

Modelos y Simulación

TEORIA DE COLAS

TEORIA DE COLAS

- Las **LINEAS DE ESPERA, FILAS DE ESPERA o COLAS**, son realidades cotidianas:
 - **Personas esperando** para realizar sus transacciones ante una caja en un banco,
 - **Estudiantes esperando** por obtener copias en la fotocopidora,
 - **Automóviles esperando** pagar ante una estación de peaje o continuar su camino, ante un semáforo en rojo,
 - **Máquinas dañadas** a la espera de ser reparadas.

Se forman debido a un desequilibrio temporal entre la demanda del servicio y la capacidad del sistema para suministrarlo.

TEORIA DE COLAS

- Los **Modelos de Líneas de Espera** son de gran utilidad tanto en las áreas de **Manufactura** como en las de **Servicio**.
- Los **Análisis de Colas relacionan:**
 - la **longitud de la línea de espera**,
 - el **promedio de tiempo de espera**
y otros factores como:
 - la **conducta de los usuarios** a la llegada y en la cola,

Los **Análisis de Colas** ayudan a *entender el comportamiento* de estos sistemas de servicio (la atención de las cajeras de un banco, actividades de mantenimiento y reparación de maquinaria, el control de las operaciones en planta, etc.).

TEORIA DE COLAS

- Desde la **perspectiva** de la Investigación de Operaciones, los pacientes que esperan ser atendidos por el odontólogo o las prensas dañadas esperando reparación, tienen mucho en común.
- Ambos (gente y máquinas) **requieren de recursos humanos y recursos materiales como equipos** para que se los cure o se los haga funcionar nuevamente.

TEORIA DE COLAS

Costos de Servicio y Costos de Espera

- Los Administradores reconocen el **equilibrio** que debe haber entre el **COSTO DE** proporcionar buen **SERVICIO** y el **COSTO** del tiempo **DE ESPERA** del cliente o de la máquina que deben ser atendidos.
- Los Administradores desean que las colas sean lo **suficientemente cortas** con la finalidad de que los clientes no se enojen e incluso se retiren sin llegar a utilizar el servicio o lo usen pero no retornen más.
- Sin embargo los Administradores contemplan tener una longitud de cola **razonable en espera**, que sea balanceada, para obtener ahorros significativos en el **COSTO DEL SERVICIO**

TEORIA DE COLAS

Costos de Servicio vs Nivel de Servicio

- Los ***COSTOS DE SERVICIO*** se incrementan si se mejora el ***NIVEL DE SERVICIO***. Los Administradores de ciertos centros de servicio pueden **variar su capacidad** teniendo personal o máquinas adicionales que son asignadas a **incrementar la atención** cuando crecen excesivamente los clientes.
 - En supermercados se habilitan **cajas adicionales** cuando es necesario.
 - En bancos y puntos de revisión/chequeo de documentos en los aeropuertos, se contrata **personal adicional** para atender en ciertas épocas del día o del año.

TEORIA DE COLAS

- Cuando el **servicio mejora**, disminuye el **costo de tiempo perdido** en las líneas de espera.
- Este costo puede reflejar **pérdida de productividad** de los operarios que están esperando que compongan sus equipos o puede ser simplemente un estimado de los **clientes perdidos** a causa de mal servicio y colas muy largas.
- En ciertos servicios el **costo de la espera puede ser intolerablemente alto**.

TEORIA DE COLAS

COLAS MAS COMUNES		
SITIO	ARRIBOS EN COLA	SERVICIO
Supermercado	Compradores	Pago en cajas
Peaje	Vehículos	Pago de peaje
Consultorio	Pacientes	Consulta
Sistema de Cómputo	Programas a ser corridos	Proceso de datos
Compañía de teléfonos	Llamadas	Efectuar comunicación
Banco	Clientes	Depósitos y Cobros
Mantenimiento	Máquinas dañadas	Reparación
Muelle	Barcos	Carga y descarga

TEORIA DE COLAS

Características de una LINEA DE ESPERA

- Una cola de espera está compuesta de tres elementos:
 1. Arribos o ingresos al sistema
 2. Disciplina en la cola
 3. Servicio
- Estos tres componentes tienen ciertas características que deben ser examinadas antes de desarrollar el aspecto matemático de los modelos de cola.

TEORIA DE COLAS

Características de una LINEA DE ESPERA

- **1. CARACTERISTICAS DE ARRIBO:**
- La fuente de ingreso que genera los arribos o clientes para el servicio tiene tres características principales:
 - a. Tamaño de la población que arriba
 - b. Patrón de llegada a la cola
 - c. Comportamiento de las llegadas.

