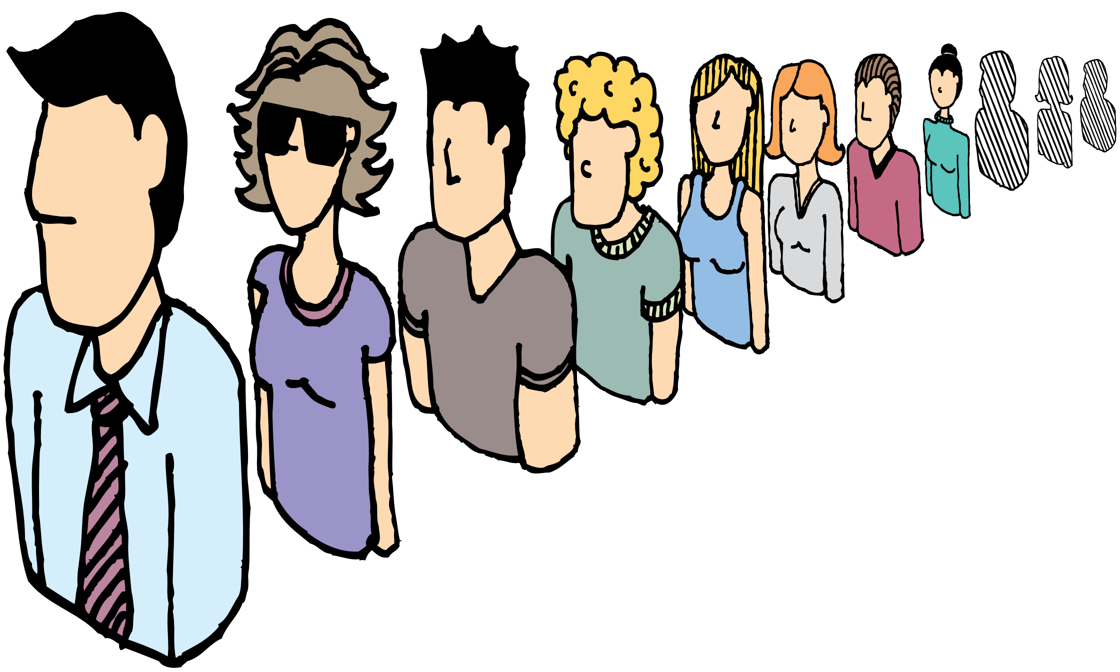
UNIDAD 3. TEORÍA DE COLAS



Teoría de colas

Tabla de contenido

Introducción

Las líneas de espera o las colas se presentan diariamente en la vida cotidiana. Por ejemplo, al realizar una llamada, a veces hay que esperar a que sea recibida porque la red del operador está ocupada. Así mismo, cuando se realiza mercado en un almacén de cadena o al registrarse en el aeropuerto para acceder a la sala de espera, también se presentan colas. En estas situaciones y numerosas circunstancias en las que se tiene que esperar se forman colas, por esta razón es importante estudiar y analizar el comportamiento de este sistema, con el fin de optimizar un servicio y unos costos.

Un modelo de formación de colas se puede nombrar como un sistema, ya que cuenta con un grupo de elementos que se comunican entre sí porque tienen una finalidad común. Cuenta con unos clientes, que son los miembros del sistema, una línea de espera y unos servidores; la interacción de estos elementos con determinadas características se puede estudiar para tomar decisiones y mejorar un proceso, a esto hace referencia la teoría de colas.

El modelo de un sistema de colas es bastante útil para tomar decisiones con respecto al servicio que se está ofreciendo en determinada situación y a la instalación del sistema. Cuando los clientes tienen que esperar en una cola para recibir ciertos servicios, están pagando un costo en tiempo más alto del esperado. Las líneas de espera largas también son costosas porque además de costos implican perder prestigio y clientes.

Objetivos

Objetivo general

Identificar en qué consiste la teoría de colas, también llamada líneas de espera, determinando las medidas de desempeño del sistema acorde a sus características y empleándola como herramienta para la toma de decisiones.

Objetivos específicos

- Identificar las consideraciones cuantitativas de costo y las cualitativas de servicio.
- Usar correctamente las fórmulas necesarias para calcular el tiempo en la línea de espera o de permanencia en el sistema.
- Caracterizar cualitativa y cuantitativamente a una cola y determinar los niveles adecuados de ciertos parámetros del sistema.

3.1 Definiciones iniciales

La **teoría de colas** es el estudio del comportamiento de líneas de espera. Para Bronson (1993, 262) "un sistema de líneas de espera es un conjunto de clientes, un conjunto de servidores y un orden en el cual los clientes llegan y son atendidos". Las líneas de espera se presentan cuando los clientes llegan a solicitar un servicio a un servidor, el cual tiene capacidad limitada de atención. Así mismo, la línea de espera se forma cuando el cliente llega y el servidor no está disponible y el cliente decide esperar. Una **cola** es una línea de espera.

Los **sistemas de colas** son modelos de sistemas que proporcionan servicio. Como modelo, pueden representar cualquier sistema en donde los trabajos o clientes llegan buscando un servicio de algún tipo y salen después de que dicho servicio haya sido atendido.

Cuando se forma una cola, se habla de **clientes**, de una **línea de espera o cola** y de unos **servidores**. Por ejemplo, cuando las personas realizan mercado en un almacén de cadena en el que posiblemente varios clientes asisten a la misma hora, éstas deberán esperar en cola para pagar en la caja si los servidores o cajeros están ocupados atendiendo a otros clientes. Pueden haber dos reacciones en la cola: la primera, los clientes pueden esperar temporalmente porque observan que el servicio es adecuado y los clientes que llegaron con anterioridad están siendo atendidos o, segundo, la cola tiende a ser explosiva, se hace cada vez más larga a medida que transcurre el tiempo y aunque unos pueden esperar, otros deciden abandonarla.

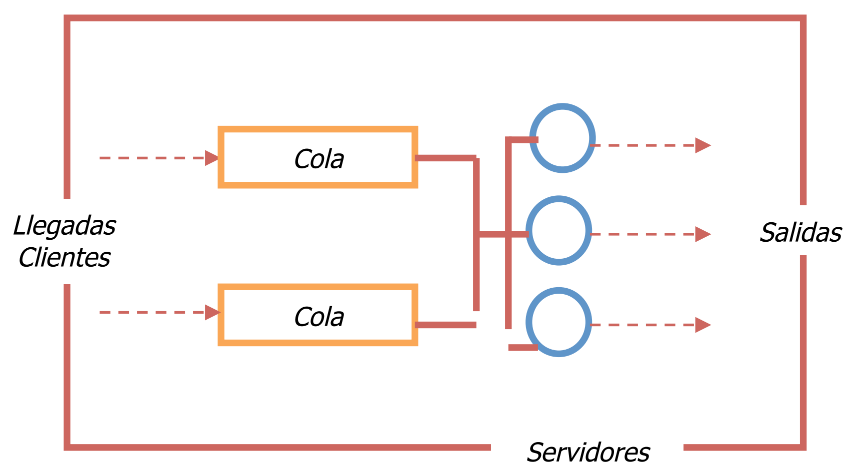


Gráfico 3.1. Representación de un sistema de colas.

3.2 Elementos existentes en un modelo de colas

En el sistema conformado en una línea de espera existen varios elementos que determinan sus características:

- **Fuente de entrada o población potencial:** Conjunto de clientes o llegadas que quieren solicitar un servicio. La fuente de entrada puede ser finita o infinita.

- **Cliente:** Miembro de la población potencial que solicita un servicio.

- **Capacidad de la cola:** Cantidad máxima de clientes que pueden estar haciendo cola antes de que sean atendidos.

