Matemáticas

Modelación Estocástica

Séptimo Semestre

Unidad 2. Teoría de colas

Clave

05144737

Universidad Abierta y a Distancia de México

DCEIT

Índice

Unidad 2. Teoría de colas

Presentación de la unidad

En la presente unidad se realizará el estudio matemático de las líneas de espera (o colas), las cuales pueden representar situaciones muy comunes en la vida cotidiana.

- Cuando se pagan las compras en la caja de un centro comercial, debido al tiempo que tarda el cajero en atender, se forma mucha gente en espera, entonces se tiene un ejemplo de una línea de espera.

  En otros campos de conocimiento:
- En informática una línea de espera puede ser útil para modelar, por ejemplo, el tráfico en redes.
- En la ingeniería puede ayudar a modelar una línea de proceso de envasado de agua, en el armado de vehículos, etcétera.

En el desarrollo de estos temas aplicarás muchos de los conocimientos previamente adquiridos sobre procesos estocásticos, teoría de probabilidades y la inferencia estadística, la cual es de suma importancia en el desarrollo de la teoría de la decisión.

El contenido de esta unidad se presentará a través de tres subtemas, de tal forma que en el primero se proporcionarán los elementos fundamentales para abordar la teoría de colas; en el segundo, los modelos de colas para procesos markovianos y, en el tercero, los modelos de colas no markovianos.

Considera que los elementos importantes se resaltan empleando un fondo de color y, por tanto, son estos conocimientos en los que deberás enfocar tus esfuerzos por desarrollar comprensión y aprendizaje con la finalidad de que los emplees para generar un nivel de conocimientos óptimo acerca de la actual materia de estudio.

Competencia específica

Aplicar los procesos estocásticos para modelar y resolver problemas relacionados con la teoría de colas utilizando sus reglas generales.

Logros

- Identificar las características que presentan los modelos de colas para procesos markovianos y no markovianos, además de su clasificación.
- Resolver problemas relacionados con la teoría de colas.

2.1. Introducción

Seguro te has formado en muchas ocasiones esperando que se te brinde un servicio, por ejemplo, cuando compras tortillas o llamas a un centro de atención a clientes para reportar el fallo de tu teléfono celular. En virtud de lo anterior, ya has tenido contacto con el tema que se tratará, pues dichas condiciones corresponden a situaciones en las que se muestran líneas de espera.

Una línea de espera se puede modelar como un proceso estocástico, dado que el tiempo en que se brinda el servicio y la llegada de “clientes” que lo reciben presenta aleatoriedad. Claramente se puede notar que la variable aleatoria de este tipo de procesos estocásticos es justamente el número de transacciones en el sistema en un momento determinado, la cual, por consiguiente, toma valores en el conjunto $\{0, 1, 2, 3, ..., N\}$, y además cada uno de ellos tiene asociada una probabilidad de ocurrencia, que cobra relevancia en la teoría de decisión, por ello es necesario que conozcas acerca de éste.

En la presente unidad se realizará el modelado de este tipo de situaciones y se tratarán de resolver algunos problemas sobre el tema. No olvides apoyarte con tu docente en línea cuando se te presente algún tipo de problemática para que no tengas “lagunas” respecto a los temas de estudio.

Recuerda que esta materia es totalmente práctica, los temas que en ella se emplearán ya fueron analizados en materias como Probabilidad I, II y III, Procesos estocásticos y Estadística I y II. Por lo anterior, es imperativo que si no recuerdas alguno de ellos te remitas a su revisión y análisis (incluyendo las demostraciones de sus reglas y propiedades). De hecho **es conveniente que revises nuevamente el tema Procesos markovianos, de la materia Procesos estocásticos**, pues se empleará en esta unidad, en el entendido de que, como ya fue tratado, no se desarrollará nuevamente.

Cabe aclarar que durante la unidad se emplearán de manera indistinta las expresiones “línea de espera” y “cola” o “fila”.

2.1.1. Fundamentos

En la sección introductoria se proporcionan dos ejemplos de colas: la espera en una fila para comprar tortillas y la de atención en una llamada a un centro de atención a clientes. Partiendo de dichos ejemplos, se puede ver que una línea de espera se presenta cuando la demanda de un servicio sobrepasa a la capacidad de proporcionarlo.

Pero te preguntarás, ¿por qué es necesario estudiar las líneas de espera?

La respuesta es simple. La idea de realizar un análisis de este tipo de situaciones tiene como objetivo implementar medidas para mejorar los procesos en que intervienen, y con ello reducir las pérdidas que presenta una determinada empresa.

Algunos elementos que se pueden obtener al analizar una línea de espera son:

- Promedio de llegadas de clientes.
- El tiempo de espera promedio.
- La longitud promedio de la cola, entre otros.

Asimismo, tomando como ejemplo la situación presentada en los bancos, donde regularmente se debe “hacer cola” con la finalidad de ser atendido para realizar alguna transacción (ya sea pago, depósito o retiro), se presenta el costo que deben considerar los socios (dueños del banco) en el pago del personal que atenderá a los clientes, la energía necesaria para que funcionen los equipos que registrarán las transacciones, su *software*, sistemas de seguridad, etcétera; además, existen clientes que, debido al largo tiempo de espera para ser atendidos, optan por llevarse su recurso a otro banco, lo cual representaría una pérdida. Por otra parte, en términos económicos, se tiene la pérdida que representa el tiempo empleado al esperar ser atendidos.

Hay que mencionar que al analizar un sistema de colas como el del caso anterior, se tiene como propósito minimizar la suma de los costos por ofrecer el servicio y por esperar, o bien, evaluar el impacto que representan las posibles alternativas de modificación de la capacidad del sistema en el costo total de éste.

De manera general se puede decir que al analizar un sistema de colas se puede establecer el nivel óptimo de servicio necesario para minimizar el costo total del sistema. Dicho nivel de servicio se puede determinar por la cantidad de servidores necesarios para brindar el servicio, o bien, por la velocidad que presentan.

Debido a que el costo total de un sistema de colas debe minimizarse, además éste se conforma por el costo del servicio y el que causa la espera en la fila, entonces se puede expresar esta situación a través de la siguiente ecuación:

$$Min\, C_t = C_e \cdot S + C_q \cdot L_q$$

Donde

$S = 1, 2, 3, \ldots$, es el número de servidores que proporcionan el servicio.

