

Teoría de Colas

Profesor:

Braulio Bernard Maltés (braulio.bernard@gmail.com)

COLAS MAS COMUNES		
SITIO	ARRIBOS EN COLA	SERVICIO
Supermercado	Compradores	Pago en cajas
Peaje	Vehículos	Pago de peaje
Consultorio	Pacientes	Consulta
Sistemas computacionales	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

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.

1. CARACTERISTICAS DE ARRIBO:

La **fente 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 limitado (finito).

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 a los vehículos que se acercan a un caseta de peaje, los aficionados a un partido del mundial de Fútbol, 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

1. CARACTERISTICAS DE ARRIBO:

1.b. Patrón de arribo al sistema:

- ▶ 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.

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 = razón promedio de arribo

$e = 2.71828$

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.

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 barberos 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 caseta de peaje que sirve a los vehículos que arriban.

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)** [PEPS (FIFO)]. Se denomina también FIFS (**First In First Served**).

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 PEPS.

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:

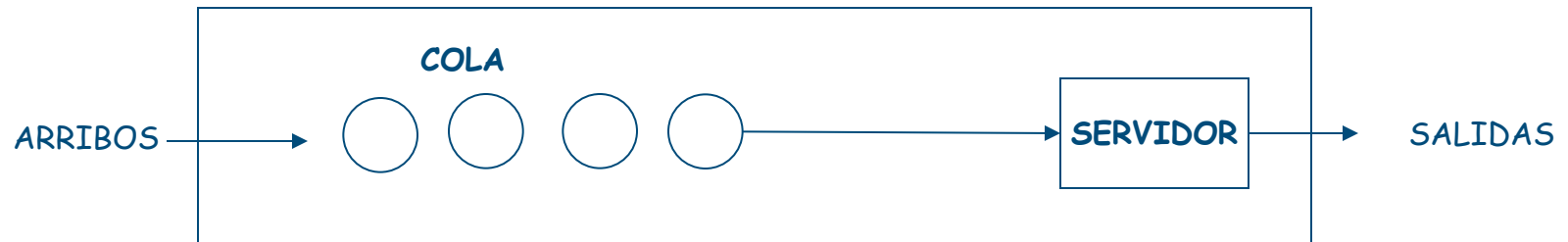
Los sistemas para el servicio son clasificados en función del numero de **canales** (servidores) y el número de **fases** (número de paradas que deben hacerse durante el servicio).

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.

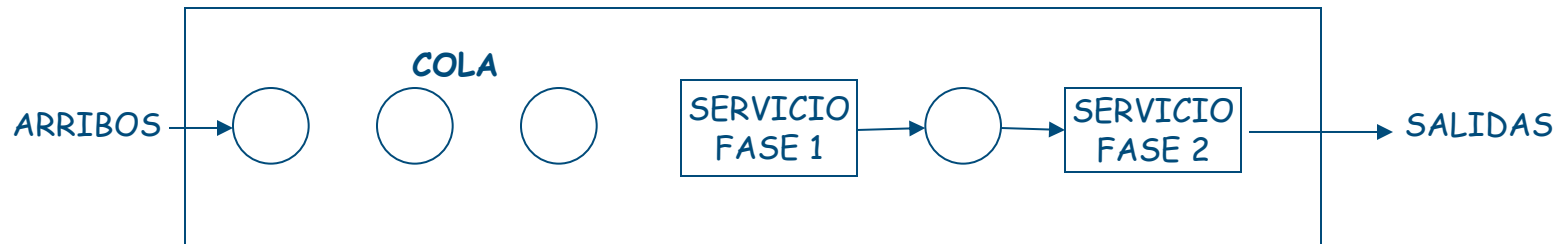
3.1. Configuraciones básicas para el Servicio

- ▶ **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 restaurante 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