1.a. Tamaño de la Población:

El tamaño de la población puede ser:

**infinito (ilimitado) o
Finito (limitado).**

TEORIA DE COLAS

Características de una LINEA DE ESPERA

1. CARACTERISTICAS DE ARRIBO:

1.a. Tamaño de la Población:

Infinito (ilimitado): Cuando el número de clientes o arribos en un momento dado es una pequeña parte de los arribos potenciales. Para propósitos prácticos poblaciones ilimitadas pueden considerarse: los vehículos que se acercan a una estación de peaje, los hinchas que van a una cancha, los clientes en un supermercado.

LA MAYORÍA DE LOS MODELOS ASUME ARRIBO INFINITO.

Población de arribo limitada o finita: cuando se tienen muy pocos servidores y el servicio es restringido. Ej.: los pacientes en un consultorio médico.

TEORIA DE COLAS

Características de una LINEA DE ESPERA

1. CARACTERISTICAS DE ARRIBO:

- 1.b. Patrón de arribo a la cola:
 - Los clientes arriban a ser atendidos de una manera **programada** (un paciente cada 15 minutos) o de una manera **aleatoria**.
 - Se consideran que los arribos son aleatorios cuando éstos son independientes de otros y su ocurrencia no puede ser predecida exactamente.
 - Frecuentemente en problemas de colas, el número de arribos por unidad de tiempo pueden ser estimados por medio de la **Distribución de Poisson** que es una distribución discreta de probabilidad.

TEORIA DE COLAS

Características de una LINEA DE ESPERA

1. CARACTERISTICAS DE ARRIBO:

- **DISTRIBUCION DE POISSON:**

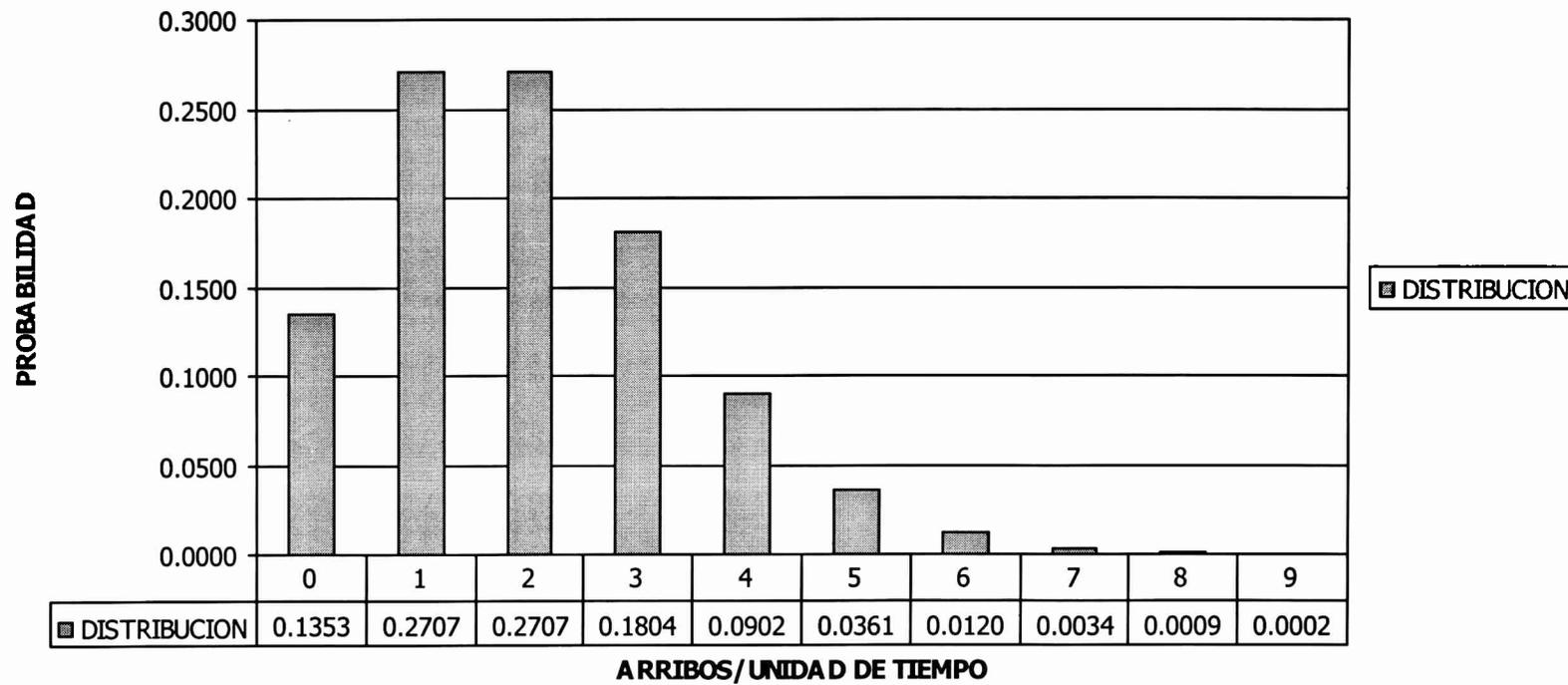
$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!} \text{ para } x = 0, 1, 2, 3, 4, \dots$$

