necesitará una discusión posterior sobre flujos y balances.

Un caso muy útil e interesante sucede cuando  $\frac{dP_n(t)}{dt} = 0$ ; es decir cuando la estructura de probabilidades no cambia en función del tiempo. Esta es la situación que ocurre cuando el sistema de filas ha estado operando por mucho tiempo, el comportamiento se ha estabilizado transitoriamente y en la que el sistema ha llegado a un punto de equilibrio. Por ese motivo se dice que el sistema está en *equilibrio* o en *estado estable*.

En este caso, al ser la derivada igual a cero, las probabilidades asociadas con el estado del sistema no cambiarán (serán constantes) y, en consecuencia, no dependerán del tiempo  $t$ . Por esa razón no es necesario escribirlas como funcione  $P_n(t)$  sino como constantes  $p_n$ . De esta manera, las ecuaciones diferenciales 1.44 y 1.45 se convierten en las expresiones 1.46 y 1.47 respectivamente.

$$0 = -(\lambda + \mu)p_n + \lambda p_{n-1} + \mu p_{n+1} \quad (1.46)$$

$$0 = -\lambda p_0 + \mu p_1 \quad (1.47)$$

Para  $n \geq 1$ .

Las ecuaciones resultantes son un conjunto de ecuaciones lineales en lugar

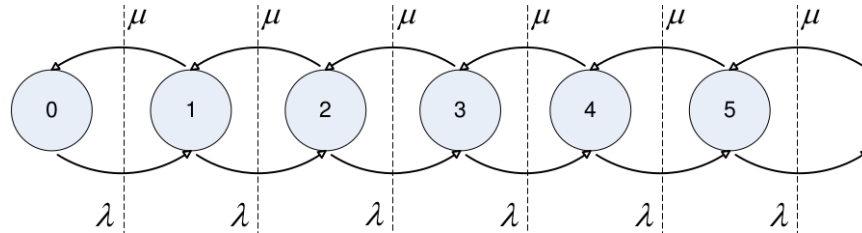


de un conjunto de ecuaciones diferenciales, lo cual simplifica su solución. Para calcular las probabilidades de equilibrio,  $p_n$ ,  $n = 0, 1, 2, \dots, N$ , que gobiernan los estados de cualquier sistema de filas con un número  $N+1$  finito de estados, se procede de la siguiente manera. Para cada uno de los  $N+1$  estados se escribe una ecuación global de balance. Esto da un sistema de  $N+1$  ecuaciones lineales con  $N+1$  incógnitas. Ninguna de estas ecuaciones es redundante y, adicionalmente, debe ser incluida la ecuación 1.48 para normalizar la solución y, al mismo tiempo, garantizar que el conjunto de todas las probabilidades encontradas constituyan una densidad de probabilidad.

$$p_0 + p_1 + p_2 + \dots + p_N = 1 \quad (1.48)$$

La solución del conjunto de ecuaciones lineales puede ser solucionado utilizando técnicas numéricas estándar. La razón por la cual es indispensable obtener todas las probabilidades a partir de ellas (y no antes) es posible determinar las medidas de desempeño, descritas en la sección 1.2.2, para cualquier sistema de líneas de espera.

La dificultad que tiene esta aproximación es que en los modelos de colas reales frecuentemente el número de posibles estados del sistema es extremadamente grande haciendo la solución computacional casi imposible. Por suerte, en esos casos se puede suponer un sistema con capacidad infinita (en la práctica número de estados muy grande) y aplicar para esa situación un resultado matemático de extrema utilidad y belleza. Supóngase que se redibuja el diagrama de Transición de Estados, *DTE*, del sistema  $(M/M/1)$  como en la Figura 1.6.



**Figura 1.6**

Balance local en la frontera entre estados

En ese *DTE* se marcan con líneas punteadas verticales las fronteras entre cada par adyacente de estados. Cada frontera separa el diagrama de estado de transición en dos mitades. Es posible equiparar el flujo de izquierda a derecha a través de tal frontera con el flujo de derecha a izquierda. Así se obtiene un conjunto de *ecuaciones locales de balance*. De cada una de esas fronteras imaginarias surge una ecuación cuyo patrón se presenta en la expresión 1.49.

$$\lambda p_{n-1} = \mu p_n \quad (1.49)$$

Para  $n = 1, 2, 3, \dots$  (hasta infinito puesto que se empezó suponiendo un sistema con infinitos estados). Y, de ahí, fácilmente se deduce la ecuación 1.50.

