INTRODUCCION A LA TEORIA DE COLAS Y SU SIMULACION

Gerardo Fabian Peraza Siqueiros

Licenciatura en Matemáticas

Director de Tesis:

Dr. Jesús Adolfo Minjárez Sosa

División de Ciencias Exactas y Naturales
Departamento de Matemáticas, Universidad de Sonora.

Hermosillo, Sonora, México

Febrero de 2013

Índice general

Dedicatoria

Con especial dedicatoria a todas las personas que han tenido que ver en mi carrera y en mi formación. Toda esa gente que me ha impulsado con su ejemplo a continuar hasta concluir esta bonita carrera.

Agradecimientos

Primero que nada a mi familia, que siempre me ha apoyado en mis proyectos, pero sobre todo que me inculcó el valor de aprender cosas nuevas y de mejorar mi persona. En segundo lugar gracias a mis maestros, que resultaron ser excelentes en su trabajo, siempre muy comprometidos con nuestra educación. Por último, pero no menos importantes, gracias a mis compañeros, ya que recorrimos todo el camino juntos, y a lo largo de este, siempre nos hemos apoyado unos con otros.

Introducción

El objetivo de la teoría de colas es modelar sistemas de espera tales que funcionan de la siguiente manera. Existe un medio al que llegan clientes demandando cierto servicio. Luego, a consecuencia de que la demanda no puede ser satisfecha inmediatamente, se forma una cola (o linea de espera) de clientes en espera de ser atendidos por el o los servidores correspondientes. Los tiempos entre arribo de clientes consecutivos al sistema y los tiempos de servicio son aleatorios, y son representados por variables aleatorias con alguna distribución de probabilidad. En particular, en este trabajo estudiaremos colas en las que ambas variables aleatorias tienen distribución exponencial. A este tipo de colas se les conoce como colas poissonianas debido a la relación entre las distribuciones de Poisson y exponencial.

El término "clientes" es general. Dependiendo del sistema a estudiar, los clientes pueden ser: llamadas telefónicas esperando ser procesadas por un conmutador, e-mails esperando entrar a un servidor, personas en un banco esperando ser atendidas, etc.

La teoría de colas, como disciplina matemática, inició con el trabajo de A. Erlang [3] quien estudió un modelo de una estación telefónica obteniendo una fórmula para la distribución del número de lineas ocupadas. A partir de este trabajo, la teoría ha sido aplicada en el estudio de un gran número de sistemas de espera como tráfico de aviones, redes eléctricas, sistemas de internet, teoría de inventarios, entre muchos otros. Sin embargo, es importante mencionar que mucha de la teoría de hoy en día fue desarrollada sin aplicación práctica sino únicamente por puro interés matemático.

Uno de los objetivos de este trabajo es presentar una introducción a los métodos matemáticos que han sido usados para modelar un sistema de espera, específicamente para colas poissonianas; así como también, presentar simulaciones por computadora con el fin de ejemplificar la teoría desarrollada.

De hecho, la tesis está estructurada de la siguiente manera.

En el Capítulo I introducimos la distribución de Poisson y su relación con la distribución exponencial para después definir el proceso de Poisson. En esta parte se presentan los elementos necesarios para describir tanto los tiempos entre arribos de los clientes como los tiempos de servicio.

Las colas poissonianas tienen la característica de que se pueden modelar mediante procesos de Markov, en particular como procesos de nacimiento y muerte. Por lo anterior, en el Capítulo II presentamos la teoría básica de procesos de Markov necesaria para el estudio de los sistemas de espera.

Finalmente, en el Capítulo III, presentamos la descripción de las colas poissonianas aplicando la teoría desarrollada en los capítulos previos. Cabe mencionar que estudiamos colas con uno y más servidores, y presentamos simulaciones las cuales se desarrollaron utilizando el lenguaje C++.

Capítulo 1

Proceso de Poisson

1.1. Introducción

El proceso de Poisson es un proceso estocástico a tiempo continuo que modela el número de veces que ocurre un evento específico a través del tiempo. Por ejemplo, los clientes que llegan a un supermercado, las llamadas que entran a un conmutador, las fallas registradas en un circuito eléctrico, entre otros.

Una característica de este proceso es que los tiempos entre la ocurrencia de eventos consecutivos tiene una distribución exponencial. En este capítulo introducimos el proceso de Poisson, así como sus propiedades más importantes, con el cual modelaremos tanto los tiempos entre arribos de los clientes como los tiempos de servicio en los sistemas de espera.

1.2. Distribuciones de Poisson y Exponencial

Definición 1.2.1 *Sea λ un número real positivo. Decimos que una variable aleatoria (v.a.) X tiene distribución de Poisson con parámetro λ ($X \sim Poisson(\lambda)$) si su función de probabilidad está dada por:*

$$f_X(x) := P\left[X = x\right] = \begin{cases} \dfrac{\lambda^x}{x!}\exp(-\lambda) & si \quad x = 0, 1, ... \\ \\ 0 & otro\ caso \end{cases}$$

Observación 1.2.1 *a) Si $N(t) \sim Poisson(\lambda t)$ entonces*

$$E(N(t)) = Var(N(t)) = \lambda t.$$

b) Una manera más práctica para fines computacionales de calcular las probabilidades de la distribución de Poisson es por medio de la siguiente ecuación recursiva. Sea

$$P_n(t) = P\left[N(t) = n\right] = \frac{(\lambda t)^n}{n!}\exp(-\lambda t).$$

Entonces

$$\begin{aligned} P_0(t) &= \exp(-\lambda t); \\ P_n(t) &= \frac{\lambda t}{n} P_{n-1}(t), \quad n = 1, 2, ... \end{aligned}$$

Definición 1.2.2 *Sea λ un número real positivo. Decimos que una v.a. X tiene distribución exponencial con parámetro λ ($X \sim \exp(\lambda)$) si su función de distribución está dada por:*

$$F_X(x) := P\left[X \le x\right] = \begin{cases} 1 - \exp(-\lambda x) & si \quad x \ge 0 \\ \\ 0 & \quad\quad x < 0 \end{cases}$$

La distribución exponencial también puede ser descrita por medio de su función de densidad

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & si \quad x \ge 0 \\ \\ 0 & \quad\quad x < 0 \end{cases}$$

Además, la distribución exponencial se obtiene al modelar el tiempo que pasa entre dos eventos que ocurren en forma aleatoria, con lo cual se relaciona con la distribución de Poisson. En efecto, si $N(t) \sim Poisson(\lambda t)$ representa el número de eventos que ocurren en $[0, t]$, y X es la v.a. que representa el tiempo que transcurre hasta que ocurre un evento, entonces los eventos $[X > t]$ y $[N(t) = 0]$ son equivalentes. Por lo tanto,

$$P\left[X \le t\right] = 1 - P\left[X > t\right] = 1 - P\left[N(t) = 0\right] = 1 - \exp(-\lambda t), \ t \ge 0, \quad (1.1)$$

lo cual representa la distribución exponencial con parámetro λ. Este punto lo discutiremos de nuevo en la Sección 1.4.

A continuación enunciamos algunas propiedades bien conocidas de la distribución exponencial las cuales incluimos para una fácil referencia.

Proposición 1.2.1 *a) Si* $X \sim \exp(\lambda)$ *entonces*

$$E(X) = \frac{1}{\lambda} \quad y \quad Var(X) = \frac{1}{\lambda^2}.$$

b) (Propiedad de pérdida de memoria) Dentro de la clase de variables aleatorias absolutamente continuas, $X \sim \exp(\lambda)$ *si y solo si para todos los números positivos* s *y* t *se tiene*

$$P\left[X > s + t | X > s\right] = P\left[X > t\right].$$

Definición 1.2.3 *Sea* λ *un número real positivo y* $n \in \mathbb{N}$*. Decimos que una v.a.* X *tiene distribución Gama con parámetros* λ *y* n $(X \sim \Gamma(n, \lambda))$ *si su función de densidad está dada por:*

$$f_X(x) = \begin{cases} \frac{\lambda^n x^{n-1} \exp(-\lambda x)}{(n-1)!} & si \quad x \geq 0 \\ \\ 0 & \quad\;\; x < 0 \end{cases}$$

Al igual que la distribución exponencial, la distribución Gama se obtiene al estudiar el tiempo que transcurre en cierto número de ocurrencias de eventos que suceden aleatoriamente en el tiempo. De hecho, observe que la distribución exponencial es un caso particular de la distribución gama tomando $n = 1$. Más aún, la distribución Gama tiene la siguiente propiedad.

Proposición 1.2.2 *Si* $X_1, X_2, ..., X_n$ *son v.a.i.i.d, tal que* $X_i \sim \exp(\lambda)$, $i = 1, 2, ..., n$, *entonces*

$$X_1 + X_2 + ... + X_n \sim \Gamma(n, \lambda).$$

Concluimos esta sección presentando la siguiente notación, la cual será utilizada para establecer algunas propiedades del Proceso de Poisson.

Notación. La expresión "$f = o(h)$ cuando $h \to 0$" significa que $f(\cdot)$ es una función tal que

$$\lim_{h \to 0} \frac{f(h)}{h} = 0,$$

es decir, para cada $\varepsilon > 0$ existe $\delta > 0$ tal que

$$|h| < \delta \quad \text{implica} \quad \left|\frac{f(h)}{h}\right| < \varepsilon$$

Ejemplo 1.2.1 *Sea $f(x) = x^r, r > 1$. Observemos que*

$$\frac{f(h)}{h} = \frac{h^r}{h} = h^{r-1} \to 0 \quad \textit{cuando } h \to 0,$$

lo cual implica que $f = o(h)$ cuando $h \to 0$.

Ejemplo 1.2.2 *Si* $f(x) = x^r, r < 1$, *entonces*

$$\frac{f(h)}{h} = \frac{h^r}{h} = h^{r-1} \nrightarrow 0 \quad \text{cuando } h \to 0.$$

Es decir, $f \neq o(h)$ cuando $h \to 0$.

Ejemplo 1.2.3 *La función $f(x) = e^{-x}$, satisface $f(h) = 1 - h + o(h)$. Esto se deduce expandiendo f en su serie de Taylor:*

$$f(x) = \sum_{n=0}^{\infty} \frac{(-x)^n}{n!} = 1 - x + \sum_{n=2}^{\infty} \frac{(-x)^n}{n!} = 1 - x + R_2(x),$$

donde $R_2 = o(h)$ cuando $h \to 0$. En efecto,

$$\frac{R_2(h)}{h} = \sum_{n=2}^{\infty} \frac{(-1)^n (h)^n}{(h)n!} = \sum_{n=2}^{\infty} \frac{(-1)^n}{n!} (h)^{n-1} \to 0 \quad \textit{cuando} \quad h \to 0;$$

entonces, $\quad f(h) = 1 - h + o(h)$.

Observación 1.2.2 *a) Si* $f = o(h)$ *y* $g = o(h)$*, entonces* $f + g = o(h)$ *y* $cf = o(h)$ *para cualquier constante* c*. En general,* $\sum_{i=1}^{n} c_i f_i = o(h)$*, si* $f_i = o(h)$ *para cada* $i = 1, 2, ...n$

b) Del Ejemplo 1.2.3, podemos ver que si $f(x) = e^{-ax}$*, entonces* $f(h) = 1 - ah + o(h)$*.*

c) Sea $X \sim exp(\lambda)$*, entonces por la observación anterior,*

$$\begin{aligned} P[X \leq h] &= 1 - e^{-\lambda h} \\ &= 1 - [1 - \lambda h + o(h)] \\ &= \lambda h - o(h) \\ &= \lambda h + o(h), \end{aligned}$$

de donde se sigue que

$$P[X > h] = e^{-\lambda h} = 1 - \lambda h + o(h).$$

1.3. Procesos estocásticos

Definición 1.3.1 *Un proceso estocástico es una familia* $\mathbf{X} = \{X(t); t \in T\}$ *de v.a., donde* T *es el conjunto de parámetros.*

Observación 1.3.1 *a) Sea* $\mathbf{X} = \{X(t); t \in T\}$ *un proceso estocástico, entonces para cada* $t \in T$*,* $X(t)$ *es una v.a. definida en algún espacio de probabilidad* $(\Omega, \mathcal{F}, P)$ *es decir,*

$$X(t) : \Omega \to \mathbb{R}.$$

b) El hecho de que $X(t)$ *sea una v.a. implica que el proceso estocástico se puede ver como una colección de funciones reales definidas sobre* $T \times \Omega$*,*

$$X(\cdot, \cdot) : T \times \Omega \to \mathbb{R}$$

que satisfacen:

(i) Para cada $t \in T$ *fijo,* $X(t, \cdot) : \Omega \to \mathbb{R}$ *es una v.a.;*

(ii) Para cada $\omega \in \Omega$ *fijo,* $X(\cdot, \omega) : T \to \mathbb{R}$ *es una función de valores reales definida en* T.
c) A la función $X(\cdot, \omega) : T \to \mathbb{R}$, $\omega \in \Omega$, *la llamaremos realización o trayectoria del proceso.*

Definición 1.3.2 *El espacio de estados del proceso* $\mathbf{X}$ *es el conjunto* S *de todos los valores posibles de las v.a.* $X(t)$ *para* $t \in T$. *Además, si* S *es finito o numerable decimos que el proceso* $\mathbf{X}$ *es discreto, y si* S *es un intervalo, decimos que el proceso* $\mathbf{X}$ *es continuo. Por otro lado, si* T *es numerable, digamos* $T = \{0, 1, 2, ...\}$, *decimos que* $\mathbf{X}$ *es un proceso con parámetro o tiempo discreto. Si* T *es un intervalo,* $\mathbf{X}$ *es un proceso con parámetro o tiempo continuo.*

Observación 1.3.2 *Si* $\mathbf{X} = \{X(t),\ t \in T\}$ *es un proceso a tiempo discreto, por ejemplo,* $T = \{0, 1, 2, ...\}$, *escribimos:*

$$X(t) = X_t,$$

y por lo tanto el proceso estocástico toma la forma $\mathbf{X} = \{X_0, X_1, X_{2,...}\}$.