- **Disciplina de la cola:** Es la forma de selección de los clientes para que sean atendidos. Las disciplinas más recurrentes son:
  - FIFO (First in first out) ó FCFS (First come first served): Se atiende al primer cliente que haya llegado.
  - LIFO (Last in first out) ó LCFS (Last come first served): Se atiende al último cliente que haya llegado.
  - RSS (Random selection of service) ó SIRO (Service in random order): La atención de los clientes se realiza al azar, de manera aleatoria.
  - Processor Sharing: Sirve a los clientes igualmente. La capacidad de la red se comparte entre los clientes.

- **Mecanismo de servicio:** Procedimiento del servicio que se le brinda a los clientes; consiste en las instalaciones de servicio, cada una de ellas con uno o más canales de servicio que reciben el nombre de servidores. Para determinar el mecanismo de servicio se debe conocer el número de servidores y la distribución de probabilidad del tiempo que toma cada servidor en brindar el servicio.

- **Cola:** Conjunto de clientes que esperan a recibir un servicio.

- **Sistema de la cola:** Conjunto formado por la cola y mecanismo de servicio junto con la disciplina de la cola. Un modelo de sistema de colas debe especificar la distribución de probabilidad de los tiempos de servicio para cada servidor. Normalmente se emplea una distribución exponencial para los tiempos de servicio, pero también se puede utilizar la distribución degenerada o determinística para tiempos de servicio constantes o la distribución Erlang (Gamma).

3.3 Proceso de servicio y de salida

Se refiere a la forma como se ejecuta el servicio y está relacionado con la estructura de la instalación.

A continuación se pueden observar los diferentes sistemas de servicio:

Una cola, un servidor

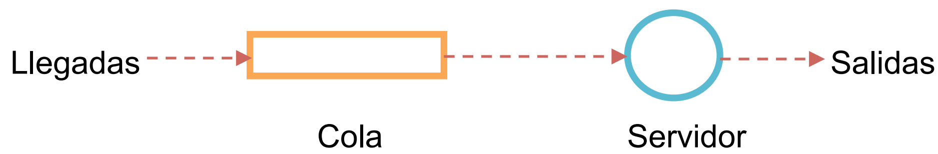


Gráfico 3.2. Una cola, un servidor.

Este primer sistema es de un servidor y una cola. Por ejemplo para hacer un pago de un producto en un almacén en donde sólo hay una caja.

Una cola, múltiples servidores

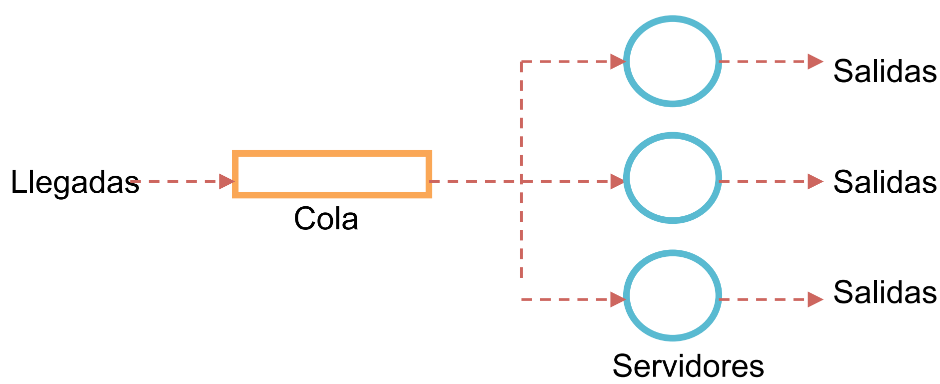


Gráfico 3.3. Una cola, múltiples servidores.

El segundo sistema es una línea con múltiples servidores. Por ejemplo, al ir a un banco y solicitar un turno, el cliente espera y puede ser atendido por cualquiera de los servidores o cajeros disponibles.

Varias colas, múltiples servidores

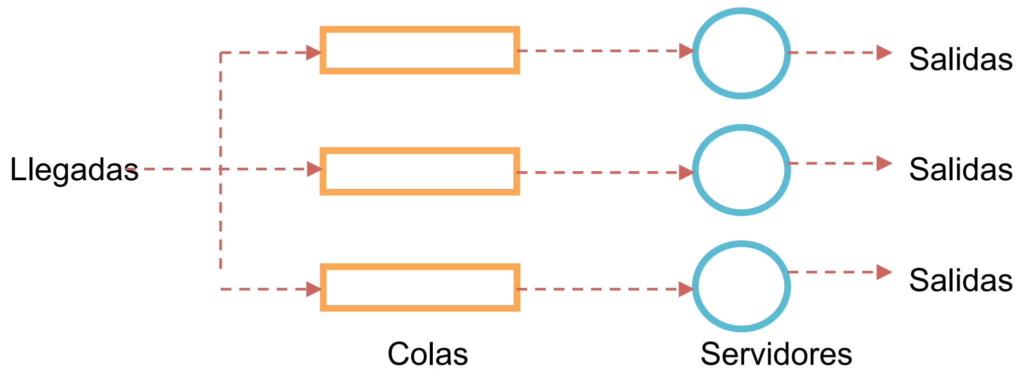


Gráfico 3.4. Varias colas, múltiples servidores.

En el tercer sistema cada servidor tiene una línea de espera. Por ejemplo, las tiendas de conveniencia o de autoservicio. Cada registradora es una estación que proporciona el mismo servicio. En estos sistemas los servidores pueden ser idénticos, ya que proporcionan la misma clase de servicio con igual rapidez o pueden no ser idénticos si todos los cajeros no tienen la misma experiencia y en este caso a unos les tomará mayor tiempo de servicio que a otros.

Una cola, servidores secuenciales

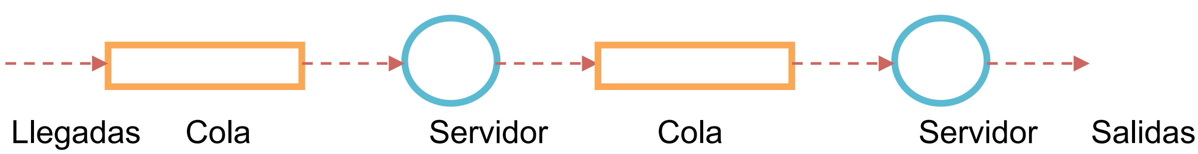


Gráfico 5. Una cola, servidores secuenciales.

El cuarto sistema es una línea con servidores en serie.

Otra característica del proceso de servicio es el número de clientes atendidos al mismo tiempo en un servidor. Así mismo si se permite o no la prioridad, es decir, si un servidor puede detener el proceso con el cliente que está atendiendo para dar paso a un cliente que acaba de llegar.

Proceso de salida

Se puede considerar cuando el cliente abandona el sistema luego de ser atendido. A esto se le puede dar el nombre de sistema de "colas de un paso" o cuando los clientes o productos reciben un servicio, pero se trasladan a otro para ser sometidos a otro proceso, lo que se llama una "red de colas".

3.4 Selección y evaluación del sistema de colas

La teoría de colas implica dos aspectos importantes, el primero, la selección del modelo adecuado con el fin de determinar las medidas de desempeño del sistema, y la segunda, el desarrollo de un modelo de decisión basado en las medidas de desempeño del sistema, con el fin de diseñar la instalación de servicio.

La selección del modelo para analizar una línea de espera, está determinado, principalmente, por las **distribuciones de los tiempos de llegada y los tiempos de servicio.** En la práctica, estas distribuciones se determinan observando las líneas de espera durante su operación y registrando los datos correspondientes.

Un sistema se puede observar cuando está operando normalmente, cuando sus partes están funcionando y teniendo en cuenta los periodos de mayor actividad para registrar momentos de congestión en donde hay mayores tasas de llegadas, es decir, mayor número de clientes por unidad de tiempo o también se puede observar cuando el sistema está en estado estable.