$$p_n = \frac{\lambda}{\mu} p_{n-1} \quad (1.50)$$

Ahora, a través de un sencillo procedimiento recursivo, que se detiene para  $n = 1$ , es fácil verificar que la solución de la ecuación en diferencias 1.50 es 1.51.

$$p_n \left( \frac{\lambda}{\mu} \right)^n p_0 \quad (1.51)$$

Finalmente, puesto que  $\sum_{n=0}^{+\infty} p_n = \sum_{n=0}^{+\infty} \left( \frac{\lambda}{\mu} \right)^n p_0 = 1$ , por ser una función de densidad de probabilidad discreta, se obtiene la probabilidad de que el sistema esté desocupado 1.52.

$$p_0 = \frac{1}{\sum_{n=0}^{+\infty} \left( \frac{\lambda}{\mu} \right)^n} \quad (1.52)$$

---

## 1.4. EL SISTEMA DE FILAS (M/M/1) EN DETALLE

### 1.4.1. Función de densidad

Es posible simplificar la expresión 1.52 para calcular  $p_0$  al emplear la “serie geométrica” dada en la identidad 1.53 con  $\rho = \frac{\lambda}{\mu}$ . Para que esa identidad pueda ser empleada es necesario imponer la restricción  $\lambda < \mu$ .

$$\sum_{n=0}^{\infty} \rho^n = \frac{1}{1-\rho} \text{ con } 0 \leq \rho < 1 \quad (1.53)$$

De esta manera, entonces,  $p_0 = (1 - \rho)$  y, en consecuencia, la función de densidad sobre los estados del sistema se obtiene a partir de 1.51 y se presenta en la ecuación 1.54.

$$p_n = (1 - \rho) \rho^n \quad (1.54)$$

#### 1.4.2. Utilización del servidor

La probabilidad de que el servidor esté ocupado (es decir que no esté libre) se denomina “la utilización del servidor”. Esta probabilidad se interpreta como el porcentaje del tiempo de operación del sistema en el cual el servidor estuvo ocupado atendiendo a los clientes del sistema. Formalmente esta medida de desempeño se calcula como se muestra en la ecuación 1.55.

$$U = P[N_t > 0] = 1 - P[N_t = 0] = 1 - p_0 = \rho \quad (1.55)$$

En consecuencia el parámetro  $\rho = \frac{\lambda}{\mu}$  se conoce como *utilización*. La magnitud  $\rho$  puede verse como la **carga ofrecida** normalizada. Esto es,  $\lambda$  puede variar desde cero (carga suave), hasta un valor ligeramente menor que  $\mu$  (carga pesada). Además,  $\rho$  puede variar entre 0 y 1.

Aunque para llegar a esta conclusión se supuso que la capacidad del sistema es infinito, en la práctica, sin embargo, las probabilidades de que en el sistema hayan muchos clientes esperando a ser atendidos es relativamente pequeña y depende, obviamente, de la “carga ofrecida” por el sistema.

#### 1.4.3. Número medio de clientes en el sistema

El número medio  $\mu_{N_t}$  de clientes que se encuentran en el sistema está dado en la ecuación 1.56.

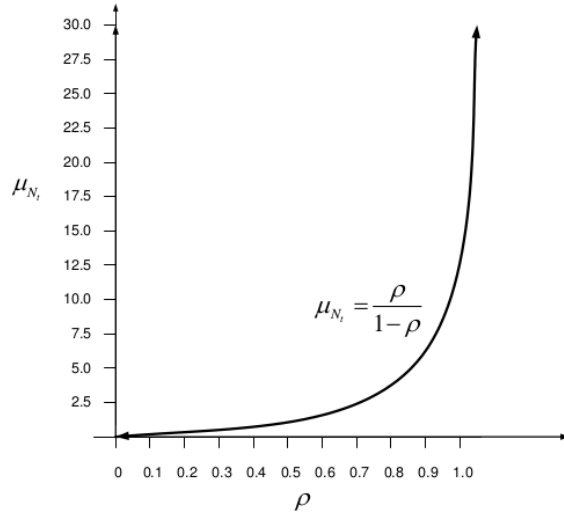
$$\mu_{N_t} = E[N_t] = \frac{\rho}{1 - \rho} \quad (1.56)$$

Para mostrar que 1.56 es cierta, considérense los dos comentarios que siguen:

1.  $\sum_{n=0}^{\infty} n \rho^n = \rho \sum_{n=1}^{\infty} n \rho^{n-1} = \rho \frac{d}{d\rho} \left( \frac{1}{1 - \rho} \right) = \frac{\rho}{(1 - \rho)^2}$ . Es decir que  $\sum_{n=0}^{\infty} n \rho^n = \frac{\rho}{(1 - \rho)^2}$ . Este resultado se empleará en uno de los pasos del numeral 2 que sigue.

$$2. \mu_{N_t} = E[N_t] = \sum_{n=0}^{\infty} n \rho_n = \sum_{n=0}^{\infty} n (1-\rho) \rho^n = (1-\rho) \sum_{n=0}^{\infty} n \rho^n = (1-\rho) \left[ \frac{\rho}{(1-\rho)^2} \right] = \frac{\rho}{(1-\rho)}.$$

La función descrita por la ecuación (1.56) se presenta en la Figura 1.7. Lo que es más notable de esta función es un incremento no lineal asintótico en el número medio de clientes en el sistema conforme  $\rho \rightarrow 1$ . También es interesante comentar que esta función se utiliza con frecuencia como una “función de penalización” en optimización de algoritmos en filas largas de espera.



**Figura 1.7**

Número medio de  $\mu_{N_t}$  clientes en un sistema (M/M/1)

#### 1.4.4. Varianza del Número de clientes en el sistema.

LA varianza  $\sigma_{N_t}^2$  del número de clientes que se encuentran en el sistema está dada en la ecuación 1.57.

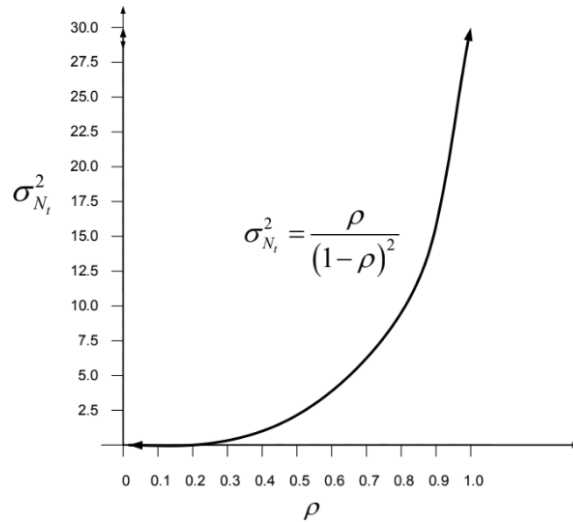
$$\sigma_{N_t}^2 = E[(N_t - \mu_{N_t})^2] = \frac{\rho}{(1-\rho)^2} \quad (1.57)$$

Para mostrar que 1.56 es cierta, considérense los dos comentarios que siguen:

1.  $\sigma_{N_t}^2 = E[(N_t - \mu_{N_t})^2] = E[N_t^2] - (\mu_{N_t})^2$ . Nótese que  $\mu_{N_t}$  está dada en la expresión 1.56 y, por lo tanto, es suficiente con calcular el segundo momento del número de clientes en el sistema  $N_t$

2. Antes de ello es fácil verificar que  $\sum_{n=0}^{\infty} n^2 \rho^n = \frac{\rho(1+\rho)}{(1-\rho)^3}$ . Este resultado se empleará en uno de los pasos del numeral 3 que sigue.
3.  $E[N_t^2] = \sum_{n=0}^{\infty} n^2 \rho_n = \sum_{n=0}^{\infty} n^2 (1-\rho) \rho^n = (1-\rho) \frac{\rho(1+\rho)}{(1-\rho)^3} = \frac{\rho(1+\rho)}{(1-\rho)^2}$ . Finalmente, reemplazando este resultado en el comentario 1, se sigue el resultado.