Ejemplo 1.3.1 *Al final de cada día se registra el número de llamadas que entraron a un conmutador durante todo el día. Definamos*

$$X_t = \textit{Número de llamadas que entraron durante el } t\textit{-ésimo día.}$$

Entonces $\mathbf{X} = \{X_0, X_1, X_{2,...}\}$ *es un proceso estocástico a tiempo discreto con espacio de estados discreto.*

Ejemplo 1.3.2 *Para cada* $t \geq 0$, $V(t)$ *es el volumen de agua que hay en una presa al tiempo t. Entonces* $V = \{V(t), t \geq 0\}$ *es un proceso a tiempo continuo y con espacio de estados continuo.*

La siguiente definición proporciona otra clase de ejemplos de procesos estocásticos a tiempo continuo con espacio de estado discreto.

Definición 1.3.3 *Un proceso de conteo es un proceso* $\mathbf{X} = \{N(t), t \geq 0\}$ *con espacio de estados* $S = \{0, 1, 2, ...\}$ *tal que* $N(0) = 0$ *y para cada* $t \geq 0$, $N(t)$ *representa el número de veces que ocurre cierto evento* E *durante un intervalo.*

En otras palabras, $\mathbf{X} = \{N(t), t \geq 0\}$ es un proceso de conteo si satisface:
(i) $N(t)$ solo toma valores enteros no negativos;
(ii) $N(t)$ es no decreciente : $N(s) \leq N(t)$ si $s < t$;
(iii) $N(t) - N(s)$ es el número de veces que ha ocurrido E, durante el intervalo $(s, t]$.

Observación 1.3.3 *En un proceso de conteo, la v.a.* $N(t)$ *puede representar, por ejemplo, el número de accidentes que ocurren hasta el tiempo* t*, o el número de clientes que llegan a un banco hasta el tiempo* t*, etc.*

Definición 1.3.4 *Sea* $\mathbf{X} = \{X(t), t \geq 0\}$ *un proceso estocástico a tiempo continuo. Decimos que:*
a) $\mathbf{X}$ *es un proceso con incrementos independientes si para cualquier selección de índices* $0 \leq t_0 \leq t_1 \leq ... \leq t_n$*, las variables aleatorias* $X(t_1) - X(t_0), X(t_2) - X(t_1), ..., X(t_n) - X(t_{n-1})$ *son independientes.*

b) $\mathbf{X}$ *es un proceso con incrementos estacionarios si las variables aleatorias* $X(t + h) - X(t)$ *y* $X(\tau + h) - X(\tau)$ *tienen la misma distribución para todo* t, $\tau \geq 0$ *y* $h > 0$.

1.4. Proceso de Poisson

Una clase particular en los procesos de conteo son los procesos de Poisson definidos de la siguiente manera.

Definición 1.4.1 *Un proceso de conteo* $\mathbf{X} = \{N(t), t \geq 0\}$ *es un proceso de Poisson con intensidad media* $\lambda > 0$ *si:*
(i) $N(0) = 0$
(ii) $\mathbf{X}$ *tiene incrementos independientes;*
(iii) $\mathbf{X}$ *tiene incrementos estacionarios tales que*

$$N(t + s) - N(s) \sim Poisson(\lambda t), \text{ si } t \geq 0, s \geq 0.$$

Observación 1.4.1 *a) La propiedad (iii) significa que el número de eventos en cualquier intervalo de longitud t tiene distribución de Poisson con parámetro λt. Es decir:*

$$P\left[N(t+s)-N(s)=n\right]=\frac{(\lambda t)^n e^{-\lambda t}}{n!},\ n=0,1,2...$$

En particular $N(t)\sim Poisson(\lambda t)$.

b) Además tenemos que (ver Observación 1.2.1),

$$E\left[N(t+s)-N(s)\right]=\lambda t \tag{1.2}$$

y

$$Var\left[N(t+s)-N(s)\right]=\lambda t. \tag{1.3}$$

c) De (1.2)

$$\begin{aligned}\lambda &= \frac{E\left[N(t+s)-N(s)\right]}{t}\\ &= \frac{\text{eventos que ocurren en promedio en } (s,t+s]}{t}.\end{aligned}$$

De aquí, podemos interpretar a λ como el número promedio de veces que ocurre un evento particular por unidad de tiempo. Esto justifica el nombre de intensidad media.

1.4.1. Una caracterización del proceso de Poisson

Para mostrar que un proceso de conteo es de Poisson debemos verificar que se cumplen las condiciones de la Definición 1.4.1. Las condiciones (i) y (ii) se pueden derivar de las características del proceso. Sin embargo, la condición (iii) puede resultar difícil de verificar. En esta parte introduciremos una caracterización del proceso de Poisson en la cual la distribución de Poisson no interviene directamente.

Teorema 1.4.1 *El proceso estocástico $\mathbf{X}=\{N(t),t\geq 0\}$ es un proceso de Poisson con intensidad $\lambda>0$ si, y sólo si $\mathbf{X}$ es un proceso de conteo tal que:*

(i) $\mathbf{X}$ tiene incrementos independientes y estacionarios;

(ii) $P\left[N(h)=1\right]=\lambda h+o(h)$ *cuando* $h\to 0^+$;

(iii) $P\left[N(h)\geq 2\right]=o(h)$ *cuando* $h\to 0^+$.

Demostración. Sea $\mathbf{X} = \{N(t), t \geq 0\}$ un proceso de Poisson con intensidad $\lambda > 0$. Entonces $\mathbf{X}$ es un proceso de conteo y por definición satisface la condición (i). Ahora mostraremos que satisface (ii). Por definición tenemos que $N(h) \sim$ Poisson(λh). De aquí tenemos que

$$\begin{aligned} P\left[N(h)=1\right] &= (\lambda h)e^{-\lambda h} = \lambda h(1-\lambda h + o(h)) \\ &= \lambda h - \lambda^2 h^2 + o(h) \\ &= \lambda h + o(h), \end{aligned}$$

Por otra parte note que:

$$\begin{aligned} P\left[N(h) \geq 2\right] &= 1 - P\left[N(h) \leq 1\right] = 1 - \left(P\left[N(h)=0\right] + P\left[N(h)=1\right]\right) \\ &= 1 - \left[\frac{(\lambda h)^0}{0!}e^{-\lambda h} + \lambda h e^{-\lambda h}\right] = 1 - (1+\lambda h)e^{-\lambda h} \\ &= 1 - (1+\lambda h)\left[1 - (\lambda h) + o(h)\right] \\ &= 1 - \left[1 - \lambda h + o(h) + \lambda h - \lambda^2 h^2 + \lambda h o(h)\right] \\ &= 1 - \left[1 - \lambda^2 h^2 + o(h)\right] = o(h). \end{aligned}$$

Para el recíproco de la demostración, supongamos que $\mathbf{X}$ es un proceso de conteo que satisface las condiciones (i)-(iii). Entonces, $N(0) = 0$, y por (i) tiene incrementos independientes y estacionarios. Por lo tanto, es suficiente mostrar que $N(t) - N(0) = N(t) \sim$Poisson (λt), para cualquier $t \geq 0$, es decir,

$$P\left[N(t)=n\right] = \frac{(\lambda t)^n}{n!}e^{-\lambda t}, n = 0, 1, 2, ... \tag{1.4}$$

Denotemos $P_n(t) = P\left[N(t)=n\right]$. Demostraremos primero que $P_0(t)$ satisface la ecuación diferencial

$$P_0'(t) = -\lambda P_0(t),\ t \geq 0, \tag{1.5}$$

con condición inicial $P_0(0) = 1$. De hecho, resolviendo esta ecuación se tiene que, para alguna constante k,

$$P_0(t) = ke^{-\lambda t}.$$

Como $P_0(0) = 1$, entonces $k = 1$, por lo tanto, la solución a (1.5) es,

$$P_0(t) = e^{-\lambda t} = P\left[N(t) = 0\right]. \tag{1.6}$$

Después demostraremos que $P_n(t)$ satisface la ecuación diferencial

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n = 1, 2, ..., \tag{1.7}$$

con condición inicial $P_n(0) = 0, n = 1, 2, ...$ Entonces,

$$P_n'(t) + \lambda P_n(t) = \lambda P_{n-1}(t), \quad n = 1, 2, ...,$$

lo cual implica

$$e^{\lambda t}\left[P_n'(t) + \lambda P_n(t)\right] = \lambda e^{\lambda t} P_{n-1}(t).$$

De aquí,

$$\frac{d}{dt}\left[e^{\lambda t} P_n(t)\right] = \lambda e^{\lambda t} P_{n-1}(t). \tag{1.8}$$

Tomando $n = 1$, y usando (1.6),

$$\frac{d}{dt}\left[e^{\lambda t} P_1(t)\right] = \lambda \text{ si, y solo si } P_1(t) = (\lambda t + c)e^{-\lambda t}.$$

Como $P_1(0) = 0$ tenemos que $c = 0$, lo cual implica $P_1(t) = \lambda t e^{-\lambda t} = P\left[N(t) = 1\right]$.

Demostraremos ahora por inducción matemática que

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 1, 2, ...$$

Supongamos que se cumple para $n - 1$. Entonces por (1.8)

$$\frac{d}{dt}\left[e^{\lambda t} P_n(t)\right] = \lambda e^{\lambda t} P_{n-1}(t) = \lambda e^{\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!}.$$

Por lo tanto,

$$e^{\lambda t} P_n(t) = \frac{\lambda^n t^n}{n!} + c$$

Como $P_n(0) = 0$ obtenemos

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 1, 2, ...$$

Por lo anterior, para concluir la demostración, es suficiente con demostrar que $P_0(t)$ y $P_n(t)$ satisfacen las ecuaciones diferenciales (1.5) y (1.7), respectivamente.

Ecuación diferencial para $P_0(t)$.

Para cualquier $t \geq 0$ y $h > 0$ tenemos

$$\begin{aligned}
P_0(t+h) &= P\left[N(t+h)=0\right] = P\left[N(t)=0, N(t+h)-N(t)=0\right] \\
&= P\left[N(t)=0\right] P\left[N(t+h)-N(t)=0\right] \qquad \text{(inc. ind.)} \\
&= P_0(t) P\left[N(h)=0\right] \qquad \text{(inc.est.)} \\
&= P_0(t)\left(1 - P\left[N(h)\geq 1\right]\ \right) \\
&= P_0(t)\left(1 - P\left[N(h)=1\right]\ - P\left[N(h)\geq 2\right]\ \right) \\
&= P_0(t)(1-\lambda h + o(h)).
\end{aligned}$$

Cabe señalar que la última igualdad se sigue de las condiciones (ii) y (iii). De aquí tenemos que

$$\frac{P_0(t+h)-P_0(t)}{h} = \frac{-\lambda h P_0(t)}{h} + \frac{o(h)}{h}.$$

donde, tomando límite cuando $h \to 0$ obtenemos la ecuaciòn diferencial (1.5), es decir,

$$P_0'(t) = -\lambda P_0(t).$$

Ecuación diferencial para $P_n(t)$, $\quad n = 1, 2, ...$

Para cualquier $t \geq 0$ y $h > 0$,

$$\begin{aligned}
P_n(t+h) &= P\left[N(t+h)=n\right] = \sum_{j=0}^{n} P\left[N(t+h)-N(t)=j, N(t)=n-j\right] \\
&= P\left[N(t+h)-N(t)=0, N(t)=n\right] \\
&\quad + P\left[N(t+h)-N(t)=1, N(t)=n-1\right] \\
&\quad + \sum_{j=2}^{n} P\left[N(t+h)-N(t)=j, N(t)=n-j\right]
\end{aligned}$$

