

Contents

1	Teoria de Colas	2
1.1	Introducción	2
1.1.1	Conceptos basicos	2
1.1.2	Notación Kendal	6
1.1.3	Preguntas y respuestas	7

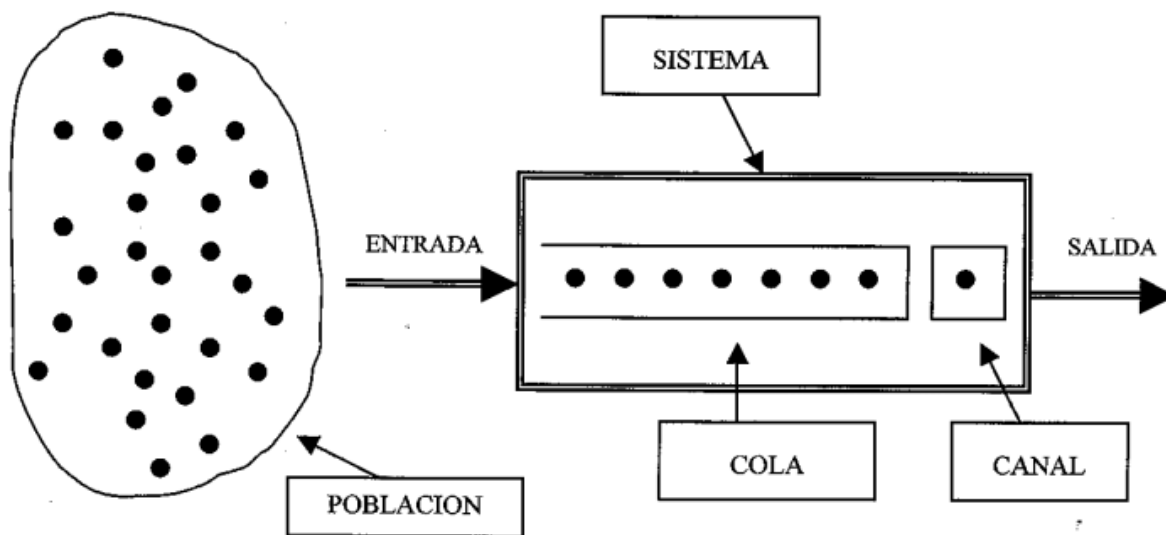
1 Teoria de Colas

1.1 Introducción

1.1.1 Conceptos basicos

Los fenómenos de *congestión* o *espera* estan relacionados con los sistemas estocasticos y pueden describirse como sistemas integrados por uno o mas *centros de atención* donde se brinda un servicio. Cada centro de atención es, a su vez, un sistema constituido por:

- *canales (ó servidores)*: Entidades que prestan el servicio.
- *clientes (ó usuario)*: Entidades que reciben el servicio.



Las colas se forman cuando la demanda de un servicio dado en un intervalo de tiempo excede la capacidad para proveerlo. El administrador del sistema debe establecer un balance apropiado entre los costos asociados a la espera de los usuarios y los costos vinculados con la mejora del servicio (mas servidores, mayor velocidad de atención, etc). En la mayoría de los procesos de atención, los tiempos entre arribos de clientes y los tiempos de los servicios no son predecibles. En estas condiciones se aplica la denominada *teoría de colas* para determinar el comportamiento del sistema bajo diferentes alternativas. Se estudiaran aquellos sistemas que describan los procesos mas generales y que son los que pueden formularse como *cadenas markovianas de primer orden*. Se analizaran los sistemas en régimen permanente, a través de variables tales como la *longitud promedio de la cola*, el *tiempo de espera promedio* del cliente para recibir el servicio, el *tiempo de permanencia* en el sistema, etc. Esta información, juntamente con los costos relevantes, permitira al directivo determinar los valores apropiados de las variables de decisión. Las variables de decisión tipicas en los sistemas de colas estan referidas a la *capacidad de servicio* (numero de canales, velocidad de canales) ó la *capacidad de espera* (número de lugares).

La **población** es el conjunto de usuario potenciales del sistema. Puede ser finito o infinito.

En los **arribos** la llegada de los clientes puede ser deterministica o aleatoria. A menudo los intervalos entre llegadas son estadisticamente independientes y estacionarios a lo largo de prolongados periodos de tiempo, por lo que se puede suponer poissonianos.