Los datos se pueden recolectar midiendo el tiempo entre llegadas o salidas sucesivas para determinar los tiempos entre arribos o de servicio, con el fin de analizar las distribuciones de los tiempos entre arribos o servicios. También se puede contar el número de llegadas o salidas durante una unidad de tiempo seleccionada, para analizar las distribuciones del número de llegadas o salidas. La información se debe resumir de manera clara y organizada, empleando herramientas como histogramas de frecuencias, gráfica de la distribución empírica o prueba de bondad de ajuste.

3.4.1 Características

Existen dos clases básicas de tiempo entre llegadas:

- **Determinístico:** En el cual clientes sucesivos llegan en un mismo intervalo de tiempo, fijo y conocido.

- **Probabilístico:** En el cual el tiempo entre llegadas sucesivas es incierto y variable. Los tiempos entre llegadas probabilísticos se describen mediante una distribución de probabilidad. Por lo general se emplea la distribución exponencial, ya que se ajusta a las características del sistema. La función de densidad para una distribución exponencial depende del parámetro λ (letra griega lambda), y está dada por:

$$f(t)=(1/\lambda)e-\lambda t$$

En donde λ (lambda) es el número promedio de llegadas en una unidad de tiempo.

Con una cantidad T, de tiempo, se puede hacer uso de la función de densidad para calcular la probabilidad de que el siguiente cliente llegue dentro de las siguientes T unidades a partir de la llegada anterior, de la siguiente forma:

$$P(\text{tiempo entre llegadas} <= T)=1-e-\lambda t$$

Por otro lado, es importante tener una idea de **cuánto tiempo se requiere para llevar a cabo el servicio.** Esta cantidad es relevante debido a que cuanto más dure el servicio, más tendrán que esperar los clientes que llegan. Este tiempo puede ser determinístico o probabilístico. Con un tiempo de servicio determinístico, cada cliente requiere, precisamente, de la misma cantidad conocida de tiempo para ser atendido. Con un tiempo de servicio probabilístico, cada cliente requiere una cantidad distinta e incierta de tiempo de servicio. Los tiempos de servicio probabilísticos se describen mediante una distribución de probabilidad. Es aconsejable seguir la distribución exponencial porque para temas prácticos ha funcionado adecuadamente. En este caso, su función de densidad depende de un parámetro, por ejemplo la letra griega m y está dada por:

$$s(t)=(1/\mu)e-\mu t$$

En la que:

μ = número promedio de clientes atendidos por unidad de tiempo, de modo que:

1/ μ = tiempo promedio invertido en atender a un cliente.

3.5 Clasificación de los modelos de colas

La clasificación de los modelos se basa en los elementos de un sistema de espera que dependen de los siguientes factores:

- Distribución de llegadas (llegadas individuales o masivas en grupo).
- Distribución del tiempo de servicio (servicio individual o masivo).
- Diseño de la instalación de servicio (en serie, en paralelo, en red).
- Disciplina de servicio (FCFS, LCFS, SIRO, por prioridad).
- Tamaño de la línea de espera (finito o infinito).
- Fuente de llamadas (población de clientes finita o infinita).
- Conducta humana (cambios, renuncias).

Para aplicar las técnicas apropiadas se deben identificar las características del sistema de colas. La clasificación se realiza empleando letras y/o símbolos a través de la **Notación de Kendall**, una forma adecuada para resumir las características principales de las líneas de espera en paralelo, empleando determinada simbología.

a/ b /c : d /e /f

Dónde:

a = Distribución de llegadas: Proceso de llegadas.
b = Distribución del tiempo de servicio (o de salidas): Proceso de servicio.
c = Número de servidores en paralelo (c = 1, 2, 3, ...,).
d = Disciplina de servicio (FCFS, LCFS, SIRO o prioridad = Disciplina General, DG).
e = Número máximo admitido en todo el sistema (en la línea de espera más en el servicio).
f = Tamaño de la población de clientes (fuente de llamadas finita o infinita).

La distribución de llegadas (a) y del tiempo de servicio (b) se reemplaza por los siguientes códigos: M, D, Ek, GI, o G (cualquiera de los 5 códigos), y significan lo siguiente:

M = Distribución de llegadas o salidas de Poisson (proceso de Markov), es decir, distribución exponencial entre llegadas o tiempos de servicio. Para que exista un proceso de llegada Poisson al menos un cliente debe llegar a la cola durante un intervalo de tiempo. Para un intervalo de tiempo dado, la probabilidad de que llegue un cliente es la misma que para todos los intervalos de tiempo de la misma longitud y la llegada de un cliente no tiene influencia sobre la llegada de otro.

D = Tiempo entre llegadas o de servicio son constantes.

Ek = Distribución Erlang o Gamma para la distribución del tiempo entre llegadas o tiempo de servicio, con el parámetro K. Si K, que determina la desviación estándar de la distribución es igual a 1, la distribución Erlang es igual a la exponencial; si es igual a ∞ la distribución Erlang es igual a la distribución degenerada con tiempos constantes.

GI = Distribución de llegadas o del tiempo entre llegadas; es general (independiente).

G = Distribución del tiempo de servicio o salidas; es general (no independiente).

Respecto a la disciplina de servicio se considera "DG" para indicar que es una disciplina general en "notación kendall", y que pudiera ser FCFS, LCFS, SIRO o cualquier procedimiento que puedan utilizar los servidores para decidir el orden en que se escogerá a los clientes de la línea de espera para iniciar el servicio.

H = Distribución hiperexponencial.

Para el orden de atención a los clientes se emplean los siguientes símbolos:

FCFS = Primeras entradas, primeros servicios.
LCFS = Últimas entradas, primeros servicios.
SIRO = Orden aleatorio.
PR = Con base en prioridades.
GD = En forma general.

3.5.1 Terminología

Usualmente siempre es común utilizar la siguiente terminología estándar:

Estado del sistema: Número de clientes en el sistema.

L: Longitud de la cola, número de clientes que esperan un servicio.
Lq: Longitud esperada de la cola, excluyendo los clientes que están en servicio.
N(t): Número de clientes en el sistema de colas en el tiempo t ($t \geq 0$).
Nq (t): Número de clientes en la cola en el instante t.
Ns (t): Número de clientes en servicio en el instante t.
W: Tiempo de espera en el sistema para cada cliente.
Wq: Tiempo de espera en la cola para cada cliente.

Pn (t): Probabilidad de que n clientes estén en el sistema en el tiempo t, dado el número en el tiempo cero.
s: Número de servidores en el sistema de colas.
Pw: Probabilidad de que un cliente que llega tenga que esperar (ningún cajero vacío).
Po: Probabilidad de que no hayan clientes en el sistema.
Pd: Probabilidad de negación de servicio o probabilidad de que un cliente que llega no pueda entrar al sistema debido que la "cola está llena".
Tq: Tiempo que un cliente invierte en la cola.

λ: Tasa media de llegadas (número esperado de llegadas por unidad de tiempo) de nuevos clientes cuando hay n clientes en el sistema.

1/ λ: Tiempo esperado entre llegadas.

μ: Tasa media de servicio para todo el sistema (número esperado de clientes que completan su servicio por unidad de tiempo) cuando hay n clientes en el sistema.

1/μ: Tiempo esperado de servicio.

$\boldsymbol{\rho} = \frac{\lambda}{s\mu}$ **:** Congestión de un sistema.

Relaciones entre las medidas L, W, Lq y Wq

Si λ **=** Número promedio de llegadas por unidad de tiempo (tasa de llegadas).
μ **=** Número promedio de clientes atendidos por unidad de tiempo en un canal (tasa de servicio), entonces:

$$L = \lambda W \qquad Lq = \lambda Wq \qquad L = Lq + \lambda/\mu$$

El tiempo medio de servicio es una constante $1/\mu$ para toda $n \geq 1$

$$W = Wq + 1/\mu \qquad L = Lq + \rho$$

Estas relaciones son importantes porque permiten determinar las cuatro variables fundamentales: L, W, Lq y Wq.