Por otro lado,

$$
\begin{aligned}
& \sum_{j=2}^{n} P\left[N(t+h)-N(t)=j, N(t)=n-j\right] \\
\leq & \sum_{j=2}^{n} P\left[N(t+h)-N(t)=j\right] \\
= & \sum_{j=2}^{n} P\left[N(h)=j\right] \qquad \text{(Inc. est.)} \\
\leq & \; P\left[N(h)\geq 2\right]=o(h) \text{ (por (iii))}. \qquad (1.9)
\end{aligned}
$$

Por lo tanto, de (1.9) se obtiene

$$
\begin{aligned}
P_n(t+h) &= P\left[N(t)=n\right]P\left[N(t+h)-N(t)=0\right]+ \\
& \quad P\left[N(t)=n-1\right]P\left[N(t+h)-N(t)=1\right]+o(h) \quad \text{(inc.ind)} \\
&= P\left[N(h)=0\right]P\left[N(t)=n\right]+P\left[N(h)=1\right]P\left[N(t)=n-1\right] \\
& \quad +o(h) \quad \text{(inc.est)} \\
&= (1-\lambda h+o(h))P_n(t)+(\lambda h+o(h))P_{n-1}(t)+o(h)
\end{aligned}
$$

De aquí

$$
\frac{P_n(t+h)-P_n(t)}{h}=\frac{-\lambda h P_n(t)}{h}+\frac{\lambda h P_{n-1}(t)}{h}+o(h)
$$

Por lo cual, haciendo $h\to 0$, obtenemos la ecuación diferencial (1.7), es decir,

$$
P_n'(t)=-\lambda P_n(t)+\lambda P_{n-1}(t), n=1,2,...
$$

■

1.4.2. Análisis del tiempo de espera

Sea $\mathbf{X}=\{N(t),t\geq 0\}$ un proceso de Poisson con intensidad $\lambda>0$ en el que $N(t)$ cuenta el número de veces que ha ocurrido un evento E durante el intervalo de tiempo $[0,t]$. En esta parte estamos interesados en estudiar la distribución del tiempo entre la ocurrencia de eventos. Generalmente en la literatura a esta v.a. se le llama tiempo entre arribos (o de llegada, o de espera).

Supongamos que un evento E ocurre durante los tiempos:

$$0 < t_1 < t_2 < ... < t_n < ...,$$

en donde t_n es una v.a. que representa el tiempo en el que ocurre E por $n-$ésima vez. Entonces, los tiempos entre arribos están definidos por las v.a.

$$\tau_n = t_n - t_{n-1},\ n = 1, 2, \ldots,$$

con $t_0 = 0$. Note que, $t_n = \tau_1 + \tau_2 + \cdots + \tau_n$.

Es fácil ver que $\{\tau_n\}$ es una sucesión de v.a. i.i.d. cuya distribución es exponencial con parámetro λ. En efecto, observemos que los eventos $[\tau_1 > t]$ y $[N(t) = 0]$ son equivalentes. Entonces, similar a (1.1), tenemos,

$$F_{\tau_1}(t) = \begin{cases} 1 - \exp(-\lambda t) & si \quad t \geq 0 \\ \\ 0 & si \quad t < 0 \end{cases}, \tag{1.10}$$

es decir, $\tau_1 \sim \exp(\lambda)$. Para mostrar que $\tau_2 \sim \exp(\lambda)$ y que es independiente de τ_1, observemos lo siguiente:

$$\begin{aligned} P\left[\tau_2 \leq t | \tau_1 > s\right] &= 1 - P\left[\tau_2 > t | \tau_1 > s\right] = 1 - P\left[\tau_2 > t | N(s) = 0\right] \\ &= 1 - P\left[N(s+t) - N(s) = 0 | N(s) = 0\right] \\ &= 1 - P\left[N(s+t) - N(s) = 0\right] \quad \text{(inc. Ind.)} \\ &= 1 - P\left[N(t) = 0\right] \quad \text{(inc. est.)} \\ &= 1 - \exp(-\lambda t). \end{aligned} \tag{1.11}$$

De aquí, la función de densidad de τ_2 dado $\tau_1$es

$$f_{\tau_2|\tau_1}(t|s) = \lambda \exp(-\lambda t), \quad t \geq 0,$$

con lo cual obtenemos

$$\begin{aligned} f_{\tau_2}(t) &= \int_0^\infty f_{\tau_2|\tau_1}(t|s) f_{\tau_1}(s) ds \\ &= \int_0^\infty \lambda \exp(-\lambda t) f_{\tau_1}(s) ds \\ &= \lambda \exp(-\lambda t) \int_0^\infty f_{\tau_1}(s) ds \\ &= \lambda \exp(-\lambda t) \end{aligned}$$

Por lo tanto, $\tau_2 \sim \exp(\lambda)$, y además, debido a que el lado derecho de (1.11) no depende de s, tenemos que τ_1 y τ_2 son independientes. Estos argumentos se repiten para mostrar que $\{\tau_n\}$ es una sucesión de v.a. i.i.d. cuya distribución es exponencial con parámetro λ.

De lo anterior tenemos que una característica del Proceso de Poisson es que el tiempo entre arribos es exponencial. Si $\{\tau_n\}$ es una sucesión de v.a i.i.d. con distribución arbitraria, al proceso estocástico resultante se le conoce como *Proceso de Renovación*. Concretamente tenemos la siguiente definición.

Definición 1.4.2 *Sea* $\mathbf{X} = \{N(t), t \geq 0\}$ *un proceso de conteo. Decimos que* $\mathbf{X}$ *es un proceso de renovación, si los tiempos entre arribos son v.a. i.i.d.*

Obsérvese que para cada $t \geq 0$ podemos definir $N(t)$ como:

$$N(t) = \max\{n : t_n \leq t\}.$$

Para cada $t \geq 0$, definamos la v.a.

$$\gamma(t) := t_{N(t)+1} - t,$$

es decir, $\gamma(t)$ representa el tiempo entre t y el próximo arribo que ocurre.

A continuación establecemos una propiedad importante de $\gamma(t)$ cuya demostración puede consultarse en [8] .

Proposición 1.4.1 *Sea* $\mathbf{X} = \{N(t), t \geq 0\}$ *un proceso de renovación con tiempos entre arribos exponencialmente distribuidos con parametro* λ*. Entonces* $\gamma(t) \sim \exp(\lambda)$*.*

Teorema 1.4.2 *Un proceso estocástico* $\mathbf{X} = \{N(t), t \geq 0\}$ *es un proceso de Poisson con intensidad media* λ *si, y solo si* $\mathbf{X}$ *es un proceso de renovación cuyos tiempos entre arribos* $\tau_n = t_n - t_{n-1}$, $n = 1, 2, \ldots,$ *son v.a. i.i.d. con* $\tau_n \sim \exp(\lambda)$*.*

Demostración. ($\Longrightarrow$) Esta parte ya se demostró previamente para (1.10) y (1.11).

($\Longleftarrow$) Supóngase ahora que $\mathbf{X}$ es un proceso de renovación con tiempos entre arribos i.i.d. y $\tau_n \sim \exp(\lambda)$. De la definición de un proceso de Poisson, para demostrar el teorema es suficiente demostrar que:
(i) $\mathbf{X}$ tiene incrementos independientes
(ii) $\mathbf{X}$ tiene incrementos estacionarios con

$$N(t+h) \ - N(t) \sim P(\lambda h), \qquad t \geq 0, h > 0.$$

Para ver esto, fijemos $s > 0$ y consideremos el proceso $\{X'(t) = N(t) - N(s),\ t \geq s\}$. Mostraremos que $X'(t)$ es un proceso de renovación. Denotemos por $\tau_1', \tau_2', ...$ los tiempos entre arribos. Es decir, τ_1' es el tiempo desde s hasta que ocurre el primer evento, τ_2' el tiempo entre el primer y el segundo eventos, etc. Es claro que

$$\tau_n' \sim exp(\lambda) \quad \text{para} \ \ n \geq 2$$

Además, por la proposicion (1.4.1), $\tau_1' = \gamma(t) \sim \exp(\lambda)$ independientemente del número de arribos que ocurren antes de s , es decir, independientemente de los valores de $N(t')$ para $t' \leq s$.

Entonces,

$$P\left[N(t) - N(s) = n | N(t') \ = k\right] = P\left[N(t) - N(s) = n\right] = P\left[N(t-s) = n\right].$$

donde la última igualdad se debe a que los tiempos entre arribos de ambos procesos X y X' tienen la misma distribución exponencial.De aquí y por la definición de probabilidad condicional obtenemos

$$\frac{P\left[N(t) - N(s) = n, N(t') \ - N(0) = k\right]}{P[N(t') \ - N(0) = k]} = P\left[N(t) - N(s) = n\right]$$

Esto implica la independencia de los incrementos $N(t)-N(s)$ y $N(t')=N(t')-N(0)$, lo cual a su vez implica la parte (i).

Para probar (ii), observemos primero que, para cualquier proceso de renovación $\mathbf{X}=\{N(t),t\geq 0\}$,

$$\begin{aligned} N(t) &\leq n \quad \text{si, y solo si} \quad t_{n+1}>t \\ N(t) &= n \quad \text{si, y solo si} \quad t_n \leq t \quad \text{y} \quad t_{n+1}>t, \quad \text{es decir,} \\ t_n &\leq t \leq t_{n+1}, \end{aligned}$$

donde t_n es el tiempo en que ocurre el $n-$ésimo arribo o renovación. Entonces,

$$\begin{aligned} P\{N(t) &\leq n\} = P\{\ t_{n+1}>t\} \\ P\{N(t) &= n\} = P\{t_n \leq\ t\} - P\{t_{n+1}\leq t\} \qquad n=1,2,... \quad (1.12) \\ P\{N(t) &= 0\} = P(\tau_1>t), \quad (1.13) \end{aligned}$$

donde (1.12) se obtiene de las siguientes relaciones:

$$\begin{aligned} P\{N(t) &= n\} = P\{N(t)\leq n\} - P\{N(t)\leq n-1\} \\ &= P\{\ t_{n+1}>t\} - P\{\ t_n>t\} \\ &= P\{t_n \leq\ t\} - P\{t_{n+1}\leq t\} \end{aligned}$$

Puesto que $\{\tau_n\}$ son v.a. i.i.d. con $\tau_n \sim \exp(\lambda)$, y $t_n=\tau_1+...+\tau_n \sim \Gamma(n,\lambda)$ (ver Proposición 1.2.2), de (1.12) y (1.13) tenemos:

$$\begin{aligned} P\{N(t) &= 0\} = P(\tau_1>t) = e^{-\lambda t}, \\ P\{N(t) &= n\} = \int_0^t \frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}dx - \int_0^t \frac{\lambda^{n+1}}{(n)!}x^n e^{-\lambda x}dx \\ &= \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \qquad n=1,2,.... \end{aligned}$$

Por lo tanto, para $0 \leq s \leq t$,

$$\begin{aligned} P\{N(t) - N(s) &= n\} = P\{N(t-s) = n\} \\ &= e^{-\lambda(t-s)}\frac{(\lambda(t-s))^n}{n!}, n = 0, 1, ..., \end{aligned}$$

con lo cual concluye la demostración del teorema. ■

Capítulo 2

Cadenas de Markov a Tiempo Continuo

2.1. Introducción

Los procesos de Markov constituyen una de las clases de procesos estocásticos más estudiados. La característica que los distingue es la llamada propiedad de Markov la cual establece, en términos generales, que dado el estado presente del sistema, la distribución de probabilidad del proceso en un tiempo futuro es independiente del pasado. Con esta propiedad, los procesos de Markov nos permiten modelar sistemas dinámicos estocásticos que evolucionan en el tiempo los cuales aparecen en diferentes áreas como economía, biología, ciencias computacionales, investigación de operaciones, etc.

El estudio de los procesos de Markov se clasifica de acuerdo al tiempo de observación (discreto o continuo) y al espacio de estados (finito numerable o no numerable). A un proceso de Markov con espacio de estados numerable se le llama cadena de Markov. De esta manera tenemos cadenas de Markov a tiempo discreto o continuo.

En este capítulo estudiaremos cadenas de Markov a tiempo continuo y analizaremos una clase particular de ellos, los llamados procesos de nacimiento y muerte. Estos procesos son importantes debido a que los sistemas de espera que estudiaremos en el próximo capítulo se pueden modelar como procesos de nacimiento y muerte.

2.2. Definición y propiedades

Definición 2.2.1 *Un proceso estocástico* $\mathbf{X} = \{X(t), t \in \mathbb{R}\}$, *con espacio de estados discreto* S, *es una cadena de Markov en tiempo continuo si para cualquier entero* n *y tiempos* $t_1 < t_2 < ... < t_{n+1}$ *en* $\mathbb{R}_+$, *se satisface que la distribución condicional de* $X(t_{n+1})$ *dados los valores de* $X(t_1), X(t_2), ..., X(t_n)$ *solo depende de* $X(t_n)$. *Es decir, para cualesquiera* $x, x_1, ..., x_n \in S$,

$$P\{X(t_{n+1}) = x | X(t_1) = x_{1,...}, X(t_n) = x_n\} = P\{X(t_{n+1}) = x | X(t_n) = x_n\}. \tag{2.1}$$

A la relación (2.1) se le conoce como Propiedad de Markov.