3.1. Configuraciones básicas para el Servicio

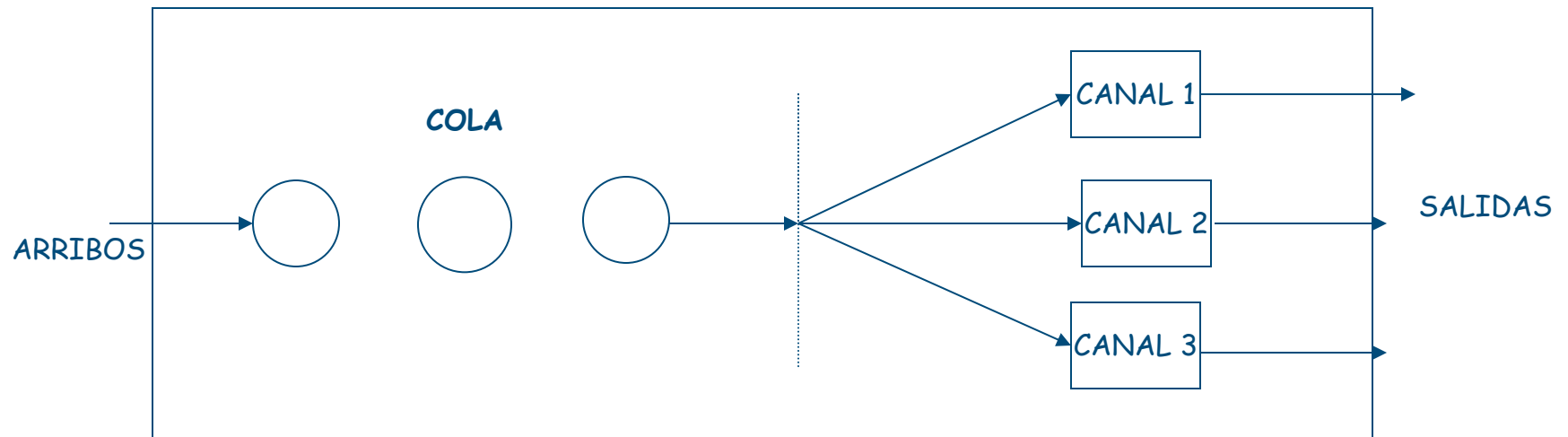


SISTEMA UN CANAL, UNA FASE



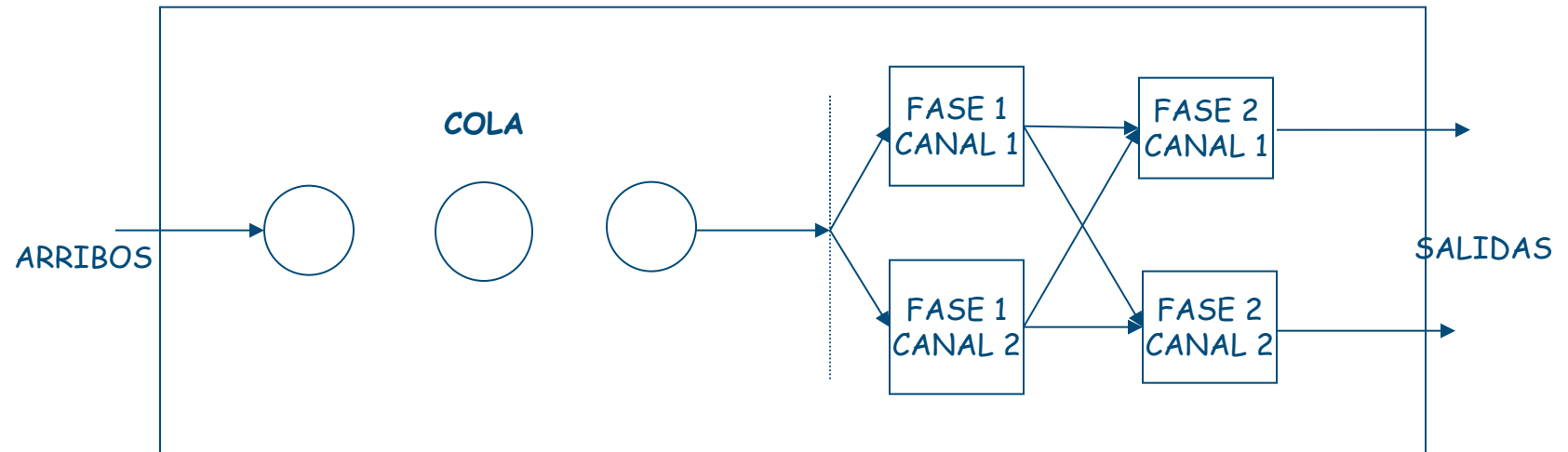
UN SOLO CANAL, MULTIFASE

3.1. Configuraciones básicas para el Servicio



SISTEMA MULTICANAL UNA FASE

3.1. Configuraciones básicas para el Servicio



SISTEMA MULTICANAL MULTIFASE

3.2. Distribución del Tiempo de Servicio

Los **patrones de servicio** son similares a los patrones de llegada. Pueden ser **constantes o aleatorios**.

