

# Unidad V: Líneas de Espera

## 5.1 Definiciones, características y suposiciones

El problema es determinar que capacidad o tasa de servicio proporciona el balance correcto. Esto no es sencillo, ya que el cliente no llega a un horario fijo, es decir, no se sabe con exactitud en que momento llegarán los clientes. También el tiempo de servicio no tiene un horario fijo.

### Definición.

Una Cola es una línea de espera y la teoría de colas es una colección de modelos matemáticos que describen sistemas de líneas de espera particulares o sistemas de colas. Los modelos sirven para encontrar el comportamiento de estado estable, como la longitud promedio de la línea y el tiempo de espera promedio para un sistema dado. Esta información, junto con los costos pertinentes, se usa, entonces, para determinar la capacidad de servicio apropiada.

## 5.2 Terminología y notación

Estos son algunos conceptos que se utilizan en los modelos de líneas de espera:

**DISCIPLINA DE SERVICIO:** La disciplina de servicio se refiere al orden en el que se seleccionan los clientes de la cola para recibir el servicio. Por ejemplo, puede ser:

**PEPS** (Primero en Entrar Primero en Salir, también conocida como **FIFO**: first in firstout) primero en entrar, primero en salir, según la cual se atiende primero al cliente que haya llegado de primero.

UEPS (Ultimo en Entrar Primero en Salir, también conocida como LIFO: last in firstout) también se conoce como pila, consiste en atender primero al cliente que ha llegado de último.

SOA (Servicio en Orden Aleatorio, también se conoce como SIRO o RSS: randomselection of service) que selecciona los clientes de manera aleatoria, de acuerdo a algún procedimiento de prioridad o a algún otro orden.

PRIORIDAD en el servicio: Los clientes se atienden de primero de acuerdo a alguna prioridad especificada. Si se forma alguna cola con prioridad, seguirá alguna disciplina de servicio.

Procesamiento equilibrado o igualitario, también llamado ProcessorSharing: Sirve a todos los clientes por igual. La capacidad del sistema se comparte entre los clientes y todos experimentan el mismo retraso.

#### TERMINOLOGÍA Y NOTACIÓN:

No se permite el faltante. Suposiciones:

1. La demanda tiene que ser constante.
2. Los costos son constantes (no se permite descuento en adquisiciones voluminosas).
3. Los proveedores entregaran con puntualidad los pedidos en el periodo comprendido.
4. El lote mínimo es igual al inventario máximo.

Nomenclatura:

Q = tamaño económico del lote. Si es muy grande o muy chico

N = número de pedido. Puedes pedir una o 2 veces

D = Demanda. Por si las dudas ten cuidado y siempre papelito habla

Ci = Costo de compra. A lo mejor te sale mas barato en otro lado

Ch = Costo de mantener un unidad en los inventarios (%).

Co = Costo de ordenar. Ya ves que hay gandallas que te cobran el envío

R = Punto de reorden.

L = Tiempo de consumo. En menos de 30 min si te cobran sino PS no

T = Tiempo para consumir el inventario máximo. el tiempo en el que te atragantas

Imáx = Inventario Máximo.

Î = Inventario Promedio.

Ct = Costo Total.

Ct = Costo de compra + Costo de ordenar + Costo de tenencia.

Costo de compra = CiD

### **5.3 Proceso de nacimiento o muerte**

Este informe tiene como finalidad presentar una teoría operacional sobre la Teoría de Colas, la cual incluye el estudio matemático de las colas o líneas de espera, siendo la de mayor aplicación potencial y sin embargo es la más difícil de aplicar. Los fenómenos de espera para recibir servicio son cosas de la vida diaria; por ejemplo, esperar en una cola para pagar el teléfono o en el supermercado. No obstante, la espera no solo se limita a personas sino a procedimientos o ensamblados de máquinas, por lo tanto en esta unidad se describen modelos matemáticos aplicables a cualquier situación donde se forme una cola.

No pretendo incluir en un solo tema todo lo que necesita saber el estudiante, sino ofrecer un marco de los conocimientos básicos presentados en forma clara y

precisa.

La formación de líneas de espera es, por supuesto, un fenómeno común que ocurre siempre que la demanda actual de un servicio excede a la capacidad actual de proporcionarlo. Con frecuencia, en la industria y en otros sitios, deben tomarse decisiones respecto a la cantidad de capacidad que debe proporcionarse. Sin embargo, muchas veces es imposible predecir con exactitud cuándo llegarán las unidades que buscan el servicio y/o cuánto tiempo será necesario para dar ese servicio; es por esto que esas decisiones suelen ser difíciles. Proporcionar demasiado servicio implica costos excesivos. Por otro lado, carecer de la capacidad de servicio suficiente causa colas excesivamente largas en ciertos momentos. Las líneas de espera largas también son costosas en cierto sentido, ya sea por un costo social, por un costo causado por la pérdida de clientes, por el costo de empleados ociosos o por algún otro costo importante. Entonces, la meta final es lograr un balance económico entre el costo de servicio y el costo asociado con la espera por ese servicio. La teoría de colas en sí no resuelve directamente este problema, pero contribuye con información vital que se requiere para tomar las decisiones concernientes prediciendo algunas características sobre la línea de espera como el tiempo de espera promedio.