SISTEMAS	EJEMPLO	CLIENTES	CANALES
TRANSPORTE	<ul style="list-style-type: none"> • PUERTOS • AEROPUERTOS • AUTOPISTAS • TERMINALES • CARGA Y DESCARGA 	BARCOS AVIONES AUTOMÓVILES ÓMNIBUS CAMIONES	MUELLES PISTAS CASILLAS DE PEAJE PLATAFORMA ISLAS
COMPUTACIÓN	<ul style="list-style-type: none"> • PROCESAMIENTO • IMPRESIÓN 	TRABAJOS TRABAJOS	CPU IMPRESORAS
TELEFONÍA	<ul style="list-style-type: none"> • CENTRALES 	LLAMADOS	LÍNEAS
PRODUCCIÓN	<ul style="list-style-type: none"> • ELABORACIÓN • REPARACIÓN • CONTROL DE CALIDAD 	PRODUCTOS SOLICITUDES PRODUCTOS	MAQUINAS OPERARIOS INSPECTORES
ATENCIÓN AL PÚBLICO	<ul style="list-style-type: none"> • BANCOS • SUPERMERCADOS • ESTACIONES DE SERVICIO • NEGOCIOS • ATENCIÓN MÉDICA • ALQUILER DE AUTOS 	PERSONAS PERSONAS AUTOMÓVILES PERSONAS PACIENTES PERSONAS	CAJEROS CAJEROS SURTIDORES EMPLEADOS AMBULANCIAS EMPLEADOS
ADMINISTRACIÓN	<ul style="list-style-type: none"> • DPTO. COMPRAS • JUZGADOS 	SOLICITUDES CAUSAS	COMPRADORES JUECES

TABLA 1.1 Ejemplos de sistemas de colas

La **impaciencia** se verifica cuando algunos usuarios que arriban al sistema se retiran sin recibir el servicio porque consideran que el tiempo de espera sera suficientemente largo. Se distinguen dos tiempos de impaciencia:

1. Un cliente que arriba, observa la cantidad de gente que esta delante de él esperando y en función de ello toma la decisión de incorporarse o no al sistema.
2. Un cliente que arriba, ingresa al sistema y al cabo de un tiempo toma la decisión de seguir esperando o no.

La **capacidad** es el número máximo de clientes que puede permanecer en el sistema simultaneamente(en espera y atendiéndose).

Para el **modo de arribo** los usuarios pueden llegar en forma individual o en masa(modos batch). En la mayoría de los sistemas que estudiaremos se hará la suposición de que los procesos de llegada son del tipo Poisson, lo que implica *arribos individuales*. Se puede considerar el arribo de grupos como clientes individuales.

Para la **prioridad de atención**, existen diversos criterios de atención en lo que se refiere al orden de selección de clientes para brindar el servicio. Ellos son:

- Base FIFO(first in, first out): Los clientes se atienden según el orden de llegada.
- Base LIFO(last in, first out): El último individuo que arriba es el primero en ser atendido.
- Base SIRO(service in random order): Es una selección aleatoria de los clientes para brindales el servicio.
- Base con PRIORIDADES: Se establecen criterios de atención conforme a los atributos de los clientes.

La **duración del servicio** es el tiempo requerido por un canal para atender un cliente. Puede ser una variable determinística o aleatoria con distribución de probabilidad conocida.

En el **modo de atención** un canal puede servicio de forma individual o múltiples(en masa). En la mayoría de los sistemas reales, el modo de atención es individual.

Los **procesos poisson** son markovianos y tienen dos distribuciones que lo describen:

1. La distribución Poisson, en donde la variable es el número de eventos que se producen en un intervalo determinado de continuo.
2. La distribución Gamma, en la que la variable es el intervalo de continuo necesario para que se verifique un número determinado de eventos. Particularmente, cuando la variable es el intervalo de tiempo de continuo necesario para que se verifique *un solo evento*, la distribución es conocida como **distribución Exponencial**.

Distribución Poisson La probabilidad de que se produzcan "n" eventos en un intervalo "t" está dada por:

$$p(n) = \frac{(\lambda \Delta t)^n \cdot e^{-\lambda \Delta t}}{n!} \quad (1)$$

siendo $n = 1, 2, \dots$ y $\lambda > 0$ La media de esta distribución es $a = \lambda \cdot t$ y el desvío estándar $\sigma = \sqrt{\lambda t}$.

Distribución Gamma La función distribución de probabilidad de la **distribución exponencial** está dada por la siguiente expresión:

$$f(t) = \lambda \cdot e^{-\lambda t} \quad (2)$$

cuya media es $\frac{1}{\lambda}$ y cuyo desvío estándar es $\frac{1}{\lambda}$

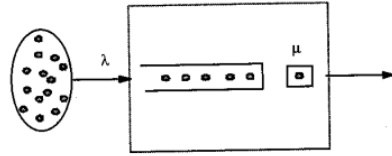
Si el proceso de arribos es de tipo Poisson significa que la variable "tiempo entre dos arribos sucesivos" tiene distribución *exponencial* y la variable "numero de clientes que arriban por unidad de tiempo" tiene distribución *Poisson*.

Ingresos y egresos de clientes: En los sistemas de capacidad finita o de población impaciente, no todos los clientes que arriban al sistema ingresan.