- $P(x)$ = Probabilidad de x arribos
- x = número de arribos por unidad de tiempo
- λ = tasa promedio de arribo
- $e = 2.71828$

TEORIA DE COLAS

DISTRIBUCION DE POISSON

DISTRIBUCION DE POISSON PARA TIEMPOS DE ARRIBO $\lambda = 2$

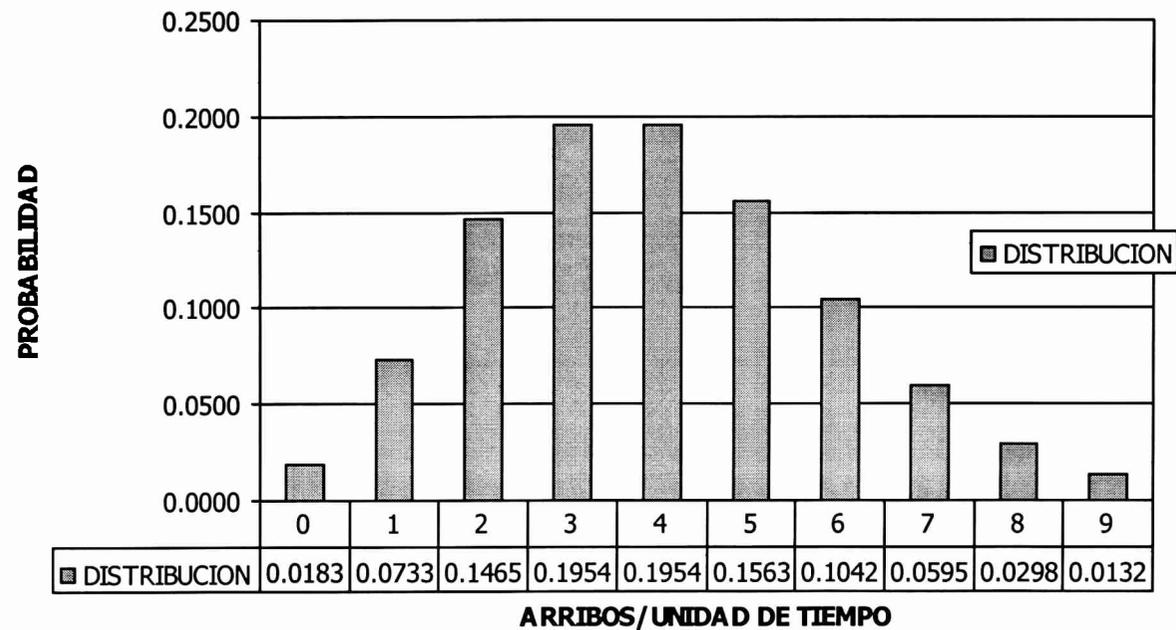


TEORIA DE COLAS

DISTRIBUCION DE POISSON

DISTRIBUCION DE POISSON PARA TIEMPOS DE ARRIBO

$$\lambda = 4$$



TEORIA DE COLAS

Características de una LINEA DE ESPERA

1. CARACTERISTICAS DE ARRIBO:

1.c. Comportamiento de los arribos:

La mayoría de los modelos de colas asume que los clientes son **pacientes** o sea que **esperan en la cola hasta ser servidos** y **no se pasan entre colas**. Desafortunadamente, la vida es complicada y la gente se reniega. Aquellos que se impacientan por la espera, **se retiran** de la cola sin completar su transacción.

Esta situación **sirve para acentuar** el estudio de la teoría de colas y el análisis de las líneas de espera, ya que un cliente no servido es un cliente perdido y hace mala propaganda de ese negocio.

TEORIA DE COLAS

2. CARACTERISTICAS DE LA LINEA DE ESPERA:

- La **LINEA DE ESPERA** es el segundo componente de un sistema de colas. La longitud de la cola puede ser también **LIMITADA** o **ILIMITADA**.
 - Cola **LIMITADA** es aquella que por aspectos físicos no puede incrementarse a tamaños infinitos. Puede ser el caso de una peluquería que tiene pocos peluqueros y sillas para atender.
 - Estudiaremos los modelos de colas asumiendo **colas de longitud infinita**. Una cola es **ILIMITADA** cuando su tamaño no tiene restricción como es el caso de una estación de peaje.