En una cadena de Markov en tiempo continuo el espacio de estados es un conjunto numerable $S = \{x_0, x_1, ...\}$, pero para simplificar la notación, supondremos que $S = \{0, 1, 2, ...\}$.

Definición 2.2.2 *Sean* $i, j \in S$ *y* $0 \leq s \leq t$. *Definimos la probabilidad de transición como*

$$P_{ij}(s,t) = P\{X(t) = j | X(s) = i\}.$$

Solo consideraremos procesos homogéneos en el tiempo, esto es, para i y j fijos, la probabilidad de transición $P_{ij}(s,t)$ depende solo de $t-s$, es decir,

$$P_{ij}(s,t) = P_{ij}(0, t-s), \quad \forall s < t.$$

En este caso las probabilidades de transición dependen sólo de una variable y escribimos

$$\begin{aligned} P_{ij}(t) \quad &: = P_{ij}(0,t) = P\{X(t) = j | X(0) = i\} \qquad (2.2) \\ &= \quad P\{X(s+t) = j | X(s) = i\} \quad \forall s \geq 0. \end{aligned}$$

Supondremos que estas probabilidades de transición satisfacen que $P_{ij}(t) \geq 0$ para todo $t \geq 0$, i, $j \in S$, y

$$\sum_{j=0}^{\infty} P_{ij}(t) = 1 \text{ para todo } t \geq 0, i \in S \text{ fijos.}$$

Además tenemos el siguiente resultado.

Teorema 2.2.1 *Sea* $\mathbf{X} = \{X(t), t \geq 0\}$ *una cadena de Markov con probabilidades de transición* $P_{ij}(t)$. *Entonces*

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(s)P_{kj}(t) \quad \forall i,j \in S, t,s \geq 0. \tag{2.3}$$

Demostración. La relación (2.3) se sigue de (2.1) y (2.2). En efecto,

$$\begin{aligned}
P_{ij}(t+s) &= P\{X(t+s) = j | X(0) = i\} \\
&= \sum_{k=0}^{\infty} P\{X(t+s) = j, X(s) = k | X(0) = i\} \\
&= \sum_{k=0}^{\infty} P\{X(s) = k | X(0) = i\} P\{X(t+s) = j | X(0) = i, X(s) = k\} \\
&= \sum_{k=0}^{\infty} P_{ik}(s) P\{X(t+s) = j | X(s) = k\} \quad \text{por (2.2) y (2.1)} \\
&= \sum_{k=0}^{\infty} P_{ik}(s)P_{kj}(t).
\end{aligned}$$

■

La ecuación (2.3) se llama la ecuación de Chapman-Kolmogorov.

Ejemplo 2.2.2 *Un proceso de Poisson* $\mathbf{X} = \{N(t), t \geq 0\}$ *con parámetro* $\lambda > 0$ *es una cadena de Markov homogénea a tiempo continuo. En efecto*

$$\begin{aligned}
& P[N(t_{n+1}) = j | N(t_1) = x_1, ..., N(t_n) = i] \\
= \; & \frac{P[N(t_1) = x_1, ..., N(t_n) = i, N(t_{n+1}) = j]}{P[N(t_1) = x_1, ..., N(t_n) = i]} \\
= \; & \frac{P[N(t_1) = x_1, N(t_2) - N(t_1) = x_2 - x_1, ..., N(t_{n+1}) - N(t_n) = j - i]}{P[N(t_1) = x_1, N(t_2) - N(t_1) = x_2 - x_1, ..., N(t_n) - N(t_{n-1}) = i - x_{n-1}]} \\
= \; & \frac{P[N(t_1) = x_1] ..., P[N(t_n) - N(t_{n-1}) = i - x_{n-1}] \, P[N(t_{n+1}) - N(t_n) = j - i]}{P[N(t_1) = x_1] \, P[N(t_2) - N(t_1) = x_2 - x_1] ..., P[N(t_n) - N(t_{n-1}) = i - x_{n-1}]} \\
= \; & P[N(t_{n+1}) - N(t_n) = j - i]
\end{aligned}$$

Por otro lado

$$\begin{aligned} P\left[N(t_{n+1})=j|N(t_n)=i\right] &= \frac{P\left[N(t_{n+1})=j, N(t_n)=i\right]}{P\left[N(t_n)=i\right]} \\ &= \frac{P\left[N(t_n)=i, N(t_{n+1})-N(t_n)=j-i\right]}{P\left[N(t_n)=i\right]} \\ &= \frac{P\left[N(t_n)=i\right]P\left[N(t_{n+1})-N(t_n)=j-i\right]}{P\left[N(t_n)=i\right]} \\ &= P\left[N(t_{n+1})-N(t_n)=j-i\right] \end{aligned}$$

Por lo tanto X es una cadena de Markov homogénea con probabilidades de transición:

$$\begin{aligned} P_{ij}(t) &= P\left\{N(t+s)-N(s)=j-i\right\} \\ &= \frac{e^{-\lambda t}(\lambda t)^{j-i}}{(j-i)!}, \quad \textit{para todo } t,s\geq 0, \quad j\geq i. \end{aligned}$$

■

En el resto del trabajo supondremos que las probabilidades de una cadena de Markov en tiempo continuo satisfacen lo siguiente:

$$\lim_{t\to 0^+} P_{ij}(t)=\delta_{ij}=\begin{cases} 1 & si \quad i=j \\ 0 & si \quad i\neq j \end{cases} \tag{2.4}$$

Es decir, pediremos que las probabilidades de transición sean continuas en $t=0$. Esta condición implica que con probabilidad 1, la cadena de Markov permanecerá un tiempo positivo en el estado i antes de moverse a un estado diferente. También asumiremos que

$$P_{ij}(0)=\delta_{ij}, \quad \forall i,j\in S. \tag{2.5}$$

Además usaremos la siguiente notación: $P_i(0)$, $i=0,1,...$, denota la distribución inicial del proceso, es decir,

$$P_i(0)=P\left\{X(0)=i\right\}, \quad i=0,1,...,$$

y

$$P_j(t) = P\{X(t) = j\}, \quad j = 0, 1, ...,$$

denotará la distribución de $X(t)$.

La distribución inicial, junto con las probabilidades de transición (2.2) determinan la distribución de $X(t)$ en cualquier tiempo $t \geq 0$, como lo establece el siguiente resultado.

Teorema 2.2.3 *Sea* $\mathbf{X} = \{X(t), t \geq 0\}$ *una cadena de Markov con probabilidades de transición* $P_{ij}(t)$ *y distribución inicial* $P_i(0)$. *Entonces*

$$P_j(t) = \sum_{i=0}^{\infty} P_i(0) P_{ij}(t). \tag{2.6}$$

Demostración. La demostración está basada en el Teorema de la Probabilidad Total. En efecto,

$$\begin{aligned} P_j(t) &= P\{X(t) = j\} = \sum_{i=0}^{\infty} P\{X(0) = i\} P\{X(t) = j | X(0) = i\} \\ &= \sum_{i=0}^{\infty} P_i(0) P_{ij}(t). \end{aligned}$$

■

2.3. Probabilidades infinitesimales

Supondremos que las probabilidades de transición $P_{ij}(t)$ son derivables en cero.

Definición 2.3.1 *Sea* $\mathbf{X} = \{X(t), t \geq 0\}$ *una cadena de Markov con probabilidades de transición (2.2). La probabilidad infinitesimal* q_{ij} *se define como la derivada de* $P_{ij}(t)$ *en* $t = 0$.

Observación 2.3.1 *i) A la probabilidad infinitesimal también se le llama intensidad de transición.*

ii) Por (2.5)

$$\begin{aligned} q_{ij} &= \frac{\partial}{\partial t} P_{ij}(t) \mid_{0^+} \\ &= \lim_{t\to 0^+} \frac{1}{t}(P_{ij}(t) - \delta_{ij}). \end{aligned}$$

Es decir,

$$\begin{aligned} (a)\quad q_{ij} &= \lim_{t\to 0^+} \frac{1}{t} P_{ij}(t) \quad para\ i \neq j \\ (b)\quad q_{ii} &= \lim_{t\to 0^+} \frac{1}{t}\left(P_{ii}(t) - 1\right). \end{aligned} \tag{2.7}$$

c) Denotando $q_i := -q_{ii}$ podemos escribir (2.7)(a)-(b) como:

$$q_i = \lim_{t\to 0^+} \frac{1}{t}\left(1 - P_{ii}(t)\right).$$

En general las probabilidades infinitesimales (2.7)(a)-(b) pueden no existir. Sin embargo, todos los procesos que estudiaremos en el trabajo tienen probabilidades infinitesimales finitas. De hecho en libros avanzados sobre procesos estocásticos se demuestra que la condición (2.4) implica que los límites en (2.7)(a)-(b) existen y son finitos.

Para no meternos en cuestiones técnicas, supondremos que los límites en (2.7)(a)-(b) siempre existen y son finitos. En este caso, observemos que

$$\begin{aligned} \text{(a)}\quad P_{ij}(t) &= q_{ij}t + o(t) \quad \text{cuando} \quad t \to 0,\ i \neq j \\ \text{(b)}\quad P_{ii}(t) &= 1 - q_i t + o(t) \quad \text{cuando} \quad t \to 0 \end{aligned} \tag{2.8}$$

Estas expresiones nos dicen que, cuando $t \to 0^+$, las probabilidades $P_{ij}(t)$ y $1 - P_{ii}(t)$ son, excepto por términos $o(t)$, proporcionales a t, con constantes de proporcionalidad q_{ij} y q_i, respectivamente.

Ejemplo 2.3.1 *Sea $\mathbf{X} = \{N(t), t \geq 0\}$ un proceso de Poisson con intensidad λ. Para calcular las probabilidades infinitesimales, primero observemos que $P_{ij}(t) = 0$ para $j < i$, y para $j \geq i$:*

$$\begin{aligned} P_{ij}(t) &= P\{N(t+s) = j | N(s) = i\} \\ &= P\{N(t+s) - N(s) = j - i | N(s) = i\} \\ &= P\{N(t+s) - N(s) = j - i\} \end{aligned}$$

Entonces, por el Teorema 1.4.1:

$$P_{ij}(t) = \begin{cases} 1 - \lambda t + o(t) & si \quad j = i \\ \\ \lambda t + o(t) & si \quad j = i+1 \\ \\ o(t) & si \quad j \geq i+2 \end{cases}$$

Luego, de (2.7) ó (2.8), vemos que las probabilidades infinitesimales para el proceso de Poisson están dadas por:

$$\begin{aligned} q_i &= -q_{ii} = \lambda \\ q_{i,i+1} &= \lambda \\ q_{i,j} &= 0 \quad para\ j \neq i, i+1. \blacksquare \end{aligned}$$

2.4. Ecuaciones de Kolmogorov

En esta sección presentaremos otras propiedades de las probabilidades de transición por medio de dos ecuaciones diferenciales llamadas ecuaciones de Kolmogorov. Estas ecuaciones serán útiles para caracterizar la distribución de $X(t)$ en cualquier tiempo t, como solución de una ecuación diferencial como se hizo con el proceso de Poisson.

Teorema 2.4.1 *Sea* $\mathbf{X} = \{X(t), t \geq 0\}$ *una cadena de Markov homogénea con espacio de estados finito* $S = \{0, 1, ..., n\}$*. Además supongamos que* $P_{ij}(t)$ *tiene derivada finita para todo* $i, j \in S$*. Entonces las probabilidades de transición satisfacen las siguientes ecuaciones diferenciales:*

$$P'_{ij}(t) = -q_i P_{ij}(t) + \sum_{k \neq i} q_{ik} P_{kj}(t) \tag{2.9}$$

y

$$P'_{ij}(t) = -P_{ij}(t) q_j + \sum_{k \neq j} P_{ik}(t) q_{kj} \tag{2.10}$$

Demostración. *Ambas ecuaciones se obtienen derivando la ecuación de Chapman-Kolmogorov y aplicando el hecho de que* S *es finito y* $P_{ij}(t)$ *tiene derivada finita.*

Derivando ambos lados de la ecuación de Chapman-Kolmogorov respecto a s y luego evaluando s=0 obtenemos

$$P'_{ij}(t+s) = \sum_{k=0}^{m} P'_{ik}(s)P_{kj}(t)$$

Entonces,

$$\begin{aligned} P'_{ij}(t) &= \sum_{k=0}^{m} P'_{ik}(0)P_{kj}(t) = \sum_{k=0}^{m} q_{ik}P_{kj}(t) \\ &= q_{ii}P_{ij}(t) + \sum_{k\neq i} q_{ik}P_{kj}(t) \end{aligned}$$

como $q_i = -q_{ii}$, *tenemos que*

$$P'_{ij}(t) = -q_iP_{ij}(t) + \sum_{k\neq i} q_{ik}P_{kj}(t)$$

lo cual demuestra (2.9).