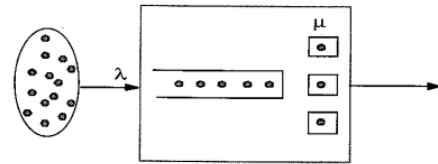
- $\bar{\lambda}$: Número promedio de clientes que ingresan efectivamente al sistema.
- \bar{R} : Número promedio de rechazados(es decir, que no ingresan al sistema).
- $\bar{\mu}$: Número promedio de clientes atendidos que egresan del sistema.
- \bar{A} : Tasa de clientes que ingresaron al sistema pero que decidieron abandonarlo sin recibir el servicio.

Estructuras de sistemas simples:

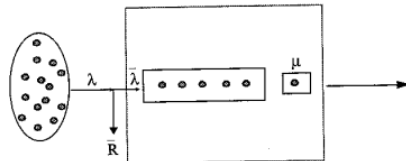
1. Cola simple, capacidad infinita, un canal de atención:



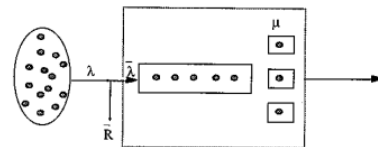
3. Cola simple, capacidad infinita, canales múltiples en paralelo:



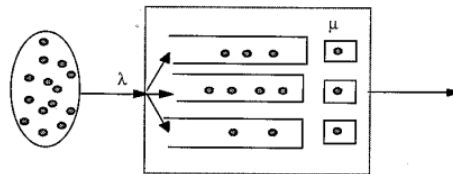
2. Cola simple, capacidad finita, un canal de atención:



4. Cola simple, capacidad finita, canales múltiples en paralelo:



5. Varias colas, una para cada canal dispuesto en paralelo:



1.1.2 Notación Kendal

Especifica las características descriptivas de una unidad operativa de un sistema de colas. Es una notación de 6 posiciones:

$$1/2/3/4/5/6 \quad (3)$$

1. La posición 1 se refiere al patrón de arribos al sistema. Puede ser:
 - P: Proceso de Poisson.
 - D: Proceso determinístico.
 - G: Cualquier otro proceso.
2. La posición 2 indica el patrón de servicio en los canales.
 - P: Proceso de Poisson.
 - D: Proceso determinístico.
 - G: Cualquier otro proceso.
3. La posición 3 indica el número de canales de atención dispuestos en paralelo en la unidad operativa.
4. La posición 4 indica la capacidad de la unidad operativa del sistema. Se asume infinita y no se indica.
5. La posición 5 indica la prioridad de atención de la cola.
 - FIFO
 - LIFO
 - SIRO
 - G: Cualquier otra modalidad de atención
6. La última posición se refiere al tamaño de la población. Si no se indica, el tamaño es infinito.

1.1.3 Preguntas y respuestas

- ¿Qué características posee la población?
- ¿En qué consiste el fenómeno de impaciencia? ¿cuántos tipos de impaciencia hay? ¿en qué consisten?
- ¿Qué características posee el sistema?
- ¿Qué se entiende por capacidad del sistema?
- ¿Con qué notación identificamos cada uno de los ítems enunciados?
 - L_c : La cantidad promedio de clientes que están esperando para recibir el servicio en un determinado momento.
 - L : La cantidad promedio de clientes que se encuentran en el sistema en un determinado momento, ya sea esperando ser atendidos como atendiéndose.
 - W_c : El tiempo promedio que un cliente debe esperar para ser atendido.
 - W : El tiempo promedio que un cliente permanece en el sistema, ya sea esperando ser atendido como atendiéndose.
 - λ : La cantidad promedio de clientes que arriban al sistema en un determinado momento.
 - $\bar{\lambda}$: la cantidad promedio de clientes que ingresan al sistema en un determinado momento.
 - $\bar{\mu}$: La cantidad promedio de clientes que egresan del sistema luego de ser atendidos.
 - μ : La velocidad promedio de atención de un canal.
 - T_a : El tiempo promedio entre arribos.
 - H : La cantidad promedio de canales ocupados.
 - PA : El porcentaje de actividad de cada canal.
- ¿Cuál es la diferencia entre “la cantidad promedio de clientes que arriban al sistema en un determinado momento” y “la cantidad promedio de clientes que ingresan al sistema en un determinado momento”?
- ¿Qué significa que se encuentren en régimen permanente o estacionario?
- ¿qué se indica cada una de las posiciones de la notación Kendall?
- ¿Qué características posee el modelo $P/P/1$?
- ¿Cuánto debe valer ρ en un $P/P/1$? ¿Por qué?
- Según el modelo $P/P/1/N$ ¿Cuáles son sus características según la notación de Kendall?
- En el modelo $P/P/1/N$, ¿debe verificarse lo mismo que en el $P/P/1$ respecto del valor de ρ ? ¿Por qué?
- ¿Qué ejemplos de la vida cotidiana se podrían plantear en cada uno de los modelos mencionados?