$E(t)$ es el tiempo promedio en que se brinda el servicio a un cliente.

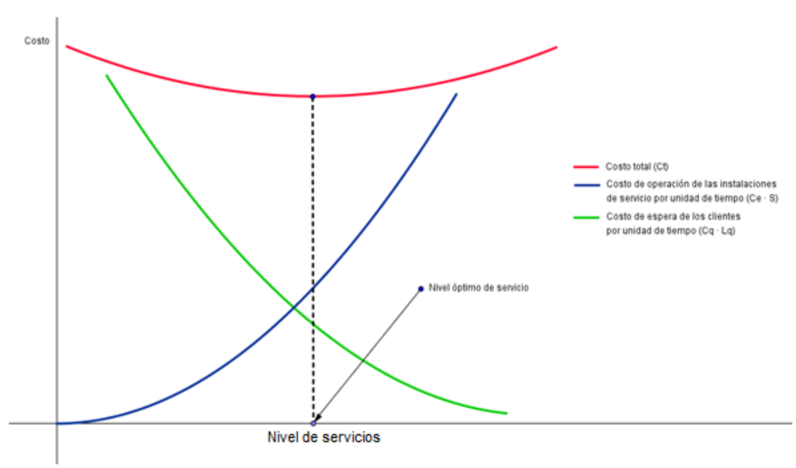
$L_q = f\{S, E(t), \ldots\}$ es el número promedio de clientes que se encuentran en la fila.

C_e es el costo promedio de servicio por cliente por unidad de tiempo.

C_q es el costo promedio de espera por cliente por unidad de tiempo.

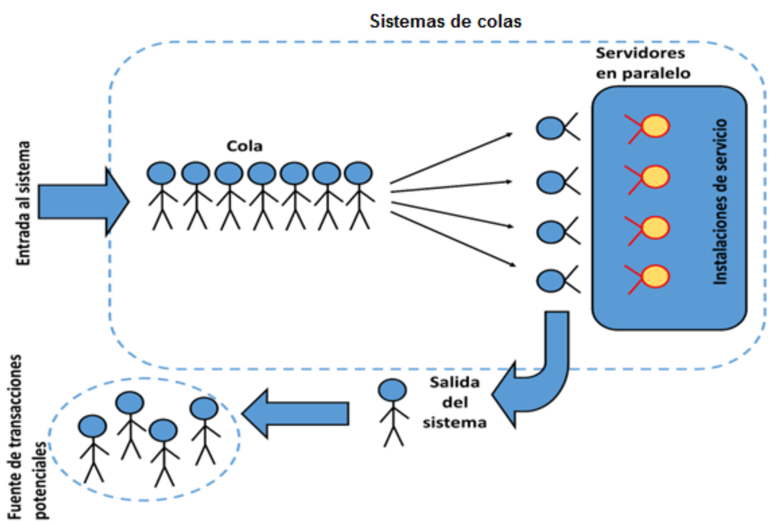
C_t es el costo total promedio del sistema de colas por unidad de tiempo.

A continuación, se muestra una gráfica típica de este tipo de modelos de costos:



En un **proceso básico de colas** existe un cliente que ingresa a un sistema requiriendo que le brinden un servicio, el cual será atendido por un servidor (o varios). Si no hay servidores disponibles, dichos clientes se deben formar para ser atendidos bajo una regla específica, la llamada disciplina de la cola. Una vez formados, puede ser que se retiren debido a diversos factores, por ejemplo, el aburrimiento, o bien son atendidos, después de lo cual salen del sistema convirtiéndose en clientes potenciales (o transacción potencial), pues pueden volver a requerir servicios. Claramente la situación descrita en este párrafo es una situación básica, pero debes tener presente que un sistema de colas puede ser mucho más complejo de lo citado aquí.

La siguiente figura muestra la estructura general de un sistema de colas.



Para efectos de hacer más explícito el desarrollo del tema, se describe cada una de las características citadas en el modelo básico.

Población potencial

Se define como el conjunto formado por los clientes potenciales, es decir, aquellos que requieran de algún servicio que proporcione el sistema. También se le llama "fuente de entrada" o "transacción potencial". La población potencial tiene dos características: su tamaño, el cual está determinado por el número de elementos que tiene dicho conjunto, que puede ser finito o infinito, y la tasa de entrada promedio, establecida por una distribución de probabilidades que representa el comportamiento probabilístico del tiempo entre llegadas.

Cliente

Cuando un elemento de la población potencial busca un servicio proporcionado por el sistema, se convierte en **cliente**.

Cola o fila

Está formada por un grupo de clientes en espera de atención o, dicho de otra forma, un grupo de transacciones en espera de atención, las cuales se agrupan debido a que todos los servidores se encuentran ocupados. Claramente los elementos de una fila presentan un cierto orden, por ello, aunque se piense que sólo se pueden colocar en línea, esto no es necesario, por ejemplo, en algunos bancos se pide que un cliente tome un número, pero que se siente en cualquier lugar de un grupo de sillas sin orden específico. Las características que tienen las filas son:

- **Capacidad de la cola:** número máximo de transacciones que puede aceptar una cola (puede ser finito o infinito).
- **Orden de la cola:** se refiere al orden en que se seleccionan los elementos de una cola para brindarles atención. También se le llama **disciplina de la cola**. Por citar algunos:
  - FCFS: consiste en atender primero al cliente que llega primero.
  - LCFS: consiste en atender primero al cliente que ha llegado al último.
  - SIRO: en este caso se seleccionan los clientes aleatoriamente.
  - PR: el orden se establece según prioridades.
  - GD: orden general.
- **Modo en que se deja una cola:** puede ser porque ya se recibió atención, por aburrimiento, otros motivos personales, etcétera.

Mecanismo de servicio

Se conforma por una o más instalaciones de servicio, las cuales se componen de uno o más canales paralelos de servicio que se denominan **servidores**, y que cuentan con dos características:

1. La cantidad asignada por cada fila del sistema.

2. La velocidad de servicio, que queda determinada por la distribución de probabilidades del tiempo de atención a los clientes.

