Facultad de Ciencias Sociales y Jurídicas

Grado en Administración y Dirección de Empresas

Universidad de Jaén

Facultad de Ciencias Sociales y Jurídicas

Trabajo Fin de Grado

TEORÍA DE COLAS

Alumno: Alejandro Gallego Campos

Mayo, 2020

Resumen

El objetivo de la Teoría de Colas es el estudio matemático del comportamiento de las colas o líneas de espera, se enmarca en el área de la Investigación Operativa. Presentes en nuestro día a día, las colas aparecen cuando la demanda de un servicio es superior a la capacidad de servicio del sistema. La Teoría de Colas proporciona información para la toma decisiones de capacidad óptima y diseño de sistemas.

En este proyecto se llevará a cabo una revisión de la Teoría de Colas y de sus modelos más utilizados, los cuales se ejemplificarán para obtener una visión práctica de cómo funciona el estudio de las líneas de espera. Se realizará un análisis de la capacidad de servicio y costes asociados a los sistemas de colas.

Abstract

The objective of the Queueing theory is the mathematical study of the behaviour of queues or lines of waiting, framed in the area of Operational Research. Present in our day to day lives, queues appear when the demand for a service is higher than the service capacity of the system. Queueing Theory provides information for decision-making regarding optimal capability and systems designs.

This project will carry out a review of the Theory of Queues and the most common queue models, which will be exemplified to obtain a practical vision of how the study of waiting lines works. An analysis of service capabilities and costs associated with queuing systems shall be carried out.

Índice

1. INTRODUCCIÓN

La Teoría de Colas tiene su origen en el comienzo del siglo pasado, basando sus primeros estudios en las redes de telefonía y los problemas de congestión que presentaban, concretamente en 1909 A.K. Erlang público el primer trabajo científico de este campo.

En la actualidad esta teoría está presente el día a día de prácticamente cualquier persona, en un contexto socioeconómico donde la expresión "*el tiempo es oro*" cobra cada vez más importancia, el tiempo perdido en esperas nos resulta siempre demasiado (semáforos, supermercados, esperas telefónicas, etc.). El objetivo de esta teoría es la optimización económica y reducción de las líneas de espera.

Se analizarán las principales partes de un sistema de colas y sus características, la distribución de llegadas de clientes y de servicio. Definiendo los conceptos matemáticos de distribución exponencial, procesos estocásticos particularizando al proceso Poisson y a procesos de nacimiento y muerte.

A partir de estos conceptos, se expondrán ejemplos de varios modelos finitos e infinitos, basados en el proceso de nacimiento y muerte. Estudiando la parte económica de la Teoría de Colas, mediante el equilibrio entre el coste de servicio y coste de espera en los diferentes modelos.

2. HISTORIA

"Aunque algunos puntos dentro del campo de la telefonía dan lugar a problemas cuya solución corresponde a la Teoría de la Probabilidad, esta última no ha sido muy utilizada en este campo, hasta donde podemos ver. En este sentido, la Telephone Company of Copenhagen constituye una excepción puesto que su director gerente, el Sr. F. Johannsen, ha aplicado durante varios años los métodos de la Teoría de la Probabilidad a la solución de varios problemas de importancia práctica, además de incitar a otros a trabajar en investigaciones de características similares. Como creo que algún que otro punto de este trabajo puede resultar de interés, y no es en absoluto necesario un conocimiento especial de los problemas telefónicos para su entendimiento, darle cuenta de ello a continuación."