El comportamiento de la variabilidad del número de clientes en el sistema se presenta en la Figura 1.8. Al igual que con el número



**Figura 1.8**

Varianza del número de clientes  $\sigma_{N_t}^2$  en un sistema (M/M/1)

## 1.5. LEY DE LITTLE

La ley de Little es el resultado teórico que permite calcular otra medida de desempeño importante en sistemas de línea de espera: *El Tiempo medio que un cliente permanece dentro del sistema*. Formalmente, sea  $T$  la variable aleatoria que representa el tiempo que un cliente “gasta” desde el momento de su llegada al sistema hasta el momento en el cual el servidor termina de atenderlo. De acuerdo con lo discutido en la sección 1.2.2, entonces, se desea

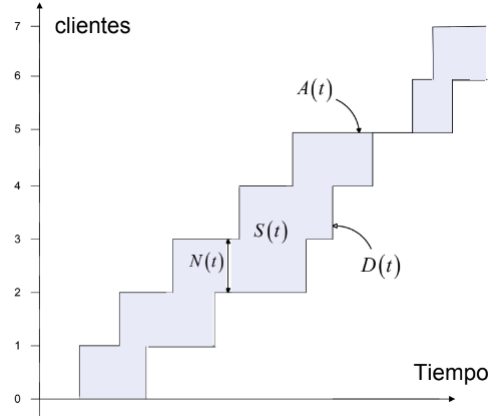
encontrar  $\mu_T = E[T]$ . Una forma de encontrar ésta constante es realizar el cálculo aplicando directamente la definición de esperanza matemática; sin embargo, para ello es necesario contar con  $f_T(t)$ , información con la cual no se cuenta.

Una forma alternativa, más directa, simple, eficiente y elegante, es emplear la ecuación llamada la “Ley de Little”<sup>1</sup>

$$\mu_N = \lambda \mu_T \quad (1.58)$$

Nótese que  $\lambda$  es la tasa de llegadas de clientes al sistema de líneas de espera. Así mismo, se cuenta  $\mu_{N_t}$  al aplicar la ecuación 1.56, en consecuencia,  $\mu_T$  se puede deducir fácilmente a partir de 1.58.

En [1] se presentan al menos cinco demostraciones diferentes de este interesante y útil resultado<sup>2</sup>. El esquema de demostración que aquí se presenta está inspirado en una de esas pruebas. La Figura 1.9 ayuda a entender ésta estrategia de demostración.



**Figura 1.9**

Esquema de demostración de la “Ley de Little”

Las bases de la prueba se enumeran a continuación:

1. Considerando el intervalo de tiempo  $[0, t]$  y sean las funciones:  $A(t)$  que denota el número de llegadas al sistema hasta el instante  $t$ .  $D(t)$  que representa el número de salidas del sistema hasta el instante  $t$ . Por esta razón  $N(t) = A(t) - D(t)$  será el número de clientes en el sistema en el instante  $t$ .

<sup>1</sup>La notación original empleada por Little para esta ley es  $L = \lambda W$ .

<sup>2</sup>La ley de Little también se tiene si consideramos la fila y no el servidor, esto es  $\mu_N^{(Q)} = \lambda \mu_T^{(Q)}$ .

2.  $S(t)$  es el área acumulada entre las curvas  $A(t)$  y  $D(t)$  hasta el tiempo  $t$ . Esta área es el tiempo total utilizado por todos los clientes que han estado en el sistema hasta el instante  $t$ .
3. Con ello, ahora, se formalizan tres relaciones evidentes en la Figura 1.9.
  - a) La primera es  $\lambda_t = \frac{A(t)}{t}$ . Aquí,  $\lambda_t$  es la tasa media de llegadas durante el  $[0, t]$ .
  - b) La segunda es  $\mu_T^{(t)} = \frac{S(t)}{A(t)}$  el tiempo medio que han empleado los cliente en el hasta el instante  $t$ .
  - c) Y, por último, la tercera ecuación es  $\mu_N^{(t)} = \frac{S(t)}{t}$  el número promedio de clientes hasta el instante  $t$ .
4. De ahí, es claro que  $\mu_N^{(t)} = \frac{S(t)}{t} = \frac{\mu_T^{(t)} \times A(t)}{t} = \lambda_t \times \mu_T^{(t)}$ . Es decir,  $\mu_N^{(t)} = \lambda_t \times \mu_N^{(t)}$ .
5. Finalmente, tomando el límite cuando  $t \rightarrow \infty$ ,  $\mu_N = \lim_{t \rightarrow \infty} (\mu_N^{(t)})$ ,  $\lambda \lim_{t \rightarrow \infty} (\lambda_t)$  y  $\mu_T = \lim_{t \rightarrow \infty} (\mu_T^{(t)})$ , De donde se concluye la ley  $\mu_N = \lambda \mu_T$ .

La ley de Little aplica, en general, para cualquier sistema de líneas de esperas ( $G/G/m$ ) y en particular para los sistemas ( $M/M/1$ ), como se muestra en el ejemplo que sigue.