Las probabilidades como medidas de desempeño permiten evaluar escenarios y establecer metas.

Dada la tasa media de llegadas λ y la tasa media de servicio μ, se define el factor de utilización del sistema ρ. Su fórmula con un servidor y con s servidores, respectivamente, es:

$$\rho = \frac{\lambda}{\mu} \qquad \rho = \frac{\lambda}{s\mu}$$

A continuación conozca algunos de los modelos que se manejan en la teoría de colas:

Modelo Simple: (M/M/1): (GD/:/:)

Este modelo es de un canal, una fase, con un origen de llegada ilimitado, una distribución de llegada Poisson, con una sola cola y una disciplina FIFO (FCFS), con una distribución de servicio exponencial y el ritmo de servicio, por lo general, es mayor al ritmo de llegada.

Ecuaciones a emplear:

Número medio de unidades en cola: $L_s = \frac{\lambda}{\mu-\lambda}$

Tiempo medio en el sistema: $W_s = \frac{1}{\mu-\lambda}$

Número medio de unidades en cola: $L_q = \frac{\lambda^2}{\mu(\mu-\lambda)}$

Tiempo medio en cola: $W_q = \frac{\lambda}{\mu(\mu-\lambda)}$

Utilización del sistema: $\rho = \frac{\lambda}{\mu}$

Probabilidad de que el sistema esté desocupado: $P_0 = 1 - \rho = 1 - \frac{\lambda}{\mu}$

Probabilidad de que haya más de k unidades en el sistema: $P_{n>k} = \left(\frac{\lambda}{\mu}\right)^{k+1}$

donde n es el número de unidades en el sistema.

A continuación podrá conocer algunos ejemplos en los cuales se aplica la teoría del modelo simple.

Ejemplo 1

Un funcionario de inmigración en el aeropuerto El Dorado atiende a una persona cada 5 minutos y la tasa media de llegada es de 9 personas por hora. Se deben obtener las medidas de desempeño para el sistema, además la probabilidad de tener 0 personas en el sistema, una cola de más de 3 personas y la probabilidad de esperar más de 30 minutos en la cola y en el sistema.

$\lambda = 9$, $\mu = 12$, $\rho = 9/12 = 0.75$

$L_s = \frac{\lambda}{\mu-\lambda} = 3$ personas

$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = 2.25$ personas

$W_s = \frac{1}{\mu-\lambda} = 0.33$ horas $= 20$ minutos

$W_q = \frac{\lambda}{\mu(\mu-\lambda)} = 0.25$ horas $= 15$ minutos

$P_0 = 1 - \rho = 1 - \frac{\lambda}{\mu} = 0.25$

$P(L_s > 3) = \rho^{3+1} = 0.32$

$P(W_s > 30/60) = e^{-\mu(1-\rho)t} = 0.22$

$P(W_q > 30/60) = \rho e^{-\mu(1-\rho)t} = 0.17$

Ejemplo 2

Un supermercado con muchas cajas registradoras para el pago de la mercancía, en donde los clientes llegan a cancelar su cuenta con una tasa de 90 por hora y que hay 10 cajas en servicio. Si hay poco intercambio entre las líneas, puede tratarse este problema como 10 sistemas separados (cajas) de una sola línea, cada una con una llegada de 9 clientes por hora. Para una tasa de servicio de 12 por hora y considerando M/M/1, evalúe el sistema.

Solución:
El cliente promedio espera 15 minutos antes de ser atendido en la caja registradora. En promedio, hay un poco más de dos o tres clientes en la línea o tres líneas ocupadas en el sistema. El proceso completo lleva un promedio de 20 minutos. La caja está ocupada el 75 % del tiempo. Y, finalmente, el 32% del tiempo habrá cuatro personas o más en el sistema (o tres o más esperando en la cola).

Ejemplo 3

Los clientes que llegan a PanpaYá son en promedio 12 por minuto. De acuerdo a la distribución Poisson, el tiempo de atención se distribuye exponencialmente con un promedio de 8 minutos por cliente. La gerencia del negocio está interesada en determinar las medidas de desempeño para este servicio.

$\lambda = 1/12$ Clientes por minuto = 60/12= 5 por hora
$\mu = 1/8$ Clientes por minuto = 60/8 = 7.5 por hora
$P_0 = 1 - (\lambda/\mu) = 1 - (5/7.5) = 0.3333$
$P_n = [1 - (\square/\square)]\,(\lambda/\mu) = (0.3333)\,(0.6667)^n$
$P_w = \lambda / \mu = 0.6667$
$L = \lambda / (\mu-\lambda) = 2$
$L_q = \lambda^2 / [\square\,(\square - \square)] = 1.3333$
$W = 1 / (\square - \square) = 0.4$ horas = 24 minutos
$W_q = \lambda / [\square\,(\square - \square)] = 0.26667$ horas = 16 minutos

Ejemplo 4

La tasa de llegada de estudiantes a una biblioteca es de 10/hora. En la biblioteca existe una sola persona y atiende con una tasa de 5 minutos/persona. ¿Cuáles son las medidas de desempeño del sistema?

$\lambda = 10$ (tasa de llegada)

$\square = \frac{60}{5} = 12$ (tasa de servicio)

$S = 1$ (número de servidores)

$L = 5$

$L_q = 4.16$

$W = 0.5$

$W_q = 0.42$

$\rho = 0.83$

$P_0 = 0.16$

$P_1 = 0.14$

$P_2 = 0.11$

$P_3 = 0.09$

$P_4 = 0.08$

Modelo de cola multicanal: (M/M/s): (GD/:/:)

Este modelo es un sistema multicanal, de origen de llegada ilimitado, con una distribución de llegada Poisson con múltiples colas ilimitadas, una disciplina de cola FIFO (FCFS), con una distribución de servicio exponencial y con un ritmo de servicio mayor al ritmo de llegada.

Ecuaciones a emplear:

Probabilidad de que hayan cero unidades o personas en el sistema:

$$P_0 = \frac{1}{\left[\sum_{n=0}^{c-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{c!}\left(\frac{\lambda}{\mu}\right)^c \frac{c\mu}{c\mu-\lambda}}$$

$$P_{n} = \frac{\left(\lambda/\mu\right)^n P_0}{n!} \quad \text{si } 0 \leq n \leq s$$

$$P_{n} = \frac{\left(\lambda/\mu\right)^n P_0}{s!\,s^{n-s}} \quad \text{si n} > s$$

$$P_{w} = \left(\lambda/\mu\right)^s \frac{P_0}{s!\,(1-\rho)}$$

Número medio de personas o unidades en el sistema: $L_S = \frac{\mu\lambda\left(\frac{\lambda}{\mu}\right)^c}{(c-1)!\,(c\mu-\lambda)} P_0 \frac{\lambda}{\mu}$

Tiempo medio que una unidad permanece en el sistema: $W_S = \frac{\mu\left(\frac{\lambda}{\mu}\right)^c}{(c-1)!\,(c\mu-\lambda)} P_0 + \frac{1}{\mu}$

Número medio de personas o unidades esperando en la cola para recibir el servicio:

$$L_q = L_s - \frac{\lambda}{\mu}$$

Tiempo medio que una persona o unidad permanece en la cola:

$$W_q = w_s - \frac{1}{\mu}$$

A continuación podrá observar la aplicación del modelo M/M/s.