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 automóviles).

Si el **tiempo de servicio es distribuido aleatoriamente**, 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.

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.

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.
- ▶ **C** = Número de servidores paralelos
- ▶ **N** = Capacidad del Sistema
- ▶ **K** = Tamaño de la población

Notación de los Modelos de Colas

Por ejemplo: **M/M/1/∞/∞** 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/∞/∞ es reducido a **M/M/1**.

Variedad de Modelos de Colas

Existe una cantidad enorme de Modelos de Colas que pueden utilizarse.
Los 4 modelos de colas más usados asumen:

- ▶ Arribos según la Distribución de Poisson
 - ▶ Disciplina PEPS (FIFO)
 - ▶ Una sola fase de servicio.
-
- ▶ Modelo **A**: Un canal, Arribos según la Distribución de Poisson; Tiempos de Servicio exponenciales
 - ▶ Modelo **B**: Multicanal
 - ▶ Modelo **C**: Tiempo de Servicio constante
 - ▶ Modelo **D**: Población Limitada

Variedad de Modelos de Colas

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.**

Modelo A: M/M/1

Asumimos que existen las siguientes condiciones:

1. Los **clientes** son servidos con una **política PEPS** 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 **razón promedio es conocida**.
5. Los **tiempos de servicio** se representan mediante la **distribución de probabilidad exponencial negativa**.
6. La razón de servicio es **más rápida** que la razón de arribo.

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**.

- **Modelo C: Modelo de Tiempo de Servicio Constante (M/D/1)**
- Algunos sistemas tienen **tiempos de servicio constantes** en lugar de exponencialmente distribuidos. Cuando los clientes son atendidos o equipos son procesados con un ciclo fijo como es el caso de una lavadora de carros automatizada o ciertos entretenimientos en los parques de diversiones, el asumir servicio constante es adecuado.
- **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.

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

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

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

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

FÓRMULAS PARA COLAS

MODELO A: SISTEMA SIMPLE O M/M/1

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

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

$$P_n = \text{Probabilidad de que "n" clientes estén en el sistema} =$$

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

$$P_o = \text{Probabilidad de cero unidades en el sistema (la unidad de servicio está vacía)} =$$

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

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

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

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	$H = FNX$
Cuantía de la Población	$N = J + L + H$

EJEMPLO: MODELO A: SISTEMA SIMPLE O M/M/1

A una línea de espera llegan 20 unidades por hora y el tiempo promedio de servicio es de 30 unidades por hora, realizar un análisis de esta línea de espera.

- Calcular la probabilidad de que el sistema esté ocupado
- Calcular la probabilidad de que el sistema no esté ocupado
- El numero esperado de unidades en el sistema
- El numero esperado de unidades que esperan ser atendidas
- El tiempo promedio que una unidad espera para ser atendida

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

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

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

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

FÓRMULAS PARA COLAS

MODELO A: SISTEMA SIMPLE O M/M/1

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

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

$$P_n = \text{Probabilidad de que "n" clientes estén en el sistema} =$$

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

$$P_o = \text{Probabilidad de cero unidades en el sistema (la unidad de servicio está vacía)} =$$

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

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

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

EJEMPLO: MODELO A: SISTEMA SIMPLE O M/M/1

A una línea de espera llegan 20 unidades por hora y el tiempo promedio de servicio es de 30 unidades por hora, realizar un análisis de esta línea de espera.