A continuación, se presentan las ecuaciones para cada uno de los elementos de un sistema de colas, las cuales se emplearán en la solución de problemas.

Formulario 2.1.1.1.

Utilización promedio del servicio (ρ)

$$\rho = \frac{\overline{\lambda}}{S\mu}$$

S : número de servidores.

μ : capacidad del servidor.

$\overline{\lambda}$: tasa promedio de llegadas.

Tasa promedio de llegadas $(\overline{\lambda})$

$$\overline{\lambda} = \sum_{n=0}^{N} \lambda_n P_n$$

λ_n : flujo de clientes que entran cuando hay n clientes en el sistema.

P_n : probabilidad en estado estable de que existan n clientes en el sistema.

N : número máximo de clientes permitido en el sistema.

Número promedio de clientes en el sistema de colas $\left(L\right)$

$$L=\sum_{n=o}^{N} n\,P_n = L_q + S\rho$$

n : número de clientes en el sistema.

P_n : probabilidad en estado estable de que existan n clientes en el sistema.

N : número máximo de clientes permitido en el sistema.

L_q : número promedio de clientes en la fila.

S : número de servidores.

ρ : utilización promedio del servicio.

Número promedio de clientes en la cola $\left(L_q\right)$

$$L_q=\sum_{n=S}^{N}\left(n-S\right)P_n$$

n : número de clientes en el sistema.

N : número máximo de clientes permitido en el sistema.

S : número de servidores.

P_n : probabilidad en estado estable de que existan n clientes en el sistema.

Tiempo promedio de espera en el sistema $\left(W\right)$

$$W=\frac{L}{\overline{\lambda}}=W_q+E\left(t\right)$$

L : número promedio de clientes en el sistema de colas.

$\overline{\lambda}$: tasa promedio de llegadas.

W_q : tiempo promedio de espera en la fila.

$E\left(t\right)$: tiempo promedio de proceso por cliente.

Tiempo promedio de espera en la fila $\left(W_q\right)$

$W_q = \frac{L_q}{\overline{\lambda}}$

L_q : número promedio de clientes en la cola.

$\overline{\lambda}$: tasa promedio de llegadas.

Coeficiente cuadrado de variación

$C_a^2 = \frac{V(a)}{\left[E(a)\right]^2}$ (Coeficiente cuadrado de variación para el tiempo entre llegadas)

$V(a)$: variancia del tiempo entre llegadas.

$E(a)$: tiempo promedio entre llegadas.

$C_S^2 = \frac{V(t)}{\left[E(t)\right]^2}$ (Coeficiente cuadrado de variación para el tiempo de servicio)

$V(t)$: variancia del tiempo de proceso.

$E(t)$: tiempo promedio de proceso por cliente.

$C_p^2 = C_a^2\left(1-\rho^2\right) + C_S^2\rho^2$ (Coeficiente cuadrado de variación para el tiempo entre salidas del servicio)

C_a^2: coeficiente cuadrado de variación para el tiempo entre llegadas.

C_p^2: coeficiente cuadrado de variación del flujo de clientes que salen del sistema.

C_S^2: coeficiente cuadrado de variación para el tiempo de servicio.

ρ : utilización promedio del servicio.

Nota: a las expresiones $L = \lambda W$ y $L_q = \lambda W_q$, se les conoce como las ecuaciones de Little.

2.1.2. Clasificación de Kendall-Lee

Actualmente existe una clasificación compuesta por seis características de las colas, se estructura de la siguiente forma (a/b/c) (d/e/f), considerando lo siguiente:

Parámetro	Definición del parámetro
a	Es la distribución de probabilidades del tiempo entre llegadas.
b	Distribución de probabilidades del tiempo de servicio.
c	Es el número de servidores.
d	Es el orden de atención de los clientes.
e	Es el número máximo de clientes que soporta el sistema.
f	Es el número de clientes potenciales del sistema de líneas de espera.

Para sustituir los parámetros a y b se emplean los siguientes símbolos:

D: constante o determinista.

E_k: distribución Erlang con parámetro k.

G: cualquier tipo de distribución.

GI: distribución general independiente.

H: distribución hiperexponencial.

M: distribución exponencial.

Asimismo, para sustituir el parámetro d se emplea lo siguiente:

FCFS: se nombra “primeras entradas, primeros servicios” y, como ya se mencionó, consiste en atender primero al cliente que llega primero.

LCFS: se denomina “últimas entradas, primeros servicios”, consiste en atender primero al cliente que ha llegado a lo último.

SIRO: en este caso se seleccionan los clientes aleatoriamente para brindarles atención.

PR: el orden se establece según prioridades.

GD: orden general.

La clasificación anteriormente descrita debe su nombre a sus autores, Kendall y Lee, quienes la propusieron en 1953.

A continuación, se presentan los parámetros de algunas distribuciones de probabilidades para el tiempo de servicio en un sistema de colas con un servidor, debido a que serán primordiales más adelante.

Distribución	Media E(t)	Variancia V(t)
Constante(K)	K	0
Exponencial($\frac{1}{\mu}$)	$\frac{1}{\mu}$	$\frac{1}{\mu^2}$
Normal($\frac{1}{\mu}$, σ^2)	$\frac{1}{\mu}$	σ^2
Erlang($\frac{1}{\mu}$, k)	$\frac{1}{\mu}$	$\frac{1}{k\mu^2}$
Weibull(γ , β , α)	$\gamma+\alpha\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]$	$\alpha^2\left[\Gamma\left(1+\frac{2}{\beta}\right)-\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2\right]$

Ejemplo 2.1.2.1.

Se tiene un sistema de colas clasificado según Kendall-Lee como (M/D/5) (PR/30/30). Describe el modelo.

Solución:

En este caso se tiene un sistema de colas en el que la distribución entre llegadas es exponencial, el tiempo de servicio es constante y se cuenta con cinco servidores en paralelo, atendiendo según prioridades, con un número máximo de 30 clientes soportados por el sistema; además de un total de 30 clientes potenciales. Cuando los clientes ingresan al sistema y encuentran todos los servidores ocupados, se forman en una fila común.

Ejemplo 2.1.2.2.