Ejemplo 5

En la biblioteca de una universidad cuyo personal está tratando de decidir cuántas fotocopiadoras debe instalar para el servicio de los estudiantes, se ha elegido una fotocopiadora particular que puede arrojar hasta 10 fotocopias por minuto. No se sabe cuál es el costo de espera para un estudiante, pero se piensa que los estudiantes no deben tener que esperar más de dos minutos en promedio. Si el número promedio de fotocopias que se sacan por usuario es cinco, ¿cuántas fotocopiadoras se deben instalar?, ¿cuál es la tasa de servicio? Si el número promedio de copias es cinco y la fotocopiadora puede sacar hasta 10 copias por minuto, entonces pueden utilizar el servicio en promedio hasta dos estudiantes por minuto. Pero en esto no se toma en cuenta el tiempo para cancelar las fotocopias y cambiar originales para que un estudiante desocupe y otro comience a fotocopiar. Suponga que se permite un 70 % del tiempo para estas actividades. Entonces la tasa de utilización de la fotocopiadora neta baja a 0.6 estudiantes por minuto. Además se supone que los periodos pico de fotocopiado tienen una tasa de llegada de 60 estudiantes por hora o 1 por minuto.

Modelo de tiempo de servicio constante: (M/D/1): (GD/:/:)

Es un sistema de un canal y de una fase. Con origen de llegada ilimitado, con una distribución de llegada Poisson, con una única cola ilimitada, con una disciplina de cola FIFO (FCFS), con una distribución de servicio exponencial y un ritmo de servicio mayor al ritmo de llegada.

Ecuaciones a emplear:

Número medio de personas o unidades esperando para recibir un servicio:

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

Tiempo medio que una persona o unidad permanece en cola:

$$W_q = \frac{\lambda}{2\mu\,(\mu - \lambda)}$$

Número medio de personas o unidades en el sistema:

$$L_s = L_q + \frac{\lambda}{\mu}$$

Tiempo medio que una unidad permanece en el sistema:

$$W_s = W_q + \frac{1}{\mu}$$

A continuación se aplicará el modelo M/D/1.

Ejemplo 6

Un banco puede atender a un cliente cada 5 minutos. La tasa media de llegadas es de 9 personas/hora. Obtener las medidas de desempeño.

$L_s = \lambda W_s$ = 1.875 clientes

$L_q = \frac{\rho^2}{2\,(1-\rho)}$ = 1.125 clientes

$W_s = W_q + \frac{1}{\mu}$ = 0.21 horas = 12.5 minutos

$W_q = \frac{L_q}{\lambda}$ = 0.125 horas = 7.5 minutos

Modelo M/Ek/1:

Ecuaciones a emplear:

$$L_s = \lambda W_s$$

$$L_q = \frac{\rho^2(k+1)}{2k\,(1-\rho)}$$

$$W_s = W_q + \frac{1}{\mu}$$

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

$$\rho < 1$$

Ejemplo 7

En una tienda de alquiler de autos se atiende a una persona cada 5 minutos. La tasa media de llegadas es de 9 autos/hora. Suponga $\sigma = 3.5$ minutos aproximadamente. Halle las medidas de desempeño de acuerdo con el modelo.

$L_s = 2.437$ personas

$L_q = 1.6875$ personas

$W_s = 0.2708$ horas = 16.25 minutos

$W_q =$ 0.1875 horas = 11.25 minutos

Modelo de población limitada: (M/M/c): (GD/N/N)

Es un sistema de un canal y una fase, con un origen de llegada limitado, una distribución de llegada Poisson, una sola cola ilimitada, con una disciplina de cola FIFO (FCFS), con una distribución de servicio exponencial y un ritmo de servicio mayor al ritmo de llegada.

M/ D / 15: DG / N /

M, significa que se tienen llegadas tipo Poisson; D, significa que se tiene tiempo de servicio o de salidas determinístico (constante); se tienen 15 servidores en paralelo; la disciplina de servicios es general; N, significa que el sistema sólo puede alojar a un máximo de N clientes.

M/ M / 4: DG /

M, significa que se tienen llegadas tipo Poisson; M, significa que se tiene tiempo de servicio o de salidas probabilístico exponencial (proceso de servicio Markov); se tienen 4 servidores o terminales en paralelo; la disciplina de servicios es general; / significa que el sistema tiene capacidad ilimitada y el siguiente / es para indicar que se tiene una población de clientes infinita.

M / D / 4: DG/

Indica que las llegadas son Poisson (el tiempo entre llegadas es probabilístico y exponencial o de Markov); el tiempo de servicio es determinístico. Existen 4 servidores, la disciplina de servicio es general y no hay límite en la capacidad ilimitada o de la fuente de llamadas.

M/M/R: DG/K/K; R

K corresponde, por ejemplo, al modelo de servicio de máquinas. Este modelo indica que se dispone de "R" técnicos en reparaciones para dar servicio a un total de "k" máquinas. Como una máquina descompuesta no puede generar nuevas llamadas mientras está en servicio, el modelo es un ejemplo de fuente de llamadas finita. Además, tanto el proceso de llegadas como el de servicio son probabilísticos y de Markov.

M / G / 1: Tiempos entre llegadas exponenciales, tiempos de servicio general y un sólo servidor, se cuenta con una población infinita y la posibilidad de infinitas filas.

Ecuaciones a emplear:

$$L_s = L_q + \frac{\lambda}{\mu}$$

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

$$W_s = W_q + \frac{1}{\mu}$$

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

$P_0 = 1 - \rho$

$P_w = \rho$

$\rho < 1$

A continuación podrá observar cómo se aplican los conceptos vistos del modelo M/G/1.

Ejemplo 8

En una central de medios en donde se emplean diariamente los computadores, se dañaron varios de estos elementos; la gerencia contrató una empresa para reparar cada uno de los computadores. El tiempo promedio para reparar un computador es de 2.25 horas. La desviación estándar del tiempo de reparación es de 45 minutos y cada computador para reparar llega a manos de los ingenieros en promedio cada 2.5 horas. De acuerdo a una distribución de Poisson, los ingenieros trabajan 9 horas diarias. En promedio el tiempo de reparación esperado debería ser de 2 horas y la desviación estándar esperada debería ser de 40 minutos. Los ingenieros desean conocer los efectos de emplear nuevos equipos para: mejorar el tiempo promedio de reparación de los computadores y mejorar el tiempo promedio que debe esperar la central de medios hasta que un computador sea reparado.

Solución:
El tiempo de atención no es exponencial ya que $\sigma = 1/\mu$

Con el sistema antiguo (sin los nuevos equipos)

λ = 1/2.5 = 0.4 computadores por hora.

μ = 1 / 2.25 = 0.444 computadores por hora.

σ = 45 / 60 = 0.75 horas.

Con el nuevo sistema (con los nuevos equipos)
μ = 1 / 2 = 0.5 computadores por hora.
σ = 40 / 60 = 0.6667 horas.

Ejemplo 9

En un cinema se puede atender a una persona que va a ingresar a las salas cada 5 minutos y la tasa media de llegada es de 9 personas/hora, σ = 2 minutos. Se deben obtener las medidas de desempeño, la probabilidad de tener 0 personas en el sistema y la probabilidad de que una persona tenga que esperar a ser atendido.

$L_s = L_q + \frac{\lambda}{\mu}$ = 1.31 + 0.75 = 2.06 personas

$L_q = \frac{\lambda^2 \sigma^2 + \rho^2}{2\,(1-\rho)}$ = 1.31 personas

$W_s = W_q + \frac{1}{\mu}$ = 0.228 horas = 13.7 minutos

$W_q = \frac{L_q}{\lambda}$ = 0.145 horas = 8.7 minutos

$P_0 = 1 - \rho = 0.25$ $\quad$ $P_w = \rho = 0.75$

Modelo (M/M/1/k)

Este modelo plantea que si el sistema está lleno, la capacidad es k y no se permite la entrada de nuevos clientes al sistema. Por tanto, la tasa de llegada efectiva no es constante y varía con el tiempo dependiendo de si el sistema está lleno o no.