TEORIA DE COLAS

2. CARACTERISTICAS DE LA LINEA DE ESPERA:

- Una segunda característica de las líneas de espera se refiere a la **DISCIPLINA EN LA COLA** mediante la cual los clientes reciben el servicio. La mayoría de los sistemas usan la regla **Primero En Entrar Primero En Salir (First In First Out)**.
- En las áreas de emergencia de hospitales sin embargo se omite esta regla dependiendo de la gravedad de las lesiones de las personas que arriban por auxilio médico.
- En supermercados, personas con menos de 10 artículos tienen la caja express que atiende a este tipo de clientes. Pero en la cola se les atiende con la política FIFO.

TEORIA DE COLAS

CARACTERISTICAS DE LA LINEA DE ESPERA

3. Características del Servicio

El **tercer elemento** de un sistema de colas es el **SERVICIO**. En él son importantes dos propiedades básicas:

1. La **configuración** del sistema de servicio.
2. El patrón de **tiempos de servicio**

3.1. CONFIGURACIONES BASICAS PARA EL SERVICIO:

Los sistemas para el servicio son clasificados en función del número de **canales** (servidores) y el número de **fases** (número de paradas que deben hacerse durante el servicio).

Sistema de cola de un solo canal: tiene un solo servidor. Ejemplos de ello son los cajeros para automovilistas o los establecimientos de comida rápida.

TEORIA DE COLAS

CARACTERISTICAS DE LA LINEA DE ESPERA

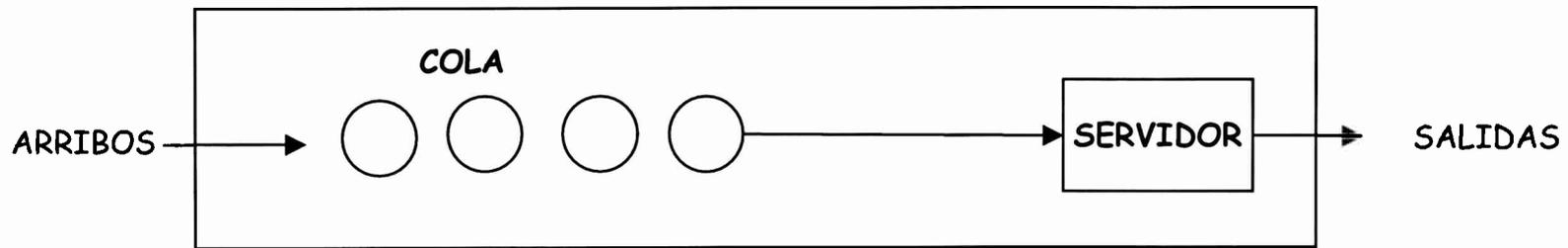
3.1. Configuraciones básicas para el Servicio

- **Sistema de cola multi-canal:** Son principalmente los cajeros de un banco en los cuales hay una sola cola y varias personas atendiendo a los clientes en diversas cajas.
- **Sistema de una sola fase:** es aquel en el cual el cliente recibe el servicio de una sola estación y luego abandona el sistema. Un restaurant de comida rápida en el cual la persona que toma la orden también le entrega el alimento y cobra, es un sistema de una sola fase
- **Sistema multifase:** cuando se pone la orden en una estación, se paga en una segunda y se retira lo adquirido en una tercera