En un establecimiento existen tres cajeros proporcionando un determinado servicio con velocidad constante. Se atiende a un máximo de 50 clientes en orden de primeras entradas, primeras salidas. El sistema tiene infinitos clientes potenciales y el tiempo de llegadas sigue una distribución exponencial, además de que cuando los clientes llegan al sistema y encuentran todos los servidores ocupados, se forman en una fila común. Clasifica este sistema según Kendall-Lee.

Solución:

(M/D/3) (FCFS/50/∞)

Cabe mencionar que en un modelo de negocios regularmente las llegadas y los tiempos de servicio no son determinísticos, es decir, no son constantes, aunque sí existen sistemas de colas en los que los tiempos de servicio y las llegadas son constantes, por ejemplo, en una fábrica, donde si un equipo realiza cortes de cristales para crear parabrisas de auto, entonces los cristales llegan de manera constante en un tiempo determinado para ser cortados, y el corte (servicio) también se realiza en tiempos iguales.

Para efecto de que comprendas mejor el tema, se te recomienda que revises, anotes y clasifiques los sistemas de colas que observas en un día cualquiera.

2.2. Modelos de colas para procesos markovianos

Los procesos de Markov se emplean para estudiar el comportamiento de los sistemas de colas. Como recordarás, un proceso de Markov es un proceso estocástico en el cual la ocurrencia de un estado futuro depende únicamente del estado inmediatamente anterior, sin importar ningún otro previo. En la teoría de colas un "estado" se refiere al número de operaciones realizadas al interior del sistema en un momento determinado.

Es necesario que, en caso de que no recuerdes el tema Procesos de Markov, lo revises, pues aquí no se explicará debido a que ya lo revisaste en la asignatura Procesos estocásticos.

2.2.1. Modelos (M/M/c) (d/N/f)

Estos modelos tienen la propiedad de que en un cierto momento (o estado) se interrumpen las entradas al sistema, es decir, tienen capacidad finita. La solución de problemas que tienen que ver con este tipo de modelos se puede realizar directamente a través de las fórmulas 2.1.1.1.

Es posible estructurar el método de solución para estos casos de la siguiente manera:

1. Se realiza el diagrama de probabilidades de transición que muestre todos los posibles estados del sistema.
2. Se genera la matriz de transición o matriz de probabilidades de un paso.
3. Se encuentran las ecuaciones de balance usando $P_j = \sum_{j=0}^{N} P_i \cdot p_{ij}$ y $\sum_{j=0}^{N} P_j = 1$, donde P_j representa las probabilidades de estado estacionario del sistema, y p_{ij} las probabilidades de transición de un paso.

4. Se resuelve el sistema de ecuaciones y se obtienen las probabilidades de estado estacionario.

5. Se calcula la tasa efectiva de entrada de clientes al sistema.

6. Usando el formulario 2.1.1.1., se calcula L, L_q, W, W_q y ρ.

7. Se debe evaluar el rendimiento del sistema analizado en términos de costo, siempre y cuando esto sea aplicable.

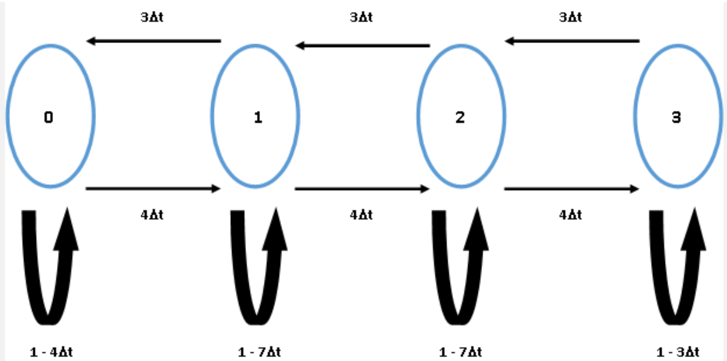
Ejemplo:

El área de atención psicopedagógica de una pequeña institución educativa cuenta con un psicólogo que atiende a un alumno en 20 minutos según una distribución exponencial. Su sala de espera acepta a dos jóvenes como máximo, y los estudiantes llegan en busca de atención con un tasa de cuatro por hora con distribución Poisson. Si algún chico necesita ayuda y encuentra la sala de espera llena, se retira sin ingresar. Si la escuela tiene 235 alumnos, determina el promedio de la utilización del servicio, el promedio de alumnos en el sistema en espera de apoyo, el número promedio de alumnos en la fila y el tiempo promedio de espera en la fila.

Solución:

En este caso el sistema se clasifica como (M/M/1)(FCFS/3/235).

El diagrama de probabilidades de transición se muestra a continuación:



De donde la matriz de un paso correspondiente es:

$$
\begin{array}{c c}
 & \begin{array}{cccc} \quad 0 \quad & \quad 1 \quad & \quad 2 \quad & \quad 3 \quad \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 1-4\Delta t & 4\Delta t & 0 & 0 \\ 3\Delta t & 1-7\Delta t & 4\Delta t & 0 \\ 0 & 3\Delta t & 1-7\Delta t & 4\Delta t \\ 0 & 0 & 3\Delta t & 1-3\Delta t \end{bmatrix}
\end{array}
$$

De aquí que se tenga el siguiente sistema de ecuaciones:

$$
\begin{cases}
P_0 = (1-4\Delta t)P_0 + 3\Delta t\, P_1 \\
P_1 = 4\Delta t\, P_0 + (1-7\Delta t)P_1 + 3\Delta t\, P_2 \\
P_2 = 4\Delta t\, P_1 + (1-7\Delta t)P_2 + 3\Delta t\, P_3 \\
P_3 = 4\Delta t\, P_2 + (1-3\Delta t)P_3 \\
1 = P_0 + P_1 + P_2 + P_3
\end{cases}
$$

Cuyas soluciones son:

$$P_1 = \frac{\lambda_0}{\mu_1} P_0 = \frac{4}{3} P_0$$

$$P_2 = \frac{\lambda_0}{\mu_1} \cdot \frac{\lambda_1}{\mu_2} P_0 = \left(\frac{4}{3}\right)^2 P_0$$

$$P_3 = \frac{\lambda_0}{\mu_1} \cdot \frac{\lambda_1}{\mu_2} \cdot \frac{\lambda_2}{\mu_3} P_0 = \left(\frac{4}{3}\right)^3 P_0$$

Y además:

$$P_0 = \left[1 + \frac{4}{3} + \left(\frac{4}{3}\right)^2 + \left(\frac{4}{3}\right)^3\right]^{-1} \approx 0.1543$$

Y así:

$$P_1 = 0.2057$$

$$P_2 = 0.2743$$

$$P_3 = 0.3657$$

Por lo que la tasa de entrada promedio será:

$\overline{\lambda} = \sum_{n=0}^{3} \lambda_n P_n = 4(0.1543) + 4(0.2057) + 4(0.2743) + 0(0.3657) = 2.5372$ alumnos/hora.