Para obtener (2.10) se sigue un procedimiento similar. Derivamos ambos lados de la ecuación de Chapman-Kolmogorov respecto a t *y luego evaluamos* $t = 0$, *para obtener:*

$$\begin{aligned} P'_{ij}(s) &= \sum_{k=0}^{m} P'_{ik}(s)P_{kj}(0) \\ &= -P_{ij}(s)q_j + \sum_{k\neq j} P_{ik}(s)q_{kj} \end{aligned}$$

esta relación demuestra (2.10). ∎

A las ecuaciones (2.9) y (2.10) se les conoce como ecuaciones de Kolmogorov hacia atrás y hacia adelante, respectivamente. En este caso se obtuvieron bajo la hipótesis de que S sea finito y $P'_{ij}(t) < \infty$. Sin embargo, son válidas en contextos mas generales (ver [1]).

Las ecuaciones de Kolmogorov se pueden escribir en forma matricial como sigue. Definamos las matrices:

$$P(t) = (P_{ij}(t)),\ Q = (q_{ij}),$$

de probabilidades de transición e infinitesimales, respectivamente. Recordemos que $q_{ii} = -q_i$. Entonces, la ecuación (2.9) se puede escribir como la ecuación diferencial matricial:

$$P'(t) = QP(t) \tag{2.11}$$

con condición inicial $P(0) = I$ (I la matriz identidad).

Similarmente, la forma matricial para la ecuación (2.10) es

$$P'(t) = P(t)Q, \;\; P(0) = I \tag{2.12}$$

Ejemplo 2.4.2 *Sea $\mathbf{X} = \{N(t), t \geq 0\}$ un proceso de Poisson con intensidad λ. Por el Ejemplo 2.3.1, la matriz de probabilidades infinitesimales toma la forma:*

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & ... \\ 0 & -\lambda & \lambda & 0 & ... \\ 0 & 0 & -\lambda & \lambda & ... \\ . & . & . & . & ... \\ . & . & . & . & ... \end{bmatrix},$$

y la ecuación de Kolmogorov hacia atrás (2.9) resulta:

$$\begin{aligned} P'_{ij}(t) &= q_{ii}P_{ij}(t) + q_{i,i+1}P_{i+1,j}(t) \\ &= -\lambda P_{ij}(t) + \lambda P_{i+1,j}(t). \end{aligned}$$

Además, la ecuación hacia adelante (2.10) en este caso es:

$$\begin{aligned} P'_{ij}(t) &= P_{ij}(t)q_{jj} + P_{i,j-1}(t)q_{j-1,j} \\ &= -P_{ij}(t)\lambda + P_{i,j-1}(t)\lambda. \end{aligned}$$

La ecuación hacia adelante (2.10) puede usarse a fín de obtener una ecuación diferencial para la distribución $P_j(t) = P\{X(t) = j\}$ del proceso en el tiempo t, tal como lo establece el siguiente resultado.

Teorema 2.4.3 *Sea $\mathbf{X} = \{X(t), t \geq 0\}$ una cadena de Markov con probabilidades de transición $P_{ij}(t)$ e infinitesimales q_{ij}. Entonces bajo las hipótesis del Teorema 2.4.1 $P_j(t)$ satisface la ecuación diferencial*

$$P'_j(t) = \sum_{k=0}^{\infty} P_k(t)q_{kj} = -P_j(t)q_j + \sum_{k\neq j} P_k(t)q_{kj}. \tag{2.13}$$

Demostración. De (2.6) tenemos que

$$P_j(t) = \sum_{i=0}^{m} P_i(0)P_{ij}(t).$$

Derivando en ambos lados de esta expresión obtenemos,

$$P_j'(t) = \sum_{i=0}^{m} P_i(0)P_{ij}'(t).$$

Sustituyendo (2.10) en esta ecuación:

$$\begin{aligned} P_j'(t) &= \sum_i P_i(0)\left(\sum_k P_{ik}(t)q_{kj}\right) \\ &= \sum_k \left(\sum_i P_i(0)P_{ik}(t)q_{kj}\right) \\ &= \sum_k P_k(t)q_{kj}, \end{aligned}$$

donde la última igualdad es una consecuencia de (2.6). Esto implica que $P_j(t)$ satisface la ecuación diferencial (2.13). ■

Al igual que el Teorema 2.4.1, este resultado es de validez general (ver [1]).

Observación 2.4.1 *Si* $\mathbf{X} = \{N(t), t \geq 0\}$ *es un proceso de Poisson con intensidad* λ*, la ecuación (2.13) se reduce a:*

$$\begin{aligned} P_j'(t) &= -\lambda P_j(t) + \lambda P_{j-1}(t), \quad P_j(0) = 0, \quad j = 1, 2, ... \\ P_0'(t) &= -\lambda P_0(t), \quad P_0(0) = 1, \end{aligned}$$

las cuales coinciden con las ecuaciones (1.5) y (1.7).

2.5. Distribuciones Estacionarias

Recúerdese que una sucesión $\{\pi_k; \quad k = 0, 1, 2, ...\}$ es una distribución de probabilidad si $\pi_k \geq 0$ para todo k, y $\sum_{k=0}^{\infty} \pi_k = 1$.

Definición 2.5.1 *Decimos que una cadena de Markov* $\{X(t),\ t \geq 0\}$ *tiene una distribución estacionaria si existe una distribución de probabilidad*

$$\{\pi_k; \quad k = 0, 1, 2, ...\}$$

tal que para todo $i,\ k \in S$,

$$\lim_{t\to\infty} P_{ik}(t) = \pi_k. \tag{2.14}$$

Observación 2.5.1 *El nombre de distribución estacionaria se refiere a que para cualquier distribución inicial, la distribución* $P_k(t) = P\left[X(t) = k\right]$ *converge a* π_k *cuando* $t \to \infty$. *En efecto, sea* $\{P_k; \quad k = 0, 1, 2, ...\}$ *cualquier distribución inicial de la cadena, es decir,* $P_k(0) = P\left[X(0) = k\right]$, *entonces de (2.6) tenemos*

$$\begin{aligned}\lim_{t\to 0} P_k(t) &= \lim_{t\to\infty} \sum_{i=0}^{\infty} P_i(0) P_{ik}(t) \qquad (2.15)\\ &= \sum_{i=0}^{\infty} P_i(0) \lim_{t\to\infty} P_{ik}(t)\\ &= \pi_k,\end{aligned}$$

donde la segunda y tercera igualdad se siguen por el hecho de que $P_{ik}(t)$ *está uniformemente acotada y que* $\sum_{i=0}^{\infty} P_i(0) = 1$.

Si la cadena de Markov tiene una distribución estacionaria, presentaremos una manera general de cómo calcularla. Consideremos la ecuación de Chapman-Kolmogorov:

$$P_{ik}(t+s) = \sum_{j=0}^{\infty} P_{ij}(s) P_{jk}(t).$$

Entonces, tomando el límite cuando $s \to \infty$, se sigue que

$$\pi_k = \sum_{j=0}^{\infty} \pi_j P_{jk}(t).$$

En principio, la solución a este sistema de ecuaciones nos daría la distribución estacionaria. No obstante, en muchos casos el cálculo de las probabilidades de transición no es fácil. En este sentido, por las condiciones que hemos venido suponiendo, lo deseable es calcular π_k en términos de las probabilidades infinitesimales. Para lo cual, derivando con respecto a t, en $t = 0$, y suponiendo que se cumple el intercambio de suma infinita y diferenciación (e.g. cuando S es finito), obtenemos

$$0 = \sum_{j=0}^{\infty} \pi_j q_{jk} = \pi_k q_{kk} + \sum_{j \neq k} \pi_j q_{jk},$$

es decir,

$$\pi_k q_k = \sum_{j \neq k} \pi_j q_{jk}, \ \ k = 0, 1, ... \tag{2.16}$$

Por lo tanto, si se sabe que los límites π_k en (2.14) existen, entonces sus valores pueden determinarse a partir del sistema de ecuaciones (2.16).

2.6. Procesos de nacimiento y muerte

Los procesos de nacimiento y muerte son un tipo de cadenas de Markov en tiempo continuo los cuales aplicaremos en el estudio de los sistemas de colas en el próximo capítulo.

Definición 2.6.1 *Un proceso de nacimiento y muerte es una cadena de Markov (homogénea) en tiempo continuo* $\mathbf{X} = \{X(t), t \geq 0\}$ *con espacio de estados* $S = \{0, 1, 2, ...\}$ *cuyas probabilidades infinitesimales son de la forma*

$$q_{ij} \ = \ \begin{cases} \lambda_i & si \ \ j = i+1 \quad (i \geq 0) \\ \mu_i & si \ \ j = i-1 \quad (i \geq 1) \\ -(\lambda_i + \mu_i) & si \ \ j = i \quad (i \geq 0) \\ 0 & si \ \ |j - i| \geq 2 \end{cases} \tag{2.17}$$

con $\mu_0 = 0$.

Observación 2.6.1 *La característica de los procesos de nacimiento y muerte es la relación*

$$q_{ij} = 0 \qquad si \quad |j - i| \geq 2,$$

la cual nos dice que las probabilidades infinitesimales desde un estado dado i solo pueden ocurrir hacia sus vecinos inmediatos $(i-1, i+1)$. De aquí, a los parámetros $\lambda_i = q_{i,i+1}$ se les llama tasas (o intensidades) de nacimiento, mientras que a los parámetros $\mu_i = q_{i,i-1}$ se les llaman tasas (o intensidades) de muerte.

De (2.8) y (2.17) tenemos que las probabilidades de transición cuando $t \to 0^+$ del proceso de nacimiento y muerte están dadas por:

$$P_{ij}(t) \quad = \quad \begin{cases} \lambda_i t + o(t) & si \quad j = i+1 \\ \\ \mu_i t + o(t) & si \quad j = i-1 \\ \\ 1 - (\lambda_i + \mu_i)t + o(t) & si \qquad j = i \\ \\ o(t) & si \quad |j-i| > 2 \end{cases} \tag{2.18}$$

Por otra parte, de (2.9)-(2.10), la ecuación de Kolmogorov hacia atrás para el proceso de nacimiento y muerte es:

$$P'_{ij}(t) = -(\lambda_i + \mu_i)P_{ij}(t) + \lambda_i P_{i+1,j}(t) + \mu_i P_{i-1,j}(t), \tag{2.19}$$

para $i = 0, 1, 2, ...$, $\mu_0 = 0$. Asimismo, de (2.10), la ecuación de Kolmogorov hacia adelante toma la forma:

$$P'_{ij}(t) = -P_{ij}(t)(\lambda_j + \mu_j) + P_{i,j-1}(t)\lambda_{j-1} + P_{i,j+1}(t)\mu_{j+1}. \tag{2.20}$$

Por otro lado, de (2.13), la ecuación diferencial para la distribución $P_j(t)$ del proceso en el tiempo t es:

$$P'_j(t) \quad = \quad -P_j(t)(\lambda_j + \mu_j) + P_{j-1}(t)\lambda_{j-1} + P_{j+1}(t)\mu_{j+1}, \tag{2.21}$$

para $j = 0, 1, ...$, con $P_{-1}(t) \equiv 0$. Más aún, la matriz de probabilidades infinitesimales se puede escribir como

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & ..... \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & ..... \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & ..... \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & ..... \\ & .... & .... & .... & & ..... \end{bmatrix}$$

Con todos estos elementos podemos calcular la distribución estacionaria, y de esta manera tener las características principales que describen un proceso de nacimiento y muerte.