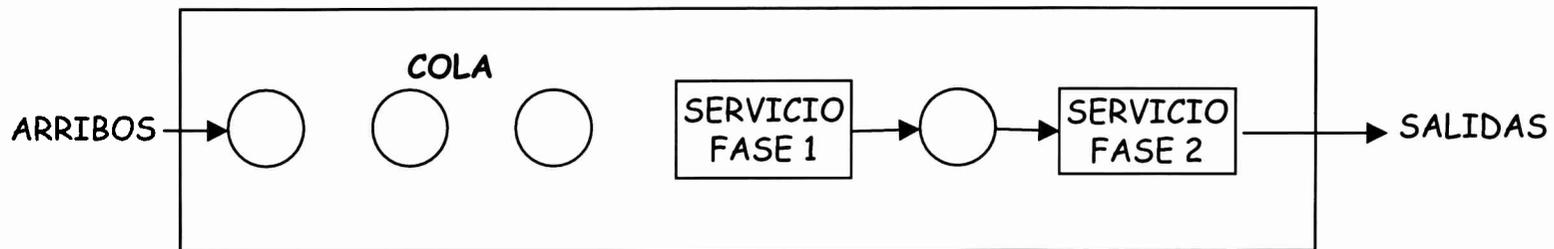
TEORIA DE COLAS

Configuraciones Básicas de Sistemas de Colas

3.1. Configuraciones básicas para el Servicio



SISTEMA UN CANAL, UNA FASE

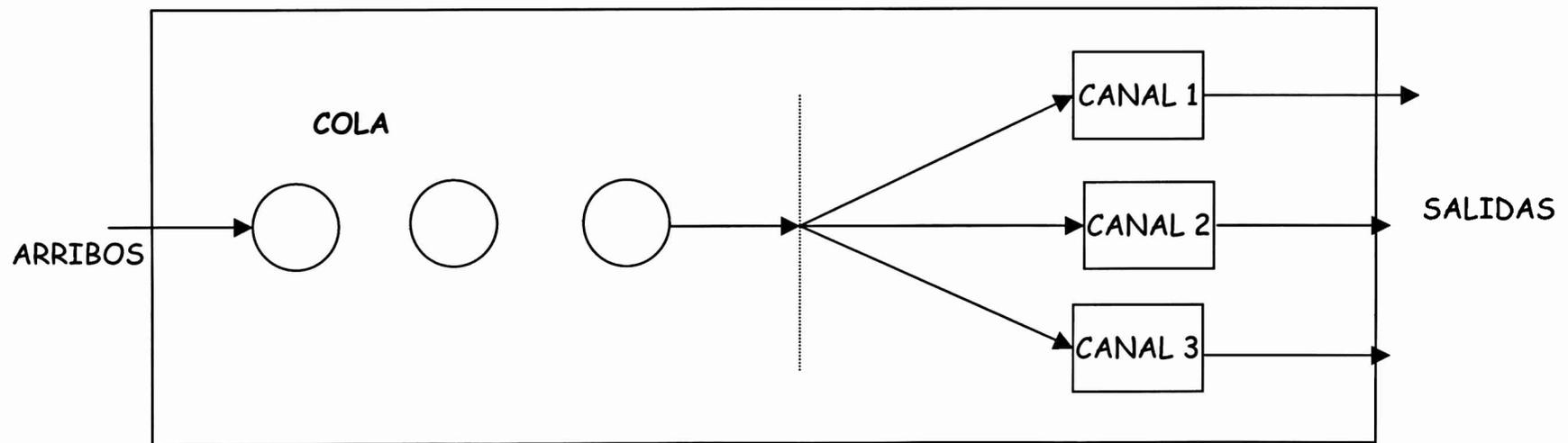


UN SOLO CANAL, MULTIFASE

TEORIA DE COLAS

Configuraciones Básicas de Sistemas de Colas

3.1. Configuraciones básicas para el Servicio

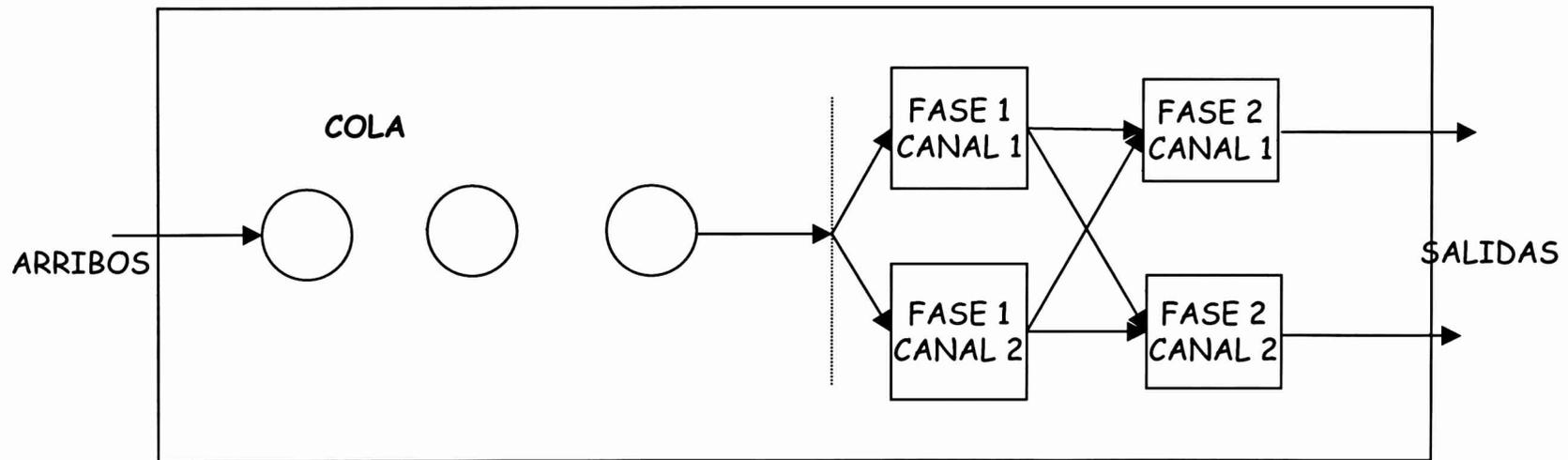


SISTEMA MULTICANAL UNA FASE

TEORIA DE COLAS

Configuraciones Básicas de Sistemas de Colas

3.1. Configuraciones básicas para el Servicio



SISTEMA MULTICANAL MULTIFASE

TEORIA DE COLAS

Configuraciones Básicas de Sistemas de Colas

3.2. Distribución del Tiempo de Servicio

- Los patrones de servicio son similares a los patrones de llegada. Pueden ser constantes o aleatorios.
 - I. Si el tiempo de servicio es constante, toma la misma cantidad de tiempo atender a cada cliente. Es común con servicios dados por medio de máquinas (Lavadora automática de carros).
 - II. Si el tiempo de servicio es distribuído aleatoriamente - que es el caso más común - se lo representa por la **DISTRIBUCION DE PROBABILIDAD EXPONENCIAL NEGATIVA** de la forma $e^{-\mu x}$ para $x \geq 0$. Esta es una hipótesis matemática muy conveniente, cuando los arribos siguen la distribución de Poisson.

TEORIA DE COLAS

Medición del Rendimiento de las Colas

- Los modelos de colas ayudan a los administradores a tomar decisiones para balancear los costos de servicio deseables con los costos de espera en la línea.
- Los principales factores que se evalúan en estos modelos son:
 1. Tiempo promedio que cada cliente u objeto permanece en la cola
 2. Longitud de cola promedio
 3. Tiempo promedio que cada cliente permanece en el sistema (tiempo de espera + tiempo de servicio).