Además, el promedio de la utilización del servicio es:

$$\rho = \frac{\overline{\lambda}}{S\mu} = \frac{2.5372}{1(3)} = 0.5487$$

El promedio de clientes en el sistema en espera de apoyo es:

$L = \sum_{n=0}^{3} n\, P_n = 0(0.1543) + 1(0.2057) + 2(0.2743) + 3(0.3657) = 1.8514$ alumnos.

Y el número promedio de alumnos en la fila es:

$$L_q = \sum_{n=1}^{3} (n-1) P_n = 0(0.2057) + 1(0.2743) + 2(0.3657) = 1.0057 \text{ alumnos.}$$

Asimismo, el tiempo promedio de espera en la fila será:

$$W_q = \frac{L_q}{\bar{\lambda}} = \frac{1.0057}{2.5372} = 0.3964 \text{ horas.}$$

2.2.2. Modelos (M/M/c) (d/∞/∞)

Estos modelos se presentan cuando el sistema de colas tiene capacidad y fuente de transacciones potenciales infinitas. En ellos se tiene que la distribución de probabilidad de estado estable no presenta límite superior considerando la capacidad del sistema, por ello es necesario que la tasa promedio de entrada sea inferior estrictamente a la capacidad promedio de servicio, con lo que se aseguraría así la convergencia del sistema a un estado estable. El método de solución para este tipo de modelos es el siguiente:

1. Se realiza el diagrama de probabilidades de transición para el grupo de estados que mejor describa al sistema.
2. Se genera la matriz de transición o matriz de probabilidades de un paso.
3. Se encuentran las ecuaciones de balance usando $P_j = \sum_{j=0}^{N} P_i \cdot p_{ij}$ y $\sum_{j=0}^{N} P_j = 1$, donde P_j representan las probabilidades de estado estacionario del sistema y p_{ij} son las probabilidades de transición de un paso.
4. Se modelan algebraicamente las probabilidades de estado estacionario P_n.

5. Se determina la utilización de los servidores y se corrobora que sea menor que 1.
6. Empleando series geométricas, se calcula la probabilidad de que el sistema se encuentre vacío.
7. Usando el formulario 2.1.1.1., se calcula L, L_q, W y W_q.
8. Se evalúa el rendimiento del sistema analizado en términos de costo, siempre y cuando esto aplique.

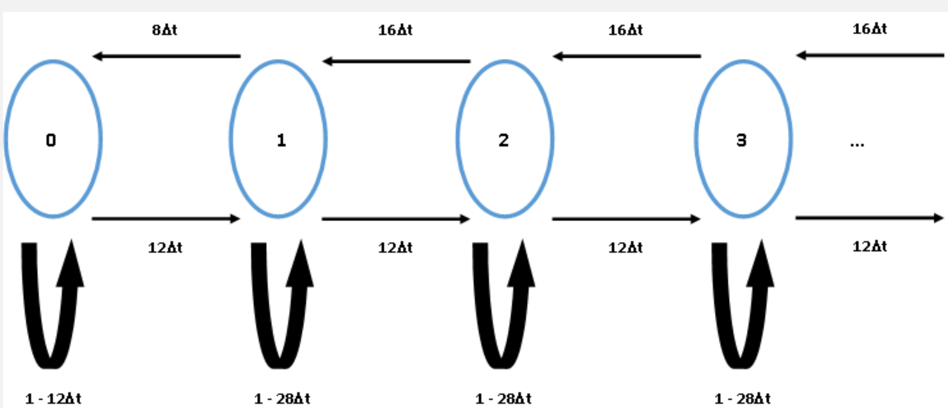
Ejemplo

En un sistema de colas con dos servidores, las llegadas siguen una distribución Poisson con media de 12 usuarios por hora, mientras que el tiempo de servicio es exponencial de 60 minutos por cliente. Si uno llega al sistema y se encuentra con que todos los servidores están ocupados, se forma en una fila común hasta ser atendido bajo el criterio FCFS. Calcula el número promedio de clientes en la fila esperando atención, además del tiempo promedio de permanencia en la fila, considerando turnos de servicio de ocho horas.

Solución:

Claramente se tiene un sistema (M/M/2)(FCFS/∞/∞).

El diagrama de probabilidades de transición para el grupo de estados que mejor describe al sistema es:



De donde la matriz de un paso correspondiente es:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ . \\ . \\ . \end{array} \begin{array}{c} \begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & \ldots & \ldots & \ldots \end{array} \\ \begin{bmatrix} 1-12\Delta t & 12\Delta t & 0 & 0 & 0 & \ldots & \ldots & \ldots \\ 8\Delta t & 1-28\Delta t & 12\Delta t & 0 & 0 & \ldots & \ldots & \ldots \\ 0 & 16\Delta t & 1-28\Delta t & 12\Delta t & 0 & \ldots & \ldots & \ldots \\ 0 & 0 & 16\Delta t & 1-28\Delta t & 12\Delta t & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 16\Delta t & 1-28\Delta t & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix} \end{array}$$

Y por tanto, se tiene el sistema:

$$\begin{cases} P_0 = \left(1-12\Delta t\right) P_0 + 8\Delta t\, P_1 \\ P_1 = 8\Delta t\, P_0 + \left(1-28\Delta t\right) P_1 + 12\Delta t\, P_2 \\ P_2 = 16\Delta t\, P_1 + \left(1-28\Delta t\right) P_2 + 12\Delta t\, P_3 \\ P_3 = 16\Delta t\, P_2 + \left(1-28\Delta t\right) P_3 + 12\Delta t\, P_4 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ 1 = P_0 + P_1 + P_2 + P_3 + P_4 + \ldots \end{cases}$$