**Ejemplo 1—3** El cálculo de  $\mu_T$  para el sistema ( $M/M/1$ ) a partir de la “Ley de Little” mostrada en la ecuación 1.58, puede realizarse usando el resultado de la ecuación 1.56, por lo tanto

---

## 1.6. TEOREMA DE BURKE

**Teorema 1.1 (Teorema de Burke)** *El proceso de salida de un proceso para un sistema ( $M/M/1$ ), en equilibrio, es un proceso de Poisson con parámetro  $\lambda$ .*

---

## 1.7. RESUMEN Y PANORAMA HISTÓRICO

En este capítulo se estudió la teoría de colas simple, una de las ramas de la investigación de operaciones. El estudio matemático de las filas que se forman

en ciertos sistemas es fundamental para optimizar la operación del sistema. Por esta razón, el principal propósito de esta disciplina es caracterizar dicho sistema desde el punto de vista de su rendimiento. Medidas de desempeño tales como el tiempo de espera medio en las colas o la utilización del sistema para que éste mantenga su normal operación sin que llegue a colapsar o, al menos, conocer las circunstancias que producirían tal situación.

Su aplicabilidad a las ciencias de la computación, las redes de computadores y las telecomunicaciones es evidente. En este sentido, la teoría es muy útil para modelar procesos tales como la llegada de datos a una cola en ciencias de la computación, la congestión de redes de computadores, redes Ad Hoc, sistemas celulares y, en general, redes de telecomunicaciones. Otras aplicaciones en el contexto de la computación son también evidentes. Por ejemplo, los procesos enviados a un servidor para su ejecución forman colas de espera mientras no son atendidos; la información solicitada, a través de Internet, a un servidor Web puede recibirse con demora debido a la congestión en la red.

El origen de esta teoría se encuentra en los estudios del matemático, estadístico, e ingeniero danés Agner Krarup Erlang quien nació en el año de 1878 y murió en 1929. A él se le da el apelativo de “Padre de la Ingeniería de Teletráfico” puesto que en 1909, para analizar la congestión del tráfico telefónico con el propósito de satisfacer adecuadamente la demanda estocástica de servicios en el sistema telefónico de Copenhague, empezó por formalizar o matematizar el comportamiento de estos sistemas.

---

## 1.8. EJERCICIOS

1. Dado el sistema de colas determinístico, tanto en su proceso de llegadas como en el de servicio,  $(\lambda/\mu/1)$  con tasas de llegadas y de atención  $\lambda, \mu \in \mathbb{R}^+$  respectivamente cuyas unidades son *paquetes/segundo*. Deduzca formalmente las siguientes medidas de desempeño (Imponga las restricciones y proponga los supuestos que sean necesarios):
  - a) Utilización.
  - b) Número medio de clientes en fila.
  - c) Número medio de clientes en el sistema.
  - d) Tiempo medio en fila.
  - e) Tiempo medio en el sistema.
2. Para un proceso binomial con  $\lambda = 5 \text{ clientes/Segt} = 10 \text{ seg}$  y  $\delta t = 0,1$ , encuentre:
  - a) La función de densidad.



- b)* La función percentil.
  - c)* La varianza.
  - d)* La función generadora de momentos.
  - e)* La probabilidad de que después de los primeros 5seg no lleguen clientes.
- 3. En la sección 1.3.3 se dedujo el modelo de nacimiento puro. Haciendo un desarrollo similar, formule los fundamentos de un proceso de muerte pura.
- 4. \* Demuestre que, en el caso continuo, la familia exponencial  $T \sim Exp(\lambda)$  es la ÚNICA familia que exhibe la propiedad de pérdida de la memoria.
- 5. Con un procedimiento análogo (esto es similar) al empleado en la sección 1.3.6 para demostrar que la familia exponencial presenta la propiedad de pérdida de la memoria,
  - a)* Demuestre que, en el caso discreto, la familia Geométrica  $T \sim Geo(p)$  presenta la propiedad de pérdida de la memoria.
  - b)* \* Demuestre que, en el caso discreto, la familia Geométrica  $T \sim Geo(p)$  es la ÚNICA familia que exhibe la propiedad de pérdida de la memoria.

---

## ***Bibliografía***

---

- [1] L. Kleinrock. *Queueing Systems. Volume 1: Theory*. Prentice-Hall, Enero de 1975.