2.6.1. Distribución estacionaria

La distribución estacionaria $\{\pi_j, \quad j = 0, 1, ...\}$ del proceso de nacimiento y muerte la calcularemos aplicando la ecuación (2.21) (ver (2.13)) considerando $P_j(t) = \pi_j$ constante. Entonces, puesto que la derivada de una constante es cero, de (2.21) tenemos que para $j = 1, 2, ...$ las π_j satisfacen

$$0 = -\pi_j(\lambda_j + \mu_j) + \pi_{j-1}\lambda_{j-1} + \pi_{j+1}\mu_{j+1} \tag{2.22}$$

y para $j = 0$:

$$0 = -\pi_0\lambda_0 + \pi_1\mu_1. \tag{2.23}$$

Supongamos que $\mu_j > 0$ para todo $j \geq 1$. Ahora consideremos el sistema de ecuaciones (2.22)-(2.23) con $j = 1$. De (2.23) se tiene que $\pi_1 = (\frac{\lambda_0}{\mu_1})\pi_0$ y de (2.22):

$$\begin{aligned}
\pi_2\mu_2 &= \pi_1(\lambda_1 + \mu_1) - \pi_0\lambda_0 \\
&= \left[\left(\frac{\lambda_0}{\mu_1}\right)\pi_0\right](\lambda_1 + \mu_1) - \pi_0\lambda_0 \\
&= \left(\frac{\lambda_0}{\mu_1}\right)\pi_0\lambda_1 + \lambda_0\pi_0 - \pi_0\lambda_0 \\
&= \left(\frac{\lambda_0\lambda_1}{\mu_1}\right)\pi_0
\end{aligned}$$

o bien,

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

Aplicando argumentos similares con $j = 2, 3, ...$, se pueden obtener π_3, π_4,..., y en general, se puede demostrar por inducción, que

$$\pi_j = (\frac{\lambda_0\lambda_1...\lambda_{j-1}}{\mu_1\mu_2....\mu_j})\pi_0, \qquad j = 1, 2, ... \tag{2.24}$$

Sin embargo, para que $\{\pi_j\}$ sea una distribución de probabilidad, debe satisfacer que $\sum_{j=0}^{\infty} \pi_j = 1$, es decir,

$$\pi_0 \left(1 + \sum_{j=1}^{\infty} \frac{\lambda_0 \lambda_1 \ldots \lambda_{j-1}}{\mu_1 \mu_2 \ldots . \mu_j} \right) = 1,$$

y por lo tanto, π_0 debe tener la forma

$$\pi_0 = \left(1 + \sum_{j=1}^{\infty} \frac{\lambda_0 \lambda_1 \ldots \lambda_{j-1}}{\mu_1 \mu_2 \ldots . \mu_j} \right)^{-1}. \tag{2.25}$$

Finalmente llegamos a que $\{\pi_j; j = 0, 1, \ldots\}$ definida por (2.24) y (2.25) es una distribución estacionaria para el proceso de nacimiento y muerte con probabilidades infinitesimales (2.17) si y solo si la serie

$$\sum_{j=1}^{\infty} \frac{\lambda_0 \lambda_1 \ldots \lambda_{j-1}}{\mu_1 \mu_2 \ldots . \mu_j} \tag{2.26}$$

converge, y en tal caso está dada por (2.24) y (2.25).

Capítulo 3

Colas Poissonianas

3.1. Introducción

En este capítulo estudiamos sistemas de espera para los cuales el tiempo entre arribos de clientes consecutivos y el tiempo de servicio tienen distribución exponencial, o equivalentemente, están modelados por procesos de Poisson. Estos sistemas se comportan como un proceso de nacimiento y muerte donde la llegada de un cliente al sistema representa un nacimiento, mientras que la salida de un cliente después de ser servido representa una muerte. Gran parte del material presentado en los capítulos anteriores será aplicado para estudiar dicha clase de sistemas de espera.

Iniciaremos el capítulo describiendo en general un sistema de espera. Después analizaremos colas poissonianas con uno o más servidores. Finalmente presentamos las simulaciones de la cola con un servidor a fin de ejemplificar la teoría.

3.2. Definiciones básicas y notación

Un sistema de espera puede describirse de la siguiente manera. Existe un sistema al que llegan clientes demandando cierto servicio. Los clientes que han arribado y que aún no han sido atendidos esperan en una cola. El sistema incluye todos los clientes, tanto los que están en la cola como los que están siendo atendidos.

De acuerdo a la descripción anterior, las componentes que intervienen en un sistema de espera son las siguientes:

- **Fuente.** Población de clientes potenciales del sistema, la cual puede ser finita o infinita.

- **Procesos de arribos y de servicio.** Supondremos que los tiempos entre arribos de los clientes así como del tiempo de servicio son aleatorios.

- **Capacidad del sistema.** El número máximo de clientes en el sistema puede ser finito o infinito.

- **Número de servidores.** Uno de los sistemas de espera más comunes es cuando existe un servidor. No obstante puede considerarse una cantidad finita o infinita de ellos.

- **Disciplina de servicio.** Existen varias maneras de seleccionar a los clientes que recibirán servicio las cuales determinan la disciplina de servicio.

Una manera abreviada de describir un sistema de espera la propuso Kendall en [7] y consiste en el siguiente arreglo:

$$A|S|c|K|F|d$$

donde cada una de estas letras describe:

- A = Distribución de tiempos entre arribos.

- S = Distribución del tiempo de servicio.

  Estas a su vez las podemos especificar como:

  - M : Distribución exponencial;
  - D : los tiempos entre arribos o de servicio son constantes o determinísticos;
  - G : distribución general.

- c = Número de servidores.

- K = Capacidad máxima del sistema.

- F = Número de individuos en la fuente.

- d = disciplina de servicio.

  Esta a su vez se especifica de la siguiente manera:

  - $FIFO$ (First In First Out);
  - $LIFO$ (Last In First Out);
  - $SIRO$ (Service In Random Order);
  - PRI (Priority queue discipline);
  - GD (General discipline).

En este trabajo estudiaremos colas del tipo $M|M|c|\infty|\infty|FIFO$. Es decir, sistemas de espera con c servidores donde los tiempos entre arribos y de servicio tienen distribución exponencial, la fuente y la capacidad del sistema son infinitas y la disciplina de servicio es, el primero que llega es el primero que se atiende. Cuando $K = F = \infty$ y $d = FIFO$ usaremos la notación más abreviada $M|M|c$.

Sea t_n la v.a. que representa el tiempo en el que llega el $n-$ésimo cliente, para $n = 1, 2, ...$, donde $t_0 = 0$. Supondremos que no hay arribos simultáneos, es decir,

$$0 = t_0 < t_1 < t_2 < ...$$

Entonces, definimos los tiempos entre arribos como

$$\tau_n = t_n - t_{n-1}, \quad n = 1, 2, ...$$

Por otro lado, supondremos que los servidores son idénticos y denotaremos por s a la v.a. que representa el tiempo de servicio.

Definamos los procesos estocásticos $\{N(t); t \geq 0\}$, $\{N_q(t); t \geq 0\}$ y $\{N_s(t); t \geq 0\}$ como

$$\begin{aligned} N(t) &: \text{ número de clientes en el sistema en el tiempo } t; \\ N_q(t) &: \text{ número de clientes en la cola en el tiempo } t; \\ N_s(t) &: \text{ número de clientes recibiendo servicio en el tiempo } t. \end{aligned}$$

Observemos que

$$N(t) = N_q(t) + N_s(t), \qquad \forall t \geq 0. \tag{3.1}$$

3.3. Sistema de espera M|M|1

Consideremos un sistema de espera del tipo M|M|1. Entonces los tiempos entre arribos $\tau_1, \tau_2, ...$ son v.a i.i.d. exponencialmente distribuidas con parámetro $\lambda > 0$:

$$A(t) := P\{\tau \leq t\} = 1 - e^{-\lambda t}, \quad t \geq 0,$$

y los tiempos de servicio son v.a.i.i.d. exponencialmente distribuidas con parámetro $\mu > 0$:

$$B(t) = P\{s \leq t\} = 1 - e^{-\mu t}, \quad t \geq 0.$$

Supondremos que los tiempos de servicio son independientes de los tiempos de arribo.

Recordemos (ver Proposición 1.2.1)

$$E[\tau] = \frac{1}{\lambda}.$$

Entonces el parámetro λ se puede interpretar como la tasa o intensidad promedio de arribos al sistema por unidad de tiempo. Similarmente, como

$$E[s] = \frac{1}{\mu},$$

el parámetro μ se interpreta como la tasa o intensidad promedio de servicio por servidor.

Bajo el planteamiento anterior tenemos que el número de clientes en el sistema $\{N(t) : t \geq 0\}$ es un proceso de nacimiento y muerte con parámetros constantes

$$\lambda_n = \lambda \quad \text{y} \quad \mu_n = \mu, \quad n \geq 0, \mu_0 = 0 \tag{3.2}$$

lo cual implica, que tanto el patrón de arribos al sistema como el de servicio, no dependen del número n de clientes en el sistema. En este caso la serie (2.26) se reduce a la serie geométrica:

$$\sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^j$$

la cual converge si, y sólo si $\dfrac{\lambda}{\mu} < 1$. Más aún, si $\dfrac{\lambda}{\mu} < 1$,

$$1 + \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^j = \frac{\mu}{(\mu - \lambda)}$$

y de (2.25), vemos que,

$$\pi_0 = \frac{\mu - \lambda}{\mu} = 1 - \frac{\lambda}{\mu}. \tag{3.3}$$

Por lo tanto, el proceso de nacimiento y muerte con tasas constantes (3.2) tiene una distribución estacionaria $\{\pi_j\}$ si, y sólo si $\frac{\lambda}{\mu} < 1$ y en tal caso está dada por (2.24), es decir,

$$\pi_n = \pi_0 \left(\frac{\lambda}{\mu}\right)^n = (1 - \rho)\rho^n, \quad n = 0, 1, ..., \tag{3.4}$$

donde

$$\rho = \frac{E(s)}{E(\tau)} = \frac{\lambda}{\mu}.$$

En el Apéndice A.1 se muestran algunas simulaciones de la evolución en el tiempo del tamaño de la cola considerando los casos $\rho < 1$ y $\rho > 1$.

3.3.1. Caso Estacionario

Nuestro objetivo ahora es analizar el sistema de espera cuando $\rho < 1$, es decir, cuando el sistema se encuentra en estado estacionario. En tal situación la distribución de las v.a. en (3.1) no dependerán de t y, abusando de la notación escribiremos

$$N = N_q + N_s. \tag{3.5}$$

Nótese que la distribución (3.4) es una distribución geométrica con parámetro ρ el cual representa la probabilidad de que el servidor esté ocupado, es decir,

$$\rho = P\left[N > 0\right].$$

De aquí que a ρ se le llame intensidad de tráfico.

Más aún, denotando:

$$L = E\left[N\right], \quad L_q = E\left[N_q\right], \quad L_s = E\left[N_s\right], \tag{3.6}$$

tenemos que

$$L = L_q + L_s. \tag{3.7}$$

Ahora definamos las v.a.

q : tiempo que un cliente espera en la cola antes de recibir servicio;
w : tiempo total que un cliente permanece en el sistema.

Entonces, de forma similar a (3.7), tenemos que se cumple la relación

$$w = q + s,$$

lo cual implica

$$W = W_q + W_s, \tag{3.8}$$

donde

$$W = E[w], \quad W_q = E[q], \quad W_s = E[s] = \frac{1}{\mu}. \tag{3.9}$$

Con la notación anterior tenemos que el número esperado de clientes en el sistema (tanto en la cola como en servicio) es:

$$L = \sum_{n=0}^{\infty} n\pi_n = (1-\rho)\rho \sum_{n=1}^{\infty} n\rho^{n-1}$$

o sea,

$$L = \frac{\rho}{1-\rho} \quad \text{si} \quad \rho = \frac{\lambda}{\mu} < 1. \tag{3.10}$$

A continuación calcularemos (en estado estacionario)

- el tiempo esperado en el sistema (W),
- el tiempo esperado en la cola (W_q)
- el número esperado de clientes en la cola (L_q)

Para ello usaremos las siguientes fórmulas demostradas por Little en [8]. Cabe señalar que omitiremos su demostración ya que no está al alcance del presente trabajo.

Proposición 3.3.1 *[8] En una cola en estado estacionario se cumple*

$$L = \lambda W \tag{3.11}$$

y

$$L_q = \lambda W_q \tag{3.12}$$

Intuitivamente esta fórmula nos dice que, en estado estacionario, el número esperado de clientes en el sistema (L) es el resultado del tiempo esperado en el sistema (W) multiplicado por la tasa de llegadas (λ).

Entonces, de (3.10) y (3.11), resulta que

$$W = \frac{L}{\lambda} = \frac{\rho}{(1-\rho)\lambda} = \frac{\lambda}{\mu(1-\rho)\lambda} = \frac{1}{(\mu-\lambda)}. \tag{3.13}$$

Además, de (3.8),

$$W_q = W - W_s = W - E(s) = \frac{\rho E(s)}{(1-\rho)}, \tag{3.14}$$

luego, usando (3.12)

$$L_q = \lambda W_q = \frac{\rho^2}{(1-\rho)}. \tag{3.15}$$

Ahora calcularemos las distribuciones de las v.a. que representan el tiempo total que pasa un cliente en el sistema (w), y el tiempo que un cliente espera en la cola antes de ser atendido (q). Denotaremos a estas distribuciones como $F_w(t)$ y $F_q(t)$, respectivamente, es decir,

$$F_w(t) = P\{w \leq t\}$$

y

$$F_q(t) = P\{q \leq t\}.$$

Proposición 3.3.2 *Para la cola $M|M|1$*

$$\begin{array}{llcll} (a) & F_w(t) & = & 1 - e^{-t/W}, & t > 0 \\ (b) & F_q(t) & = & 1 - \rho e^{-t/W}, & t > 0, \end{array}$$

donde $W = E(w) = \frac{E(s)}{(1-\rho)}$ y $F_w(t) = F_q(t) = 0$ para $t < 0$.