Cuyas soluciones son:

$$P_1 = \frac{12}{8} P_0$$

$$P_2 = \left(\frac{12}{8}\right)\left(\frac{12}{16}\right) P_0$$

$$P_3 = \left(\frac{12}{8}\right)\left(\frac{12}{16}\right)^2 P_0$$

$$P_4 = \left(\frac{12}{8}\right)\left(\frac{12}{16}\right)^3 P_0$$

…

$$P_n = \left(\frac{12}{8}\right)\left(\frac{12}{16}\right)^{n-1} P_0$$

Sean $\frac{\lambda}{\mu} = \frac{12}{8} = \frac{3}{2}$, $\rho = \frac{\lambda}{S\,\mu} = \frac{12}{16} = \frac{3}{4}$, entonces:

$$P_n = \frac{\lambda}{\mu}\rho^{n-1}P_0 \ , \ n \geq 1$$

Donde:

$$P_0 = \left[1 + \frac{3}{2} + \frac{3}{2}\cdot\frac{3}{4} + \frac{3}{2}\cdot\left(\frac{3}{4}\right)^2 + \ldots + \frac{3}{2}\cdot\left(\frac{3}{4}\right)^{n-1} + \ldots\right]^{-1} = \left[1 + \frac{3}{2} + \frac{3}{2}\cdot\rho + \frac{3}{2}\cdot\rho^2 + \ldots + \frac{3}{2}\cdot\rho^{n-1} + \ldots\right]^{-1}$$

$$= \left[1 + \frac{3}{2}\left(1 + \rho + \rho^2 + \ldots + \rho^{n-1} + \ldots\right)\right]^{-1} = \left[1 + \frac{3}{2}\left(\frac{1}{1-\rho}\right)\right]^{-1}, \ \rho < 1$$

Es decir:

$$P_0 = 0.1429$$

De tal forma que:

$$P_n = \frac{3}{2}\rho^{n-1}\left(0.1429\right) = 0.21435\,\rho^{n-1}, \ n \geq 1$$

El número promedio de clientes en la fila es:

$$L_q = \sum_{n=2}^{\infty}\left(n-2\right)\left(0.21435\,\rho^{n-1}\right) = 0.21435\ \sum_{n=2}^{\infty}\left(n-2\right)\rho^{n-1}$$

Convenientemente, hágase $j = n-2$, con lo que:

$$L_q = 0.21435 \sum_{j=0}^{\infty} j\rho^{j+1} = 0.21435\,\rho^2 \sum_{j=0}^{\infty} j\rho^{j-1} = 0.21435\,\rho^2 \sum_{j=0}^{\infty} D_\rho\left(\rho^j\right)$$

$$= 0.21435\,\rho^2\; D_\rho \sum_{j=0}^{\infty} \rho^j = 0.21435\,\rho^2\; D_\rho\left(\frac{1}{1-\rho}\right) = 0.21435\left[\frac{\rho^2}{(1-\rho)^2}\right]$$

Y por tanto:

$$L_q = 0.21435\left[\frac{\left(\frac{3}{4}\right)^2}{\left(1-\frac{3}{4}\right)^2}\right] = 1.92915 \text{ clientes.}$$

Así, el tiempo promedio de permanencia en la fila será:

$$W_q = \frac{L_q}{\lambda} = \frac{1.92915}{12} = 0.1608 \text{ horas.}$$

Es conveniente que recuerdes el resultado de este problema, pues será de utilidad para hallar la solución de los modelos (M/G/S) (d/∞/∞), los cuales se tratarán más adelante.

2.3. Modelos de colas para procesos no markovianos

Cuando los sistemas de colas se derivan de procesos no markovianos, su solución es más sencilla que los casos de la sección anterior, pues únicamente se hace necesario que apliques las fórmulas que se van a describir en cada caso, salvo en el modelo (M/G/S) (d/∞/∞), en el cual para calcular el número promedio de clientes en la cola, será necesario considerar el sistema como (M/M/S).

Recuerda que este curso es una aplicación de todo lo aprendido, por ello se enfoca más a la realización de ejercicios que a la demostración y extracción de fórmulas para cada uno de los casos que se presentarán.

2.3.1. Modelos (M/G/1) (d/∞/∞)

En este modelo se tiene una entrada Poisson, es decir, existen tiempos de llegada exponenciales con una tasa media de llegadas λ y presenta tiempos de servicio independiente e idénticamente distribuido de cualquier clase. Asimismo, se cuenta con un servidor que atiende a los clientes bajo un cierto orden determinado.

Las fórmulas empleadas en estos modelos se muestran a continuación.

Formulario 2.3.1.1.

Número promedio de clientes en la fila $\left(L_q\right)$

$$L_q = \frac{\lambda^2 V\left(t\right) + \rho^2}{2\left(1-\rho\right)}$$ (Fórmula de Pollaczek-Khintchine)

Utilización promedio del servicio $\left(\rho\right)$

$$\rho = \frac{\overline{\lambda}}{\mu}$$

μ : capacidad del servidor.

$\bar{\lambda}$: tasa promedio de llegadas.

Número promedio de clientes en el sistema de colas (L)

$$L = L_q + \rho$$

L_q : número promedio de clientes en la fila.

ρ : utilización promedio del servicio.

Tiempo promedio de espera en el sistema (W)

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

L : número promedio de clientes en el sistema de colas.

$\bar{\lambda}$: tasa promedio de llegadas.

Tiempo promedio de espera en la fila (W_q)

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

L_q : número promedio de clientes en la cola.

$\bar{\lambda}$: tasa promedio de llegadas.

Ejemplo 2.3.1.1.

Una fábrica de refrescos cuenta con un robot que llena de líquido una botella en un tiempo aproximado de un minuto de manera constante. La entrada de botellas vacías sigue una distribución Poisson con media de 59.5 piezas por hora. Clasifica este sistema según Kendall-Lee y calcula el promedio de botellas que "se forman" para ser llenadas (L_q) y el tiempo promedio que permanecen en el sistema (W).

Solución:

Este sistema se clasifica como $(M/G/1)(FCFS/\infty/\infty)$, en virtud de lo cual se pueden aplicar las relaciones del formulario 2.3.1.1., usando $\lambda = 59.5$ botellas/hora, $\mu = 60$ botellas/hora, $E(t) = 1$ minuto/botella, $V(t) = 0$.