 4. Número de clientes promedio en el sistema.
 5. Probabilidad de que el servicio se quede vacío
 6. Factor de utilización del sistema
 7. Probabilidad de la presencia de un específico número de clientes en el sistema.

TEORIA DE COLAS

Notación de los Modelos de Colas

- Reconociendo la diversidad de los sistemas de colas, Kendall (1953) propuso un **sistema de notación** para sistemas de servidores paralelos que ha sido adoptado universalmente.
- Una versión **resumida** de esta convención está basada en el formato **A/B/c/N/K**. Estas letras representan las siguientes características del sistema:
 - **A** = Distribución de tiempo entre arribos.
 - **B** = Distribución del tiempo de servicio.

Los siguientes son símbolos comunes para **A** y **B**:

M = exponencial o Markov ₍₁₎

D = constante o determinística

c: no de servidores paralelos

N: capacidad del sistema

K: tamaño de la población

TEORIA DE COLAS

Notación de los Modelos de Colas

- E_k = Erlang de orden k
- $P H$ = Tipo fase
- H = Hiperexponencial
- G = Arbitrario o general
- GI = General independiente
- c = número de servidores paralelos
- N = Capacidad del sistema .
- K = Tamaño de la población.

Nota₍₁₎: A causa de las suposiciones de distribución exponencial en los procesos de arribo, estos modelos son llamados **MARKOVIANOS**

de servicios

TEORIA DE COLAS

Notación de los Modelos de Colas

- Por ejemplo: $M/M/1/\infty/\infty$ significa *un solo servidor*, capacidad de *cola ilimitada* y *población infinita* de arribos potenciales. Los tiempos entre arribos y los tiempos de servicio son **distribuidos exponencialmente**.
- Cuando N y K son *infinitos*, pueden ser *descartados* de la notación. $M/M/1/\infty/\infty$ es reducido a $M/M/1$.