Demostración. Primero demostraremos la parte (b). Para esto hay que tener en cuenta que,

$$F_q(t) = P\{q \leq t\} = P\{q = 0\} + P\{0 < q \leq t\}. \tag{3.16}$$

Observemos que por (3.4)

$$P\{q = 0\} = P\{N = 0\} = \pi_0 = 1 - \rho. \tag{3.17}$$

Por otro lado, recordemos que $\rho = P[N > 0]$. Entonces

$$\begin{aligned} P\{0 < q \leq t\} &= P\{q > 0\} P\{q \leq t|q > 0\} \\ &= P\{N > 0\} P\{q \leq t|N > 0\} \\ &= \rho P\{q \leq t|N > 0\} \\ &= \rho \sum_{n=1}^{\infty} \frac{P\{N = n\} P\{q \leq t|N = n\}}{P\{N > 0\}} \\ &= \sum_{n=1}^{\infty} \pi_n P\{q \leq t|N = n\}, \end{aligned} \tag{3.18}$$

ya que $\pi_n = P\{N = n\} \quad \forall n \in \mathbb{N}$.

Por otro lado, si tenemos $N = n$ clientes, el próximo cliente deberá esperar n tiempos de servicio para ser atendido, y éstos están distribuidos exponencialmente con parámetro μ. En otras palabras, la distribución condicional de q dado $N = n$ se puede escribir como la distribución de una suma de v.a. i.i.d. exponencialmente distribuidas con parámetro μ, es decir, como una distribución Gama con parámetros n y μ, cuya densidad está dada por (ver Proposición 1.2.2)

$$f_q(t) = \frac{\mu^n t^{n-1} e^{-\mu t}}{(n-1)!}.$$

Por lo tanto, para $n \geq 1$,

$$P\{q \leq t \ / \ N = n\} = \int_0^t f_q(r) dr.$$

Sustituyendo esta expresión en (3.18) y por (3.4) tenemos

$$\begin{aligned}
P\{0 < q \le t\} &= \sum_{n=1}^{\infty} \pi_n P\{q \le t | N = n\} \\
&= (1-\rho)\sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!}\int_0^t r^{n-1}e^{-\mu r}dr \\
&= (1-\rho)\lambda \int_0^t e^{-\mu r}\sum_{n=1}^{\infty}\frac{(\lambda r)^{n-1}}{(n-1)!}dr \\
&= (1-\rho)\lambda \int_0^t e^{-(\mu-\lambda)r}dr \\
&= \rho(1-e^{-\mu(1-\rho)t}) = \rho(1-e^{-t/W}), \qquad (3.19)
\end{aligned}$$

donde la última igualdad se sigue de (3.13).

Sustituyendo (3.17) y (3.19) en (3.16) obtenemos

$$\begin{aligned}
F_q(t) &= 1-\rho+\rho(1-e^{-t/W}) \\
&= 1-\rho+\rho-\rho e^{-t/W} \\
&= 1-\rho e^{-t/W},
\end{aligned}$$

lo cual demuestra la parte (b).

La parte (a) se demuestra de manera análoga. En efecto, observe que

$$\begin{aligned}
F_w(t) &= \sum_{n=0}^{\infty} P\{N=n\} P\{w \le t | N = n\} \\
&= \sum_{n=0}^{\infty} \pi_n P\{w \le t | N = n\}.
\end{aligned}$$

En este caso el cliente que va llegando debe esperar $n+1$ tiempos de servicio, los cuales son v.a. i.i.d. exponencialmente con parámetro μ. Por lo tanto, usando la densidad Gama con parámetros $n+1$ y μ tenemos:

$$P\{w \le t | N = n\} = \int_0^t \frac{\mu^{n+1} r^n e^{-\mu r}}{n!}dr.$$

Luego,

$$
\begin{aligned}
F_w(t) &= \sum_{n=0}^{\infty} \pi_n \int_0^t \frac{\mu^{n+1} t^n e^{-\mu t}}{n!} \\
&= (1-\rho)\mu \int_0^t e^{-\mu r} \sum_{n=0}^{\infty} \frac{(\mu\rho r)^n}{n!} dr \\
&= (1-\rho)\mu \int_0^t e^{-\mu r} e^{\rho\mu r} dr \\
&= 1 - e^{-\mu(1-\rho)t} \\
&= 1 - e^{-t/W},
\end{aligned}
$$

donde la última desigualdad se sigue de (3.13). ■

En el Apéndice A.3 presentamos simulaciones de las v.a. w y q tomando en cuenta las distribuciones dadas en la Proposición 3.3.2. Con estas simulaciones estimaremos el tiempo promedio que pasa un cliente en el sistema y el tiempo promedio que espera para ser atendido.

3.4. Sistema de espera M|M|c

Consideraremos un sistema de espera con c servidores idénticos donde los tiempos entre arribos y de servicio son exponenciales con medias $E[\tau] = \frac{1}{\lambda}$ y $E[s] = \frac{1}{\mu}$, respectivamente.

Definamos el factor de utilización como:

$$u = \frac{\rho}{c},$$

donde $\rho = \frac{\lambda}{\mu}$. Suponiendo que el tráfico está uniformemente distribuido entre los servidores, tenemos que u representa la fracción de tiempo que cada servidor está ocupado.

De acuerdo a la descripción anterior, podemos modelar el sistema $M|M|c$ como un proceso de nacimiento y muerte con parámetros

$$
\begin{aligned}
\lambda_n &= \lambda, \quad n = 0, 1, 2, ...; \\
\mu_n &= \begin{cases} n\mu & si \quad n = 0, 1, ..., c \\ \\ c\mu & si \qquad n \geq c \end{cases}.
\end{aligned}
$$

La razón por la cual $\mu_n = n\mu$ para $n \leq c$ es que cuando existen menos de c clientes en el sistema, cada uno de ellos está siendo atendido y por lo tanto, la tasa de servicio es igual al número de clientes por μ. Si existen más de c clientes en el sistema, entonces exactamente c servidores están ocupados y por lo tanto la tasa de servicio es $c\mu$.

3.4.1. Caso estacionario

De (2.26) sabemos que el proceso tendrá una distribución estacionaria si, y solo si, la siguiente serie converge:

$$\begin{aligned} C &= 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 ... \lambda_{n-1}}{\mu_1 \mu_2 .... \mu_n} \\ &= \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \sum_{n=0}^{\infty} (\frac{\rho^c}{c!}) u^n \end{aligned}$$

Por lo tanto si $u < 1$, tendremos que

$$C = \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!(1-u)}$$

En este caso, sus probabilidades estacionarias son

$$\begin{aligned} \pi_n &= \pi_0 \frac{\rho^n}{n!} \quad \text{si} \quad n = 0, 1, ..., c, \\ &= \pi_0 \frac{\rho^n}{c! c^{n-c}} \quad \text{si} \quad n \geq c, \end{aligned}$$

con $\pi_0 = \dfrac{1}{C}$.

Ahora calcularemos las cantidades L_q (número esperado de clientes en la cola), L (número esperado de clientes en el sistema), W (tiempo esperado en el sistema) y W_q (tiempo esperado en la cola).

Proposición 3.4.1 *El número esperado de clientes en la cola $L_q = E[N_q]$ es:*

$$L_q = \pi_0 \rho^c \frac{u}{c!(1-u)^2}. \tag{3.20}$$

Demostración. Recordemos que si $u < 1$ entonces

$$\sum_{n=1}^{\infty} n u^{n-1} = \frac{1}{(1-u)^2}. \tag{3.21}$$

Entonces,

$$\begin{aligned} L_q &= \sum_{n=c}^{\infty} (n-c)\pi_n \\ &= \sum_{n=0}^{\infty} n\pi_{n+c} \\ &= (\pi_0 \frac{\rho^c}{c!}) \sum_{n=1}^{\infty} n u^n \quad \text{ya que} \;\; u = \frac{\rho}{c} \\ &= (\pi_0 \rho^c \frac{u}{c!}) \sum_{n=1}^{\infty} n u^{n-1}. \end{aligned}$$

Entonces, por (3.21) tenemos que

$$L_q = \pi_0 \rho^c \frac{u}{c!(1-u)^2},$$

lo cual demuestra la Proposición. ■

Proposición 3.4.2 *La probabilidad de que un cliente que llega al sistema tenga que esperar en la cola está dada por:*

$$P\left[N \geq c\right] = \frac{\pi_0 \rho^c}{c!(1-u)},$$

si $u < 1$.

Demostración. Esta expresión es consecuencia de las siguientes relaciones:

$$\begin{aligned}
P\left[N \geq c\right] &= \sum_{n=c}^{\infty} \pi_n \\
&= \left(\frac{\pi_0}{c!}\right) \sum_{n=c}^{\infty} \frac{\rho^n}{c^{n-c}} \\
&= \left(\frac{\pi_0 \rho^c}{c!}\right) \sum_{n=c}^{\infty} \left(\frac{\rho}{c}\right)^{n-c} \\
&= \left(\frac{\pi_0 \rho^c}{c!}\right) \sum_{n=0}^{\infty} u^n \\
&= \frac{\pi_0 \rho^c}{c!(1-u)}.
\end{aligned}$$

■

Considerando las Proposiciones 3.4.1 y 3.4.2, y las fórmulas de Little (ver Proposición 3.3.1) tenemos que

$$L = \lambda W$$

y

$$W_q = \frac{L_q}{\lambda}, \tag{3.22}$$

donde

$$W = W_q + E(s) = W_q + \frac{1}{\mu}.$$

Además, denotando

$$\mathcal{E}(c, \rho) = \frac{\pi_0 \rho^c}{c!(1-u)},$$

de (3.20) obtenemos,

$$L_q = \frac{\mathcal{E}(c, \rho) u}{1-u} \tag{3.23}$$

y de (3.22)

$$W_q = \frac{\mathcal{E}(c, \rho) u}{\lambda(1-u)} = \frac{\mathcal{E}(c, \rho) E(s)}{c(1-u)}. \tag{3.24}$$

Apéndice A

Simulaciones

A.1. Simulación del tamaño de la cola

Presentaremos simulaciones del tamaño de la cola para valores arbitrarios del factor ρ , es decir, en los casos cuando el sistema de espera sea estacionario ($\rho < 1$) o no estacionario ($\rho > 1$).

Para simular el tamaño de la cola partimos de los valores de las tasas de arribo λ y de servicio μ, así como el número de eventos n. En este caso, consideramos como evento a una llegada o salida de un cliente.

Recordemos que si $X \sim \exp(\lambda)$ entonces

$$X = -\frac{1}{\lambda} \log u, \quad \text{donde } u \sim u(0,1)$$

Entonces, para simular los tiempos entre arribos y de servicio definimos las variables

$$llega = \frac{-1}{\lambda} \log u$$

$$sale = \frac{-1}{\mu} \log u_2$$

donde u y u_2 son números aleatorios entre 0 y 1.

Ahora definimos las siguientes variables:

t: tiempo total transcurrido

clientes: clientes en el sistema
na: número de clientes llegados

nd:número de clientes atendidos

Ahora comparamos el tiempo de llegada(*llega*) con el de servicio(*sale*), y tenemos los siguientes 2 casos:

1.Si $llega \leq sale$

$t = llega$

$na = na + 1$ (ya que hay una llegada más)

$clientes = clientes + 1$ (hay un cliente más)

$llega = \frac{-1}{\lambda} \log u$ (generamos el siguiente tiempo entre llegadas)

si $clientes = 1, sale = t + \left(\frac{-1}{\mu}\right) * \log u_2$ (como solo hay un cliente, hay que generar su tiempo de servicio y sumárselo al tiempo actual)

2. Si $llega > sale$

$t = sale$

$clientes = clientes - 1$ (un cliente menos en el sistema)