Primero se verifica que el sistema puede alcanzar un estado estable, es decir, $\rho < 1$:

$$\rho = \frac{\bar{\lambda}}{\mu} = \frac{59.5}{60} \approx 0.99 < 1$$

Y como el sistema se puede tornar estable, entonces se calcula L_q y W:

$$L_q = \frac{\lambda^2 V(t) + \rho^2}{2(1-\rho)} = \frac{(59.5)^2(0) + (0.99)^2}{2(1-0.99)} \approx 49.005 \text{ botellas.}$$

Y ahora:

$$W = W_q + E(t) = \frac{L_q}{\bar{\lambda}} + E(t) = \frac{49.005}{59.5^2} + 1 \approx 1.014 \text{ minutos.}$$

2.3.2. Modelos (M/G/S) (d/∞/∞)

Estos sistemas cuentan, al igual que en el caso anterior, con una entrada Poisson, es decir, existen tiempos de llegada exponenciales con una tasa media de llegadas λ y presenta tiempos de servicio independientes e idénticamente distribuidos de cualquier clase, pero tienen S servidores que atienden a un número ilimitado de clientes potenciales bajo un determinado orden con media $E(t)$ y varianza $V(t)$. En este caso es necesario que $\rho < 1$ con la finalidad de que se tenga un estado constante.

Las fórmulas aplicables a estos modelos son:

Formulario 2.3.2.1.

Número promedio de clientes en la fila $\left(L_q\right)$

$$L_q \le \left(\frac{1+C_S^2}{2}\right)\left(L_{q(M/M/S)}\right)$$

C_S^2 : coeficiente cuadrado de variación para el tiempo de servicio.

$L_{q(M/M/S)}$: número promedio de clientes en la fila del sistema (M/M/S).

Utilización promedio del servicio (ρ)

$$\rho = \frac{\overline{\lambda}}{S\mu}$$

S : número de servidores.

μ : capacidad del servidor.

$\overline{\lambda}$: tasa promedio de llegadas.

Número promedio de clientes en el sistema de colas (L)

$$L = L_q + \rho$$

L_q : número promedio de clientes en la fila.

ρ : utilización promedio del servicio.

Tiempo promedio de espera en el sistema (W)

$$W = \frac{L}{\overline{\lambda}}$$

L : número promedio de clientes en el sistema de colas.

$\overline{\lambda}$: tasa promedio de llegadas.

Tiempo promedio de espera en la fila $\left(W_q\right)$

$$W_q = \frac{L_q}{\overline{\lambda}}$$

L_q : número promedio de clientes en la cola.

$\overline{\lambda}$: tasa promedio de llegadas.

Ejemplo

En un sistema de colas con dos servidores las llegadas siguen una distribución Poisson con media de 12 usuarios por hora, mientras que el tiempo de servicio es en promedio de 60 minutos por cliente con una varianza de 0.02. Si un cliente llega al sistema y se encuentra con que todos los servidores están ocupados, se forma en una fila común hasta ser atendido bajo el criterio FCFS. Calcula el número promedio de clientes en la fila esperando atención, además del tiempo promedio de permanencia en la fila, considerando turnos de servicio de ocho horas, y al tomarse en cuenta que si el sistema funciona durante ocho horas por día, se clasifica como (M/G/7)(FCFS/∞/∞). Calcula el número de clientes en la fila esperando atención, así como el tiempo promedio que permanecen en la fila y el valor máximo que puede tomar L_q.

Solución:

Claramente se tiene que $\lambda = 12$ usuarios/día, $\mu = 8$ usuarios/día, $S = 2$ servidores, $E\left(t\right) = \frac{1}{12}$ día/usuario y $V\left(t\right) = 0.02$ día/usuario.

Se analiza primero la estabilidad del sistema.

$$\rho = \frac{\bar{\lambda}}{S\mu} = \frac{12}{2\cdot 8} = \frac{12}{16} = \frac{3}{4} = 0.75 < 1$$

Por tanto este sistema no colapsa, con lo que se puede continuar calculando los demás parámetros solicitados.

$$L_q \le \left(\frac{1+C_S^2}{2}\right)\left(L_{q(M/M/S)}\right)$$

Observa que esta fórmula implica resolver el problema considerando al sistema como un modelo markoviano y al proceder como se hizo en un ejemplo anterior.

Cabe aclarar que este problema fue redactado de tal manera que la solución que se brinda en el ejemplo de referencia sea de utilidad para resolver el sistema actual. Así:

$$L_q \le \left(\frac{1+\left(\frac{V(t)}{\left[E(t)\right]^2}\right)}{2}\right)\left(L_{q(M/M/S)}\right) \Rightarrow L_q \le \left(\frac{1+\left(\frac{0.02}{\left[1/12\right]^2}\right)}{2}\right)(1.92915)$$

$$\Rightarrow L_q \le \left(\frac{1+\left(\frac{0.02}{\left[1/12\right]^2}\right)}{2}\right)(1.92915) \Rightarrow L_q \le 3.742551$$

Con lo anterior, a lo más 3.742551 clientes permanecen en la fila en espera de atención.

De aquí se tiene que el tiempo promedio que permanecen en la fila, el valor máximo que puede tomar L_q, está dado por…

$W_q = \frac{L_q}{\bar{\lambda}} = \frac{3.742551}{12} = 0.31187925$ horas.

2.3.3. Modelos (G/G/1) (d/∞/∞)

En este caso se tiene un sistema con número de clientes potenciales y capacidad del sistema ilimitados, debido a lo cual nuevamente se debe tener que $\rho < 1$ para que el sistema presente un estado estable. Asimismo, cuenta con un servidor, y el tiempo entre llegadas y el tiempo de servicio son distribuciones generales con media $E(t)$ y varianza $V(t)$.

Las fórmulas empleadas en este tipo de sistemas son las siguientes:

Formulario 2.3.3.1.