La teoría de colas proporciona un gran número de modelos matemáticos para describir una situación de línea de espera. Con frecuencia se dispone de resultados matemáticos que predicen algunas de las características de estos modelos.

Como ejemplo prototipo expondré la sala de emergencia del Hospital General, el cual proporciona cuidados médicos rápidos a los casos de emergencia que llegan en ambulancia o vehículos particulares. En cualquier momento se cuenta con un doctor de guardia. No obstante, a causa de la mala situación económica que vive nuestro país existe una creciente tendencia a usar estas instalaciones para casos de emergencia en lugar de ir a una clínica privada, es por ello que el hospital ha

venido experimentando un aumento continuo en el número de pacientes anuales que llegan a la sala de emergencia. Como resultado, es bastante común que los pacientes que llegan durante las horas pico (temprano en la tarde) tengan que esperar turno para recibir el tratamiento del doctor. Por esto, se ha hecho una propuesta para asignar un segundo doctor a la sala de emergencia durante esas horas pico, para que se puedan atender dos casos de emergencia al mismo tiempo. Se ha pedido al ingeniero administrador del hospital que estudie esta posibilidad.

#### 5.4 Modelos Poisson

Para una única variable independiente  $X$ , es un modelo de la forma:  $\ln \lambda = a_0 + a_1 X$ , para simplificar la notación donde  $\ln$  significa logaritmo neperiano,  $a_0$  y  $a_1$  son constantes y  $X$  una variable que puede ser aleatoria o no, continúa o discreta. Este modelo se puede fácilmente generalizar para  $k$  variables independientes:

Por lo tanto  $a_0$  es el logaritmo de  $\lambda$  (probabilidad de que ocurra un evento en un intervalo de tamaño unidad) cuando todas las variables independientes son cero, y  $a_i$  es el cambio en el logaritmo de  $\lambda$  (o logaritmo del cociente de  $\lambda$ ) cuando la variable  $X_i$  aumenta una unidad, manteniéndose constantes las demás, dicho de otro modo, es la probabilidad de que ocurra un evento en un intervalo unidad cuando todas las variables independientes son cero y  $\lambda$  el cociente de dicha probabilidad para un aumento de una unidad en la variable  $X_i$  (riesgo relativo). Obsérvese que, al igual que en la regresión logística, el modelo supone efectos multiplicativos, es decir, si la variable  $X_i$  aumenta  $n$  unidades, la probabilidad para la variable de Poisson se multiplica por es decir, la potencia  $n$ -ésima

#### **5.4.1 Un servidor**

Existen una gran variedad de modelos para los sistemas de colas, las dos características más importantes serán:

a) Los tiempos de llegada. b) Los tiempos de servicio.

En los sistemas de colas reales no es posible determinar con exactitud estos dos tiempos, es decir no son determinísticos, los más comunes son los modelos probabilísticos, donde se dan un promedio de estos tiempos, por lo tanto tenemos que usar una distribución de probabilidad que se ajuste lo más cercano a la realidad. Para calcular la probabilidad de cuál será el tiempo entre llegadas se utiliza la distribución exponencial, esta distribución tiene una función de densidad de probabilidad: (densidad de probabilidad continua)

#### **5.4.2 Múltiples servidores**

### **5.5 Análisis de costos**

Los costos en que se incurre al realizar una actividad, un proyecto, una empresa, se le nombra costos pertinentes, que son los costos directos de una elección, y son de dos tipos: los costos fijos que están disociados de la producción, esto es, se presentan independientemente de la escala productiva, produzca o no, tales como la renta o pagos de arrendamiento de un automóvil. Los costos que varían en proporción directa a la escala de producción de la empresa se les llaman costos variables.

En el corto plazo la función de costos es una relación lineal del tipo:

$$C_t = C_f + C_v * Q$$

En el largo plazo la función se expresa como una función potencia:

$$C_t = aQ^b$$

donde a y b son parámetros estimados y Q es la escala de producción realizada.

Cuando el valor de b es mayor a 1 describe costos marginales crecientes positivos,

Cuando el valor de b es igual a 1 los costos son constantes positivos,

Cuando el valor de b es mayor a 0 y menor a 1 los costos comportan incrementos marginales decrecientes positivos.

Al realizar una elección y trabajar en ella los costos pertinentes implican la renuncia implícita de otras posibles alternativas, las cuales se ordenan de forma transitiva, es decir de la mejor elección a la menor, la renuncia de la alternativa inmediata a la elegida es el costo de oportunidad.

El costo pertinente más el costo de oportunidad dan el costo económico que es el implícito de toda elección. El costo económico es un tema de reflexión necesario en la toma de decisiones, sobre todo en la evaluación o valoración de un proyecto de inversión.

Todo lo anterior es un análisis del costo total y sus componentes, así como sus implicancias económicas en una cartera de inversión. Ahora bien interesa el enfoque relativo de los costos, especialmente relativo a los marginales, puesto que los costos marginales de los insumos o factores productivos implicados establecen el criterio de combinación óptima eficiente de los mismos.