Ecuaciones a emplear:

$\lambda_{ef} = \lambda\,(1-P_k)$

En este caso,

$P_0 = \rho^n p_0$ para n = 0,1,…..,k

Y no existe estado k + 1

Por tanto,

$$p_0 + p_1 + p_2 + \cdots p_k = 1$$

De esta manera:

$p_0 = \frac{1-\rho}{1-\rho^{k+1}}$, si $\lambda \neq \mu$

$p_0 = \frac{1}{1+k}$, si $\lambda = \mu$

Y siempre existe una distribución estacionaria.
Se obtienen las siguientes relaciones:

$L = \frac{\rho\,(1-(k+1)\,\rho^{k}+k\rho^{k+1}}{(1-\rho)(1-\rho^{k+1})}$ si $\lambda \neq \mu$

$L = \frac{k}{2}$ si $\lambda = \mu$

$L_q = L - (1 - p_0)$

$W = \frac{L}{\lambda_{ef}}$

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

Ejemplo 10

En una empresa metalmecánica, en una de sus secciones, se fabrican puertas metálicas para ascensores. Las puertas para ascensores se fabrican en una gran variedad (formatos, colores y huecos para vidrios variables). Se conoce que el proceso de producción se compone de 4 etapas consecutivas pero independientes. La empresa trabaja al año 220 días. Cada día de 7 horas y 30 minutos de trabajo efectivo. La demanda anual, según cálculos del pasado año, fue por una cantidad de 8.500 puertas. Los pedidos tienen una cantidad variable de unidades y los ajustes de cambio de partida, aunque importantes en ocasiones, no parecen repercutir en los ritmos de producción promedio de las diferentes etapas de trabajo.

La primera etapa se realiza simultáneamente por dos equipos de trabajo, con un ritmo promedio cada uno de ellos de una puerta cada 20 minutos. La segunda etapa la realiza un equipo de trabajo con un tiempo de ciclo promedio de 11 minutos por puerta. La tercera etapa requiere del uso de otra máquina con un tiempo de ciclo promedio de 10 minutos por puerta.

Por último, la cuarta etapa es de embalaje. Como es un trabajo manual, que realiza un operario, tiene un tiempo de ciclo de 18 minutos por unidad y se dispone de tantos trabajadores como se requiera, pues irán viniendo de otras secciones siempre que haya una puerta por preparar.

- Modele el problema según teoría de colas, estableciendo los parámetros básicos y asumiendo tiempos promedio exponenciales.

- ¿Cuál será el número promedio de puertas que habrá en el sistema?

- ¿Cuántos trabajadores serán necesarios normalmente en la cuarta etapa?

- Si un pedido tiene 30 puertas, ¿cuánto tiempo tardará en promedio en ser servido?

- ¿Cuál es el tiempo o plazo de entrega promedio previsto de una puerta? Si le dicen que el tiempo de entrega promedio es de 5 días, ¿a qué puede ser debido? Proponga un mecanismo de corrección.

- ¿Cuál será la consecuencia sobre la cantidad de puertas en la primera etapa si en lugar de dos equipos de trabajo con tiempos de ciclo como los citados se establece un único equipo más eficiente con un tiempo de ciclo de 9 minutos por unidad?

Al exponer el funcionamiento de la tercera etapa se ha simplificado el proceso. Realmente existen tres máquinas que pueden realizar la misma función, aunque en realidad nunca hay más de una fabricando. Las citadas máquinas se averían cada 5 horas en promedio (distribución negativa exponencial). Se dispone de 2 equipos de mantenimiento en la empresa que pueden poner en funcionamiento la máquina de nuevo en un tiempo promedio de 1 hora.

Solución:

El problema propuesto es una serie de colas con una entrada λ = 38’6 puertas/día.

a) La primera etapa es una cola M/M/2 con λ = 38’6 puertas/día. y μ = 22,5 puertas/día.

   La segunda etapa es una cola M/M/1 con λ = 38’6 puertas/día. y μ = 40,9 puertas/día
   La tercera etapa es una cola M/M/1 con λ = 38’6 puertas/día. y μ = 45 puertas/día
   La cuarta etapa es una cola M/M/∞ con λ = 38’6 puertas/día. y μ = 25 puertas/día

b) L1= 6,53 L2= 17 L3=6,07 L4= 1,54 LT=31,14 puertas

c) L=l W WT= 0,806 días = 6,05 horas

d) Habrá 1,54 trabajadores por término medio.

e) El tiempo que tardará será el de salir la primera 0,806 días más el que tardan en salir las 29 restantes. 0,806 + 29/λ =0,806+0,751=1,557 días = 11,68 horas

f) Si en realidad tardan 5 días en salir, es porque en el sistema hay puertas de más. La cantidad de puertas que hay es L= $\lambda \cdot (W-29/\lambda)$=162 puertas de más.

3.6 Costos en los sistemas de colas

Los dos elementos más importantes de un sistema de colas son la instalación del servicio y propiamente la cola. Estos componentes asumen unos costos que se deben considerar.

Sistema de costo mínimo

Es interesante analizar, además de las medidas de desempeño que permiten describir el comportamiento del sistema, modelos de decisión que minimicen los costos totales asociados con las líneas de espera.

Un modelo de costos en líneas de espera busca equilibrar los costos de espera contra los costos de incrementar el nivel de servicio. A medida que crece el nivel de servicio, los costos también crecerán y disminuirá el tiempo de espera de los clientes. El nivel de servicio óptimo se presenta cuando la suma de los dos costos es un mínimo. Para tasas bajas de servicio, se experimentan largas colas y costos de espera elevados. A medida que aumenta el servicio disminuyen los costos de espera, pero aumenta el costo de servicio y el costo total disminuye, sin embargo, se llega a un punto de disminución en el rendimiento.

El **costo de espera** o costo de clientes en espera por unidad de tiempo, está dado por:

$$Cw * L$$

Dónde:
Cw = costo de espera por llegada por unidad de tiempo.
L = longitud promedio de la línea en el sistema.

El **costo de servicio** o costo de operación de la instalación de servicio por unidad de tiempo, está dado por:

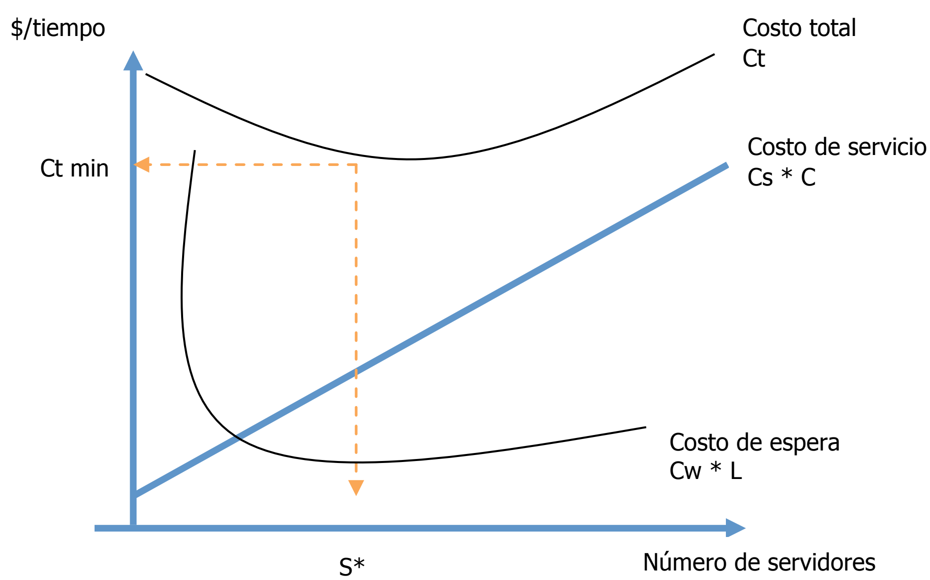
$$Cs * C$$

Dónde:

Cs = Costo por servidor por llegada por unidad de tiempo.
C = Número de servidores o cajeros.

La capacidad del servicio se puede aumentar añadiendo más servidores o haciendo servidores más eficientes.