Número promedio de clientes en la fila (L_q)

$$L_q = \frac{\rho^2\left(C_a^2 + C_S^2\right)}{2(1-\rho)} \cdot k$$

Donde

$$k = \begin{cases} \exp\left(-\frac{2(1-\rho)\left(1-C_a^2\right)^2}{3\rho\left(C_a^2 + C_S^2\right)}\right) & , si\ C_a^2 < 1 \\ 1 & , si\ C_a^2 = 1 \\ \exp\left(-\frac{(1-\rho)\left(C_a^2 - 1\right)^2}{\left(C_a^2 + 4C_S^2\right)}\right) & , si\ C_a^2 > 1 \end{cases}$$

C_a^2 : coeficiente cuadrado de variación para el tiempo entre llegadas.

C_S^2 : coeficiente cuadrado de variación para el tiempo de servicio.

ρ : utilización promedio del servicio.

Utilización promedio del servicio (ρ)

$$\rho = \frac{\overline{\lambda}}{\mu}$$

μ : capacidad del servidor.

$\overline{\lambda}$: tasa promedio de llegadas.

Número promedio de clientes en el sistema de colas (L)

$$L = L_q + \rho$$

L_q : número promedio de clientes en la fila.

ρ : utilización promedio del servicio.

Tiempo promedio de espera en el sistema (W)

$$W = \frac{L}{\overline{\lambda}}$$

L : número promedio de clientes en el sistema de colas.

$\overline{\lambda}$: tasa promedio de llegadas.

Tiempo promedio de espera en la fila (W_q)

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

L_q : número promedio de clientes en la cola.

$\bar{\lambda}$: tasa promedio de llegadas.

Los problemas que se presentan sobre estos modelos implican la aplicación directa de las relaciones mostradas en el formulario 2.3.3.1., al igual que en el modelo (M/G/1) (d/∞/∞).

Ejemplo

El tiempo entre llegadas de clientes a una consultoría con un asesor sigue una distribución Erlang con parámetros $\mu = 0.7$ y $k = 0.3$. El tiempo promedio de la asesoría es de 1.2 horas con una desviación estándar de 0.7 horas. Determina el número promedio de clientes en la cola y el tiempo promedio de espera en la fila.

Solución:

Como puedes observar, este sistema se trata de un modelo (G/G/1)(FCFS/∞/∞), por tanto es posible aplicar las fórmulas 2.3.3.1.

Se deben calcular inicialmente la esperanza y la varianza para el tiempo entre llegadas.

$E(a) = \frac{1}{0.7} \approx 1.4286$ horas/clientes.

$V(a) = \frac{1}{0.3(0.7^2)} \approx 6.8027$ hora/clientes.

$\lambda = 0.7$ clientes/hora.

Además:

$$C_a^2 = \frac{V(a)}{\left[E(a)\right]^2} = \frac{1.1534}{0.5882^2} \approx 3.3337 > 1$$

Se debe hacer lo mismo para el tiempo de servicio:

$E(t) = 1.2$ horas/cliente.

$V(t) = 0.49$ hora/cliente.

$\mu = 0.83$ clientes/hora.

$$C_S^2 = \frac{V(t)}{\left[E(t)\right]^2} = \frac{0.49}{1.2^2} \approx 0.3403$$

Se debe analizar la estabilidad del sistema:

$$\rho = \frac{\overline{\lambda}}{\mu} = \frac{0.7}{0.83} \approx 0.8434 < 1$$

Como consecuencia de lo anterior, es posible proseguir con la búsqueda de L_q y W_q.

$$L_q = \frac{\rho^2\left(C_a^2 + C_S^2\right)}{2(1-\rho)} \cdot \exp\left(-\frac{(1-\rho)\left(C_a^2 - 1\right)^2}{\left(C_a^2 + 4C_S^2\right)}\right)$$

$$= \frac{0.8434^2\left(3.3337 + 0.3403\right)}{2(1-0.8434)} \cdot \exp\left(-\frac{(1-0.8434)(3.3337-1)^2}{\left(3.3337 + 4(0.3403)\right)}\right)$$

$= 8.3442(0.3759) \approx 3.1366$ clientes.

Asimismo:

$W_q = \dfrac{L_q}{\overline{\lambda}} = \dfrac{3.1366}{0.7} \approx 4.4809$ horas.

Antes de continuar con las actividades, realiza un formulario con todas las relaciones presentadas en la unidad.

Cierre de la unidad

En esta unidad aprendiste a analizar las líneas de espera para optimizar costos a través de las reglas correspondientes de este tipo de sistemas.

En la unidad 3 la meta es simular números pseudoaleatorios y variables aleatorias a través de diversas técnicas de simulación.

Para saber más

Análisis de líneas de espera a través de teoría de colas y simulación:

Portilla, LM, Arias Montoya, L., & Fernández Henao, SA (2010). Análisis de líneas de espera a través de teoría de colas y simulación. Scientia et Technica, XVII (46), 56-61. Disponible en http://www.redalyc.org/articulo.oa?id=84920977012

Aplicaciones de la teoría de colas a la provisión óptima de servicios sociales: el caso del servicio de teleasistencia:

Peláez Fermoso, FJ, Gómez García, IM, & García González, a. (2011). Aplicaciones de la Teoría de Colas a la provisión óptima de servicios sociales: El caso del servicio de Teleasistencia. Estudios de Economía Aplicada, 29 (3), 1-25. Disponible en https://www.redalyc.org/pdf/301/30122405014.pdf

Fuentes de consulta

Básica

- Azarang, M. y García, E. (1997). *Simulación y análisis de modelos estocásticos*. México: McGraw-Hill Interamericana.
- Evans, M. y Rosenthal, J. (2005). *Probabilidad y estadística: la ciencia de la incertidumbre*. España: Reverté.
- Hillier, F. y Lieberman, G. (1997). *Introducción a la investigación de operaciones. México*: McGraw-Hill Interamericana.
- Ibe, O. (2009). *Markov Processes for Stochastic Modeling*. Reino Unido: Elsevier Academic Press.
- Springer, C., Herlihy, R., Mall, R. y Beggs, R. (1972). *Modelos probabilísticos: serie de matemáticas para la dirección de negocios.* México: Unión tipográfica editorial hispano-americana.
- Taha, H. (1998). *Investigación de operaciones: una introducción*. México: Prentice Hall.
- Taylor, H. y Karlin, S. (1998). *An Introduction to Stochastic Modeling*. EUA: Academic Press.