TEORIA DE COLAS

Variedad de Modelos de Colas

- Existe una cantidad enorme de Modelos de Colas que pueden utilizarse. Nos vamos a concentrar en **4 de los modelos más usados**. Modelos más complejos pueden ser desarrollados mediante el uso de la Simulación y se los encuentra en textos especializados sobre el tema.
- Los 4 modelos de colas a estudiar asumen:
 - o Arribos según la Distribución de Poisson
 - o Disciplina FIFO
 - o Una sola fase de servicio.
- **Modelo A:** Un canal, Arribos según la Distribución de Poisson; Tiempos de Servicio exponenciales

TEORIA DE COLAS

Variedad de Modelos de Colas

- Modelo B: Multicanal
- Modelo C: Tiempo de Servicio constante
- Modelo D: Población Limitada
- **Modelo A: Modelo de Colas de un solo canal, con arribos que siguen la distribución de Poisson y Tiempos de Servicio Exponenciales: (Modelo M/M/1)**
- Los casos más comunes de problemas de colas incluyen la línea de espera de canal único o servidor único. En este caso los arribos crean una sola cola a ser servida por una sola estación.

TEORIA DE COLAS

Modelo A: M/M/1

- Asumimos que existen las siguientes condiciones:
 1. Los clientes son servidos con una política **FIFO** y cada arribo espera a ser servido sin importar la longitud de la línea o cola.
 2. Los arribos son **independientes** de arribos anteriores, pero el promedio de arribos, **no cambia** con el tiempo.
 3. Los arribos son descritos mediante la distribución de probabilidad de **Poisson** y proceden de una **población muy grande o infinita**.
 4. Los **tiempos de servicio** varían de cliente a cliente y son **independientes** entre sí, pero su ~~ta~~ **ta** **promedio es conocida**.

TEORIA DE COLAS

Modelo A: (M/M/1) - Modelo B: (M/M/S)

5. Los tiempos de servicio se representan mediante la ~~distribución de probabilidad exponencial negativa.~~
6. La rata de servicio es más rápida que la ~~rata~~ de arribo.

~~Tabla 5.3 Render Pág. 192~~

- **Modelo B: Modelo de cola multicanal (M/M/S)**
- **Dos o más servidores** o canales están disponibles para atender a los clientes que arriban.
- Los clientes forman una sola cola y se los atiende de acuerdo al servidor que queda libre.
- Asumimos que los arribos siguen la **distribución de probabilidad de Poisson** y los tiempos de servicio son **distribuidos exponencialmente.**

TEORIA DE COLAS

Modelo B: (M/M/S) Modelo C: (M/D/1)

- Los servicios se los hace de acuerdo a la política *primero en llegar primero en ser servido* (PEPS) y todos los servidores atienden a la misma ^{Fifo} tasa.
- **Modelo C: Modelo de Tiempo de Servicio Constante (M/D/1)**
- Algunos sistemas tienen **tiempos de servicio constantes** en lugar de exponencialmente distribuídos. Cuando los clientes son atendidos o equipos son procesados con un ciclo fijo como es el caso de una lavadora de ~~café~~ automatizada o ciertos entretenimientos en los parques de diversiones, el asumir servicio constante es adecuado.

TEORIA DE COLAS

Modelo D: Población limitada

- **Modelo D: Modelo de Población limitada.-**
- Este modelo puede ser usado por ejemplo si estamos considerando reparaciones de equipo en una fábrica que tiene 5 máquinas. Este modelo **permite cualquier número de reparadores** a ser considerados.
- La razón por la cual este modelo difiere de los otros tres es que ahora hay una relación de **dependencia entre la longitud de la cola y la tasa de arribo**. La situación extrema sería si en la fábrica tenemos 5 máquinas, todas se han dañado y necesitan reparación; siendo en este caso la tasa de arribo **CERO**. En general, si la línea de espera crece, la rata de llegada tiende a cero

RESUMEN DE LOS MODELOS DE COLAS DESCRITOS

MODELO	NOMBRE	Nº DE CANALES	Nº DE FASES	PATRÓN DE ARRIBO	PATRÓN DE SERVICIO	TAMAÑO DE LA POBLACIÓN	DISCIPLINA DE COLA
A	SIMPLE M/M/1	UNO	UNA	POISSON	EXPONENCIAL	INFINITA	FIFO
B	MULTI-CANAL M/M/S	MULTI-CANAL	UNA	POISSON	EXPONENCIAL	INFINITA	FIFO
C	SERVICIO CONSTANTE (M/D/1)	UNO	UNA	POISSON	CONSTANTE	INFINITA	FIFO
D	POBLACION LIMITADA	UNO	UNA	POISSON	EXPONENCIAL	FINITA	FIFO

FÓRMULAS PARA COLAS

MODELO A: SISTEMA SIMPLE O M/M/1

λ = Número promedio de arribos por período de tiempo

μ = Número promedio de gente o cosas servidos por período de tiempo

n = número de unidades en el sistema

L_s = Número promedio de unidades (clientes) en el sistema $L_s = \frac{\lambda}{\mu - \lambda}$ $\frac{\lambda}{\lambda - \mu}$