- Calcular la probabilidad de que el sistema esté ocupado
- Calcular la probabilidad de que el sistema no esté ocupado
- El número esperado de unidades en el sistema
- El número esperado de unidades que esperan ser atendidas
- El tiempo promedio que una unidad espera para ser atendida

EJEMPLO: MODELO A: SISTEMA SIMPLE O M/M/1

Suponga que un cajero bancario puede atender a los clientes a una velocidad promedio de diez clientes por hora. Además, suponga que los clientes llegan a la ventanilla del cajero a una tasa promedio de 7 por hora. Se considera que las llegadas siguen la distribución Poisson y el tiempo de servicio sigue la distribución exponencial. Determine:

- Utilización del sistema
- % del tiempo en que no habrán clientes en el sistema
- Número promedio de clientes en la cola
- Número promedio de clientes en el sistema (en cola y servicio)
- Tiempo promedio que el cliente espera en la cola
- Tiempo promedio que el cliente espera en el sistema
(en la cola y en el servicio)

EJEMPLO: MODELO A: SISTEMA SIMPLE O M/M/1

A un peaje con un único cajero llega un promedio de 10 vehículos por hora. Suponga que el tiempo promedio de servicio para cada cliente es de 4 minutos. Se considera que las llegadas siguen la distribución Poisson y el tiempo de servicio sigue la distribución exponencial. Conteste las siguientes preguntas.

- ¿Cuál es la probabilidad de que el peaje se encuentre vacío?
- ¿Cuál es el número promedio de automóviles que esperan en la cola su turno?, se considera que un vehículo que está pagando el peaje, no está en la cola esperando.
- ¿Cuál es el tiempo promedio que un cliente pasa en el proceso de llegada a la fila hasta salida del peaje?
- En promedio, ¿Cuántos clientes por hora serán atendidos por el cajero del peaje?

EJEMPLO: MODELO A: SISTEMA SIMPLE O M/M/1

Wester Nacional Bank está considerando si debe abrir una ventanilla para el servicio a clientes. La administración estima que los clientes llegarán con una tasa de 15 por hora. El cajero que atenderá la ventanilla puede atender a los clientes con una rapidez de uno cada tres minutos. Suponiendo llegadas con una distribución del tipo Poisson y un servicio exponencial, encuentre:

Utilización del cajero.

- El número promedio en la fila
- El número promedio en el sistema.
- El tiempo promedio de espera en la fila.
- El tiempo promedio de espera en el sistema, incluyendo el servicio.

EJEMPLO:
MODELO A: SISTEMA SIMPLE O M/M/1

Encargado de Bibliotecas

Un estudiante trabaja como encargado de una biblioteca por las noches y es el único en el mostrador durante todo su turno de trabajo. Las llegadas al mostrador siguen una distribución de Poisson con una media de 8 por hora. Cada usuario de la biblioteca es atendido de uno en uno, y el tiempo de servicio sigue una distribución exponencial con una media de 5 minutos.

- a) ¿Cuál es la probabilidad de que se forme cola?
- b) ¿Cuál es la longitud media de la cola?
- c) ¿Cuál es el tiempo medio que un cliente pasa en la biblioteca hasta que le han atendido?
- d) ¿Cuál es el tiempo medio que un cliente pasa en la cola esperando a que le atiendan?

EJEMPLO: MODELO A: SISTEMA SIMPLE O M/M/1

La oficina de Desarrollo Comunitario de la Municipalidad de Puente Alto, es atendida por un Trabajador Social que recibe un promedio de 3 personas por hora buscando una clase de ayuda (alimento , casa , etc.). El Trabajador Social habla con ellos y los ayuda directamente o los envía a otras oficinas o departamentos de la Municipalidad. Esto le toma un promedio de 15 minutos por persona. Suponiendo que las llegadas siguen una distribución de Poisson y el tiempo de servicio sigue una distribución exponencial, y la disciplina de la cola el primero que llega es el primero en atenderse.

Se solicita:

- a) La probabilidad de tener cero personas buscando ayuda.
- b) La probabilidad de tener dos personas buscando ayuda a la vez.
- c) El tiempo promedio que uno de ellos gasta en el sistema , incluyendo el tiempo de servicio.
- d) El tiempo promedio que una persona gasta en la espera para ser atendido por el Trabajador Social.
- e) El número promedio de personas que esperan ser atendidos.