Costo total del sistema: Costo de espera más costo de servicio = CwL + CsC



Gráfica 31. Optimización de costos

El costo de servicio aumenta con el incremento en el nivel del servicio, pero el costo por espera disminuye con el nivel.

Se debe buscar el nivel de servicio que minimiza el costo total.

A continuación podrá observar cómo se aplican los conceptos sobre costos de un sistema de colas.

Ejemplo 11

Una bomba de gasolina dispone de 3 servidores. Las personas llegan a una tasa de 40 por hora. El tiempo de servicio es de 3 minutos por persona.

La bomba se plantea si le conviene aumentar el número de servidores para satisfacer mejor a los clientes.

El costo para tener otro servidor es de 6 euros por hora. El costo de espera es de 18 euros por cliente.

$\lambda = 40$ tasa de llegadas

$\mu = 60/3 = 20$ tasa de servicio

$S = 3$ número de servidores

$Cs = 6$

$Cw = 18$

¿Cuántos servidores se deben emplear?

	S = 3	S = 4	S = 5
L	2.88	2.17	2.03
Costo de servicio	18	24	30
Costo de espera	52	39.13	36.72
Costo total	70	63.13	66.72

Tabla 3.1. Costos del sistema

La bomba debe emplear únicamente un servidor de más.

Ejemplo 12

Abbot, uno de los laboratorios de la industria farmacéutica, tiene un estacionamiento de carga en una de sus sedes que sirve a las farmacias cercanas a las instalaciones y existe sólo un trabajador para buscar los medicamentos del pedido de cada furgoneta y cargarlos en ella. Ocasionalmente las furgonetas de transporte se acumulan en el estacionamiento formando cola y de vez en cuando el trabajador está ocioso. Después de examinar las llegadas de las camionetas durante varias semanas, se determina que la tasa media de llegada es de 4 camionetas por hora y que la tasa de servicio es de 6 camionetas por hora. Los gestores del almacén están considerando añadir un trabajador adicional o incluso dos de ellos, para aumentar la tasa de servicio. El problema está en evaluar las diferentes opciones planteadas. Si se contrata un trabajador, el sistema seguirá siendo de cola simple, porque sólo una única camioneta puede cargarse a la vez. Si se emplean dos trabajadores, la tasa de servicio será igual a 12. Si se utilizan 3 trabajadores, la tasa de servicio será igual a 18.

En la siguiente tabla se han utilizado las ecuaciones adecuadas, acorde al sistema de colas, para obtener las medidas de eficiencia del sistema.

	Con 1 trabajador	Con 2 trabajadores	Con 3 trabajadores
Número medio de camionetas en la cola	1333	0.167	0.063
Número medio de camionetas en el sistema	2000	0.500	0.286
Tiempo medio de la camioneta en cola	0.333	0.042	0.016
Tiempo medio de la camioneta en el sistema	0.500	0.125	0.071
Ocupación del servicio	0.667	0.333	0.222

Tabla 3.2. Medidas de eficiencia del sistema

Suponiendo que los costos de operación de cada camioneta por hora son de 2.000 pesos y los trabajadores cobran 1.800 pesos por hora de trabajo con un horario de trabajo de 8 horas/día, en la siguiente tabla se presentan los costos asociados.

Trabajadores	Costo de camioneta por día	Costo de mano de obra por día	Costo total por día
1	320.000	144.000	464.000
2	80.000	288.000	368.000
3	46.000	432.000	478.000

Tabla 3.3. Costos asociados

Los empleadores tendrían que contratar un nuevo trabajador al sistema ya que esto representa una disminución en costos totales operacionales, aunque el factor de utilización pasará a ser de un 33%. Es decir, que los trabajadores tendrán 5 horas y 20 minutos para dedicarse a otras tareas dentro del laboratorio farmacéutico.

Ejemplo 13

El gerente del restaurante Abra Kadabra ha observado que cada minuto que un cliente tiene que esperar antes de terminar su servicio le cuesta un promedio de 30 centavos en negocio futuro perdido. Por lo tanto, desea estar seguro de que siempre tiene suficientes cajas abiertas para que la espera sea mínima. Un empleado de tiempo parcial opera cada caja, obtiene la orden del cliente y cobra. El costo total de cada empleado es $9 por hora. Durante la hora del almuerzo, los clientes llegan según un proceso de Poisson con tasa media de 66 por hora. Se estima que el tiempo necesario para servir a un cliente tiene distribución exponencial con media de 2 minutos. Determine cuántas cajas debe abrir el gerente de Abra Kadabra para minimizar su costo total esperado por hora.

λ = 66 clientes/hora   2 minutos $\longrightarrow$ 1 cliente

μ = 30 clientes/hora   60 minutos $\longrightarrow$ 30 clientes

$\rho = 66/30 = 2.2$

Costo de un cliente que espera antes de terminar su servicio.

Cs = $\frac{30}{100}$ \$/min = 0.3\$/min = 18\$/hora

Costo de un empleado en caja = 9$/hora.
Costo total = Cw*L + Cs*c.

Medidas de desempeño	3 cajas	4 cajas	5 cajas
ρ	73.33%	55%	44%
P0	0.08147	0.10456	0.109437
Lq	1.490936	0.277199	0.06594
Wq	0.02258	0.0041999	0.0009915
Ws	0.055923	0.037533	0.03433
Ls	3.69094	2.47720	2.26594

Tabla 3.4. Medidas de desempeño

No. De cajas	Cs*Ls	Cc*c	Costo Total ($/hora)
3	18*3.690.94	9*3	93.4369
4	18*2.47720	9*4	80.5896
5	18*2.26594	9*5	85.7869

Tabla 3.5. Costo mínimo

Para minimizar el costo total por hora se deberían abrir 4 cajas.

Ejemplo 14

En un centro de idiomas hay 3 fotocopiadoras para uso de los empleados. Sin embargo, debido a quejas de la cantidad de tiempo que pierden esperando que se desocupe una fotocopiadora, la gerencia planea comprar una o más. Durante las 2.000 horas de trabajo al año, los empleados llegan al área de fotocopiado según un proceso de Poisson con una tasa media de 30 por hora. Se cree que el tiempo que cada empleado necesita una fotocopiadora tiene una distribución exponencial con media de 5 minutos. El costo promedio de la productividad perdida debido al tiempo que pasa un empleado en el área de fotocopiado se estima en $25 por hora. La renta de cada fotocopiadora es de $3.000 por año. Determine cuántas fotocopiadoras debe tener el centro de idiomas para minimizar su costo total esperado por hora.

λ = 30 empleados/hora 5 minutos $\longrightarrow$ 1 empleado

μ = 12 empleados/hora 60 minutos $\longrightarrow$ 12 empleados

$\rho = 30/12 = 2.5$

Costo de un cliente que pasa en área de fotocopiado:
Cs = 25$/hora

Costo de la renta de cada fotocopiadora = 3000$/año * 1año/2000 horas = 1.5$/hora.

Medidas de desempeño	3 fotocopiadoras	4 fotocopiadoras	5 fotocopiadoras	6 fotocopiadoras	7 fotocopiadoras	8 fotocopiadoras
ρ	83.33%	62.5%	50%	41.67%	35.71%	31.25%
P_0	0.04494	0.11212	0.08010	0.08162	0.08198	0.08206
L_q	3.51124	0.81104	0.13037	0.03389	0.00858	0.00205
W_q	0.11704	0.02703	0.00435	0.00129	0.00029	0.000068
W_s	0.20037	0.11037	0.08768	0.08446	0.08362	0.08340
L_s	6.01124	3.31104	2.630.7	2.53389	2.50858	2.50205

Tabla 3.6. Medidas de desempeño.

No. De cajas	Cs*Ls	Cc*c	Costo Total ($/hora)
3	25*6.01124	1.5*3	154.7810
4	25*3.31104	1.5*4	88.7760
5	25*2.63037	1.5*5	73.2593
6	25*2.53389	1.5*6	72.3473
7	25*2.50858	1.5*7	73.2145
8	25*2.50205	1.5*8	74.5513

Tabla 3.7. Costo mínimo

Para minimizar el costo total por hora la compañía debe tener 6 fotocopiadoras.