ρ = Factor de utilización del sistema = $\frac{\lambda}{\mu}$ λ/μ

W_s = Tiempo promedio que una unidad permanece en el sistema =
(tiempo de espera + tiempo de servicio)

$W_s = \frac{1}{\mu - \lambda}$ $\frac{1}{\lambda - \mu}$

FÓRMULAS PARA COLAS

MODELO A: SISTEMA SIMPLE O M/M/1

$$L_q = \text{Número promedio de unidades en la cola} = \frac{\lambda^2}{\mu(\mu - \lambda)} = \rho * L_s$$

$$W_q = \text{Tiempo promedio que una unidad espera en la cola} = \frac{\lambda}{\mu(\mu - \lambda)} = \rho * W_s$$

P_n = Probabilidad de que "n" clientes estén en el sistema =

$$P_n = \left(1 - \frac{\lambda}{\mu}\right) * \left(\frac{\lambda}{\mu}\right)^n = (1 - \rho) * \rho^n$$

P_o = Probabilidad de cero unidades en el sistema (la unidad de servicio está vacía) =

$$P_o = 1 - \frac{\lambda}{\mu} = (1 - \rho)$$

$P_{n>k}$ = Probabilidad de que más de "k" unidades estén en el sistema =

$$P_{n>k} = \left(\frac{\lambda}{\mu}\right)^{k+1}$$

$\lambda - \mu$

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

FÓRMULAS PARA COLAS

MODELO B: SISTEMA MULTICANAL O

M/M/S

M = número de canales abiertos

λ = tasa promedio de arribo

μ = tasa promedio de servicio en cada canal

P_o = Probabilidad de que existan CERO personas o unidades en el sistema =

$$P_o = \frac{1}{\left[\sum_{n=0}^{M-1} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n \right] + \frac{1}{M!} \left(\frac{\lambda}{\mu} \right)^M \frac{M\mu}{M\mu - \lambda}} \text{ para } M\mu > \lambda$$

L_s = número promedio de personas o unidades en el sistema :

$$L_s = \frac{\lambda \mu \left(\frac{\lambda}{\mu} \right)^M}{(M-1)! (M\mu - \lambda)^2} P_o + \frac{\lambda}{\mu}$$

FÓRMULAS PARA COLAS

MODELO B: SISTEMA MULTICANAL O

M/M/S

W_s = Tiempo promedio que una unidad permanece en el sistema,
(en la cola y siendo servida (atendida)) =

$$W_s = \frac{\mu \left(\frac{\lambda}{\mu} \right)^M}{(M-1)!(M\mu - \lambda)^2} P_0 + \frac{1}{\mu} = \frac{L_s}{\lambda}$$

L_q = Número promedio de personas o unidades en la línea o cola, en espera de servicio =

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

W_q = Tiempo promedio que una persona o unidad se
tarda en la cola esperando por servicio =

$$W_q = W_s - \frac{1}{\mu} = \frac{L_q}{\lambda}$$

FÓRMULAS PARA COLAS

MODELO C: SERVICIO CONSTANTE O

MODELO M/D/1

Longitud promedio de la cola, $L_q = \frac{\lambda^2}{2\mu(\mu - \lambda)}$

Tiempo promedio de espera en la cola, $W_q = \frac{\lambda}{2\mu(\mu - \lambda)}$

Número promedio de clientes en el sistema, $L_s = L_q + \frac{\lambda}{\mu}$

Tiempo promedio de espera en el sistema, $W_s = W_q + \frac{1}{\mu}$

FORMULAS PARA COLAS

MODELO D: POBLACIÓN LIMITADA

NOTACIÓN :

D = Probabilidad de que una unidad tenga que esperar en la cola

F = Factor de eficiencia

H = Número promedio de unidades siendo servidas

J = Número promedio de unidades que no están en cola o en el sector de servicio

L = Número promedio de unidades esperando el servicio

M = Número de canales de servicio

N = Número de clientes potenciales

T = Tiempo de servicio promedio

U = Tiempo de servicio entre requerimientos de atención a la unidad

W = Tiempo promedio que una unidad espera en la cola

X = Factor de servicio

FORMULAS PARA COLAS

MODELO D: POBLACIÓN LIMITADA

FÓRMULAS

Factor de Servicio.....	$X = \frac{T}{T + U}$
Número promedio en espera	$L = N(1 - F)$
Tiempo promedio de espera	$W = \frac{L(T + U)}{N - L} = \frac{T(1 - F)}{XF}$
Número promedio en funcionamiento	$J = NF(1 - X)$
Número promedio siendo servido	$H = FNX$
Cuántía de la Población	$N = J + L + H$