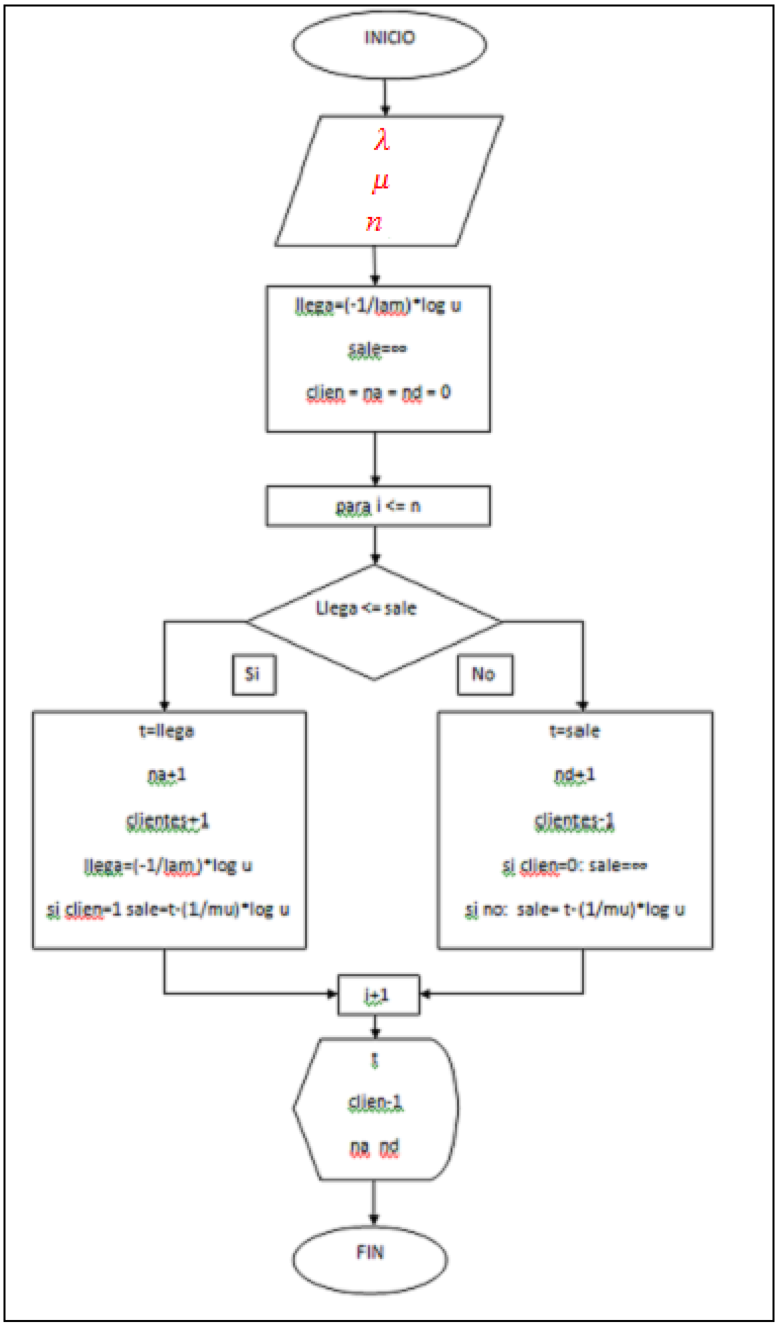
$nd = nd + 1$ (ha ocurrido una salida al tiempo t)

si $clientes = 0 \qquad sale = \infty$

si no $sale = t + \left(\frac{-1}{\mu}\right) * \log u_2$ (generamos el tiempo del siguiente servicio y lo sumamos al tiempo actual)

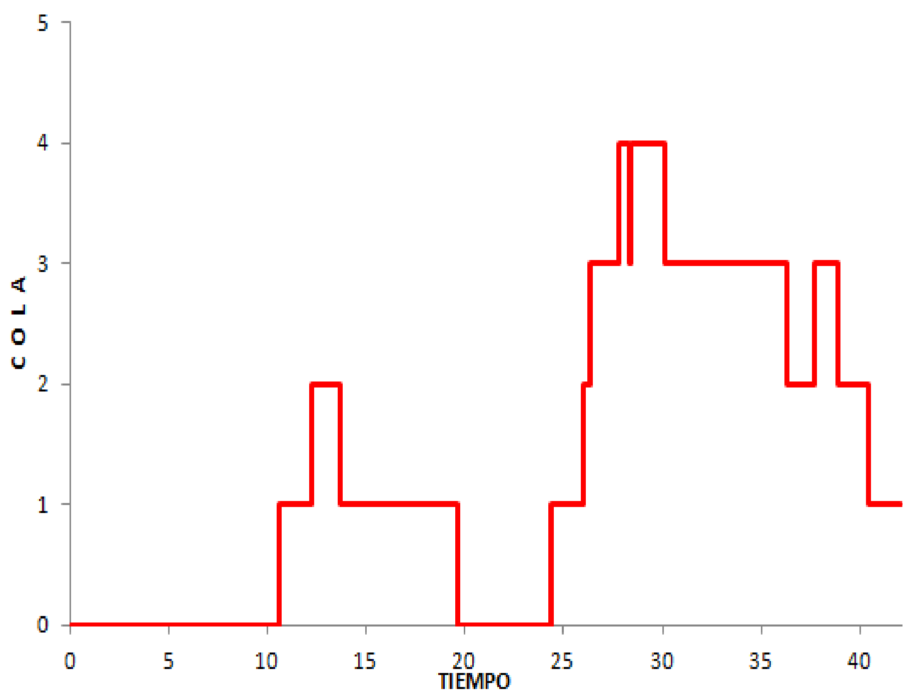
La gráfica que proporciona el programa representa el número de clientes en la cola ($clientes - 1$) al tiempo t.

A.1.1. Diagrama de flujo



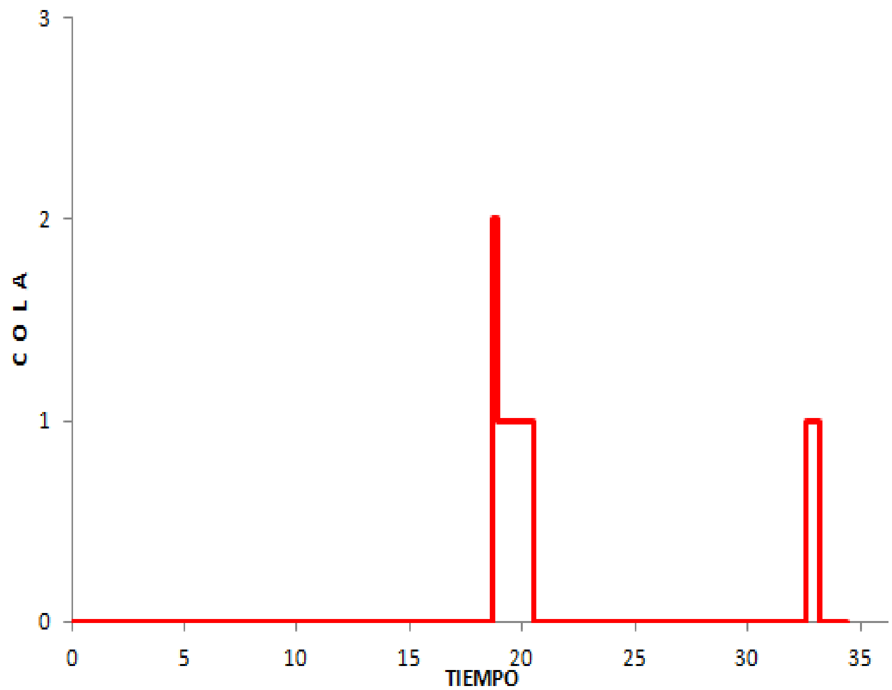
A.1.2. Gráficas

A continuación mostramos algunas gráficas de la evolución en el tiempo del tamaño de la cola variando los párametros λ y μ.



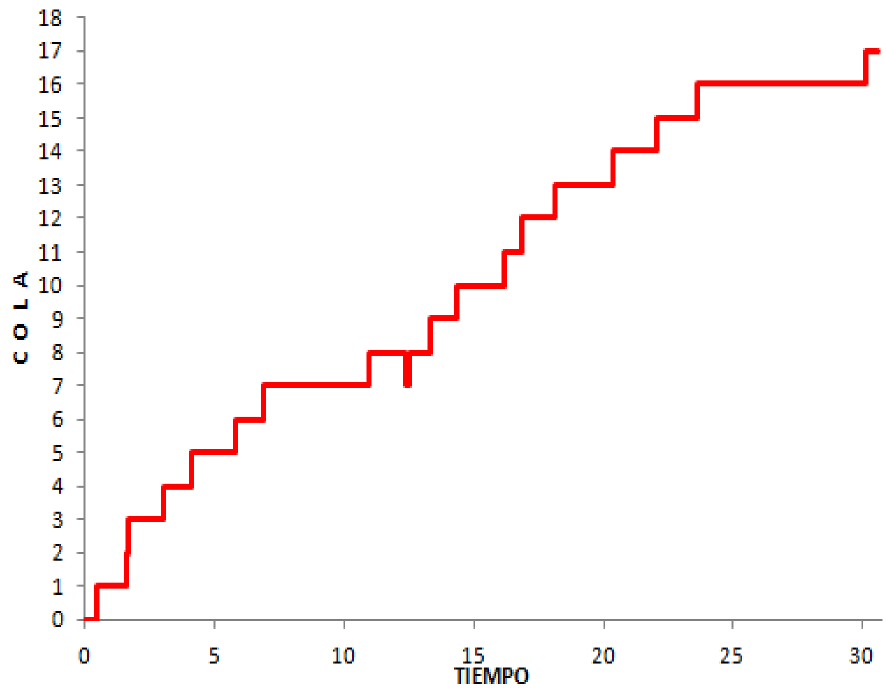
$$\rho = \frac{4}{5}$$

Como ρ es cercana a 1 vemos que la cola tiene un comportamiento regular, no crece ilímitadamente, pero tampoco permanece mucho tiempo vacía.



$$\rho = \frac{1}{5}$$

Como la ρ es muy pequeña vemos que la cola casi siempre esta vacía, es decir, los clientes son atendidos de inmediato.



$$\rho = \frac{14}{5}$$

Con $\rho > 1$ vemos que la cola crece conforme avanza el tiempo.

A.2. Tiempo de espera del cliente

Supondremos que existe un número determinado de clientes en la cola al momento de la llegada del cliente. En este caso simularemos el tiempo de espera del cliente para ser servido. En vez de simular eventos, le preguntaremos al usuario cuántos clientes hay antes que él, así tenemos un uso práctico para el programa

En el ejemplo el número de personas en la cola era 10 cuando llegó el usuario, corremos las simulaciones hasta que el número de clientes atendidos séa igual a 10.

Tiempo	arribos	servidos	cola
0.8137	0	1	9
3.4363	0	2	8
5.0417	0	3	7
5.0885	1	3	8
8.9306	2	3	9
12.9939	3	3	10
13.2641	3	4	9
14.8829	4	4	10
19.9114	4	5	9
20.5284	4	6	8
20.8867	5	6	9
20.9054	6	6	10
22.716	6	7	9
22.946	6	8	8
25.0791	6	9	7
26.2134	7	9	8
26.3373	7	10	7

Así obtenemos que se tardaría 26,3373 minutos en ser atendido y además quedarían 7 clientes en la cola detrás de él.

A.3. Estimación en estado estacionario

En esta parte estimaremos los tiempos promedios en el sistema y en la cola, así como el número de clientes, en el caso que $\rho < 1$. Para esto simularemos las variables aleatorias

$$
\begin{aligned}
w &: \text{Tiempo total en el sistema} \\
q &: \text{Tiempo en la cola} \\
N &: \text{clientes en la cola}
\end{aligned}
$$

considerando la Proposición 3.3.2 y que N tiene distribución geométrica con parámetro ρ, las fórmulas son:

$$
\begin{aligned}
w_i &= (-W) * \log(1-u) \\
q_i &= (-W) * \log\frac{(1-u)}{\rho} \\
n_i &= \left\| \frac{\log(1-u)}{\log\rho} \right\|
\end{aligned}
$$

donde $\|X\|$ representa la parte entera de X.

Resultados:

	N=100	RO=1/2
	Promedios	Esperados
Tiempo en el sistema	2.0681	2
Tiempo en la cola	1.0358	1
Clientes en la cola	1.0501	0.5

	N=500	RO=1/5
	Promedios	Esperados
Tiempo en el sistema	1.272	1.25
Tiempo en la cola	0.267	0.25
Clientes en la cola	0.264	0.05

	N=500	RO=4/5
	Promedios	Esperados
Tiempo en el sistema	19.1583	20
Tiempo en la cola	15.121	16
Clientes en la cola	3.818	3.2

	N=500	RO=99/100
	Promedios	Esperados
Tiempo en el sistema	9737.354	9899.9795
Tiempo en la cola	9638.41	9800.9502
Clientes en la cola	97.372	98.0095

Bibliografía

[1] R.N. Bhattacharya, E.c. Waymire (1990) *Stochastic Processes with Applications*, Wiley, New York.

[2] L. Breiman (1968). Probability, Addison-Wesley.

[3] A.K. Erlang (1918) *Solution of some problems in theory of probabilities of significance in automatic telephone exchanges.* Post Office Electrical Enginners Journal, 10, pp. 189-197.

[4] R.M. Feldman, C. Valdez-Flores (2010) Applied Probability and Stochastic Processes, 2nd ed., Springer.

[5] I. I. Gikhman, A.V. Skorokhod (1969) *Introduction to the theory of random processes*, W.B. Saunders Co.

[6] O. Hernández-Lerma (1981) Procesos Estocásticos: Introducción a la Teoría de Colas. 2do Coloquio del Departamento de Matemáticas, Cinvestav.

[7] D.G. Kendall (1953) *Stochastic processes occuring in the theory of queues and their analysis by the method of imbedded Markov chains.* Annals of Mathematical Statistics, 24, pp. 338-354.

[8] J.D.C. Little (1961) *A proof of the queueing formula* $L = \lambda W$. Operations Research, 9, pp. 338-387.

[9] Parzen, th. 2a, p.173