Ejemplo 15

En un almacén de muebles se realiza inventario al finalizar cada año. Un grupo de 4 personas carga y descarga cada uno de los camiones que llegan. El gerente está despidiendo personal para disminuir costos y debe decidir el tamaño del futuro grupo. Los camiones tienen llegadas Poisson con tasa media de 1 por hora. El tiempo que requiere el grupo para cargar y/o descargar un camión tiene distribución exponencial. El tiempo esperado con 4 hombres es de 15 minutos. Si cambia el tamaño del grupo se estima que la tasa media de servicio (ahora 4 clientes por hora) sería proporcional al tamaño.

El costo por cada persona adicional del grupo es $20/hora. El costo atribuible a la espera de un camión se estima en $30/hora.

- Identifique los clientes y servidores de este sistema. ¿Cuántos servidores se tienen por hora?
  Clientes: Los camiones que llegan al muelle para ser cargados y/o descargados, población infinita.
  Servidores: Cada uno de los grupos que cargan y/o descargan los camiones, actualmente se tiene un grupo, es decir, un servidor.

- Encontrar las medidas de desempeño para este sistema de cola con un grupo de cuatro.
- Hacer el mismo ejercicio b con un grupo de tres.
- Hacer el mismo ejercicio b con un grupo de dos.
- ¿Se debe considerar una sola persona?
- Dados los resultados, ¿qué grupo debe elegir el gerente?
-

	Cantidad de personas en el grupo			
	4	**3**	**2**	**1**
Tasa de llegada	λ =1 camión/hora	λ =1 camión/hora	λ =1 camión/hora	λ =1 camión/hora
Tasa de servicio	μ = 15 minutos/camión	μ = 20 minutos/camión	μ = 30 minutos/camión	μ = 60 minutos/camión
Costo por cada grupo	80 $/hora	60 $/hora	40 $/hora	20 $/hora
Costo por cada camión que espera antes de terminar su servicio	30 $/hora	30 $/hora	30 $/hora	30 $/hora

Tabla 3.8. Medidas de desempeño

Conversión de la tasa de servicio a camión/hora para un grupo de 4 personas:

1 cliente $\longrightarrow$ 15 minutos
μ $\longrightarrow$ 60 minutos = 1 hora
μ = 4 camión/hora

Conversión de la tasa de servicio a camión/hora para un grupo de 3 personas:

1 cliente $\longrightarrow$ 20 minutos
μ $\longrightarrow$ 60 minutos = 1 hora
μ = 3 camión/hora

Conversión de la tasa de servicio a camión/hora para un grupo de 2 personas:

1 cliente $\longrightarrow$ 30 minutos
μ $\longrightarrow$ 60 minutos = 1 hora
μ = 2 camión/hora

	Cantidad de personas en el grupo			
	4	**3**	**2**	**1**
Utilización (%)	25%	33.33%	50%	100%
ρ: Factor de utilización del servicio	0.25	0.33	0.50	1.00
P0: Probabilidad de que no existan clientes en el sistema	0.75	0.67	0.50	0.00
Lq: No. Promedio de clientes en la cola	0.08333	0.16667	0.50000	-
Wq: Tiempo promedio que los clientes pasan en la cola	0.08333	0.16667	0.50000	-
Ws: Tiempo promedio que los clientes pasan en el sistema	0.33333	0.50000	1.00000	-
Ls: No. Promedio de clientes en el sistema	0.33333	0.50000	1.00000	-

Tabla 3.9. Medidas de desempeño

Conversión de la tasa de servicio a camión/hora para 1 persona:

1 cliente $\longrightarrow$ 60 minutos

μ $\longrightarrow$ 60 minutos = 1 hora

μ = 1 camión/hora

No se debe considerar una persona, ya que es imposible trabajar al 100% de utilización porque se deben dar algunas demoras.

No. de cajas	Cs*Ls	Cc*c	Costo Total ($/hora)
4	30*0.33333	20*4	90.00
3	30*0.50000	20*3	75.00
2	30*1.00000	20*2	70.00

Tabla 3.10. Costo total mínimo por hora

El gerente debe optar por un grupo de 2 personas para lograr minimizar costos hasta $70 por hora.

Modelo del nivel de aceptación

Este modelo analiza las características de la operación del sistema para decidir sobre los valores óptimos de los parámetros del diseño. El nivel o los límites de aceptación lo define la persona que conozca el sistema y busque equilibrar el tiempo promedio de espera en el sistema (Ws) y el porcentaje X de tiempo inactivo de los servidores.

3.7 Aplicaciones

La teoría de colas o líneas de espera son de gran utilidad en áreas como la ingeniería, ya que se pueden modelar sistemas en los que hay una demanda de servicio en el que se llega a un mismo servidor y se pueden registrar esperas desde que se llega al sistema y el servidor realiza la atención. Otros procesos que se pueden modelar son la llegada de datos a una red de computadores, la implementación de una cadena productiva en una industria y la información solicitada en los motores de búsqueda como Google que se puede concebir como una demora a causa de la congestión en la red.

Resumen

La teoría de colas es el estudio matemático de las líneas de espera. La formación de colas es un fenómeno común que se presenta siempre que la demanda efectiva de un servicio excede a la oferta efectiva.

Para las empresas es imprescindible tomar decisiones respecto a los servicios que quiere ofrecer y cómo hacer para lograrlo. No obstante, muchas veces es imposible calcular con certeza cuándo llegarán los clientes que demanden un servicio y/o cuánto tiempo será necesario disponer para dar ese servicio. Ahora bien, que una empresa esté lista para ofrecer cualquier servicio que los clientes soliciten en cualquier momento puede implicar mantener recursos ociosos y costos elevados. Pero, por otro lado, no disponer de la capacidad de servicio suficiente puede conllevar que se presenten colas excesivamente largas en ciertos momentos. Cuando los clientes tienen que esperar en una cola para ser atendidos y conseguir lo que necesitan, están obligados a pagar un costo en tiempo, más alto del que esperaban y así mismo sucede con las empresas. Las líneas de espera largas se vuelven costosas para las empresas, ya que producen pérdida de prestigio y pérdida de clientes.

Por eso, la teoría de colas viene a ser una herramienta fundamental, ya que contribuye con la información vital que se requiere para la toma de decisiones, especialmente las relacionadas en predecir algunas características sobre las líneas de espera.

Bibliografía

- Azarang, M. Simulación y Análisis de Modelos Estocásticos. Líneas de Espera.
- Bronson, R. (1993). Investigación de operaciones. México, Ed McGraw-Hill.
- Chediak, F. (2005). Investigación de operaciones. Colombia, Ed El Poira
- Izar, J. (2012). Investigación de operaciones. México, Ed Trillas.
- K Roscoe, D. (1984), Modelos Cuantitativos para Administración. México, Editorial Iberoamérica.
- Lieberman, G. (2002), Investigación de operaciones. México, Ed McGraw-Hill.
- Taha, H.(1998). Investigación de operaciones. México, Ed Alfaomega.
- Winston, W. (2005). Investigación de operaciones. México, Editorial Thomson.
- Winston, W. (1997). Investigación de Operaciones, Aplicaciones y Algoritmos. México, Grupo Editorial Iberoamericana.

Referencias de Internet

- http://www.lcc.uma.es/~av/Libro/CAP5.pdf
- http://www.andrew.cmu.edu/user/mgoic/files/documents/optimization/pdinamica.pdf
- http://cms.dm.uba.ar/materias/1ercuat2009/optimizacion/Maurette_Ojea.pdf
- http://www.angelfire.com/oz/rubincelis/Pdinamica.pdf
- http://www.unamerida.com/archivospdf/337%20Lectura6.2.pdf