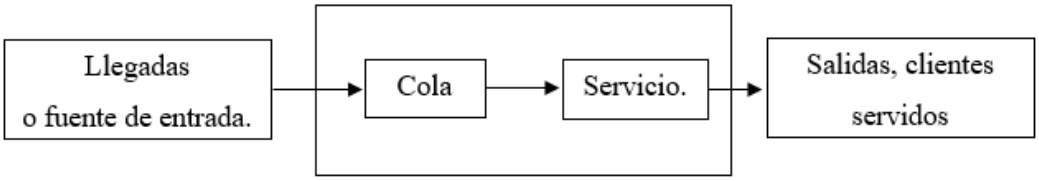
(Delgado, 2009) Introduce en su análisis de la historia de la teoría de colas mediante el comienzo del artículo que el matemático danés Agner Krarup Erlang (Longborg, Dinamarca, 1878-1929) público en 1909, considerado el primer artículo de la Teoría de Colas. El trabajo de Erlang fue ampliado por diversos investigadores en la primera mitad del siglo XX, como Pollaczek, Kolmogorov y Khintchine, entre otros. A partir de los años 50, se puede observar un gran crecimiento en este área, debido a que su interés e importancia han ido aumentando, en parte, al gran crecimiento del ámbito de las telecomunicaciones, fundamentales para el desarrollo y la globalización, ámbito donde la Teoría de Colas tiene una mayor implicación.

Actualmente, la sofisticada teoría desarrollada durante estos años tiene un abanico muy amplio de aplicaciones, desde el campo de la telefonía y las telecomunicaciones donde se inició, hasta la ciencia de la computación, los procesos industriales manufactureros (cadenas de producción), el control de tráfico (por carretera, aéreo, de personas, de información a través de Internet,) o la logística militar y civil, pasando por las empresas de servicios (hospitales, oficinas bancarias, restaurantes de comida rápida, supermercados, etc.).

3. DESCRIPCIÓN DE UN SISTEMA DE COLAS

Como se expone en la *Figura 1* el proceso básico supuesto por la mayor parte de los modelos de colas es el siguiente: Los clientes que requieren un servicio se generan a través del tiempo de una fuente de entrada. Estos clientes entran al sistema y se unen a una cola. En un determinado momento se selecciona un miembro de la cola para proporcionarle un servicio, mediante alguna regla conocida como disciplina de servicio**.** Luego, se lleva a cabo el servicio requerido por el cliente en un mecanismo de servicio, después de lo cual el cliente sale del sistema de colas (Martín, 2003).

Figura 1: *Estructura básica de un sistema de colas.*



Fuente: Elaboración propia

La complejidad de esta teoría requiere un análisis de detallado de cada uno de los elementos del proceso.

3.1 FUENTE DE ENTRADA

Es el conjunto de individuos o de población potencial que pueden demandar el servicio en un determinado momento. Esto se conoce como población de entrada, esta puede ser finita o infinita.

La población potencial suele tratarse como infinita, esto, aunque no sea realista, se debe a que permite resolver de forma más simple muchas situaciones en las que la población es finita pero muy grande, esto se tendrá que tomar como una suposición implícita si no se indica lo contrario. La suposición de infinitud no resulta restrictiva cuando, aun siendo finita la población potencial, su número de elementos es tan grande que el número de individuos que ya están solicitando el citado servicio prácticamente no afecta a la frecuencia con la que la población potencial genera nuevas peticiones de servicio.

3.2 MECANISMO DE ENTRADA

Por otro lado, se debe especificar el patrón estadístico por el que se generan los clientes a través del tiempo, normalmente se utiliza una distribución Poisson. Esto significa que las llegadas son aleatorias, pero con cierta tasa media y sin importar el número de clientes que ya se encuentran en el sistema. Cuando el número de ocurrencias de un determinado evento aleatorio se modeliza según una distribución de Poisson, el tiempo que transcurre entre dos ocurrencias del suceso aleatorio se modeliza mediante una distribución exponencial, es decir a que la distribución de probabilidad del tiempo que trascurre entre dos llegadas consecutivas es exponencial. También debemos conocer si los clientes entran en el sistema de forma individual o en grupo y, en caso de que sea en grupo, se debe conocer la distribución de probabilidad que modeliza el tamaño del mismo.

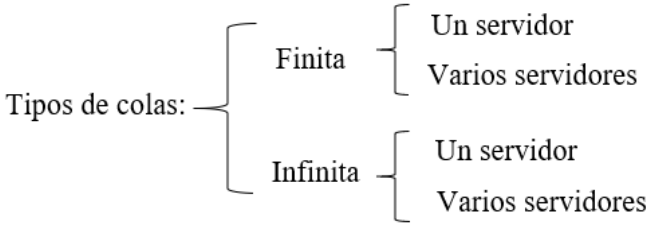
3.3 COLA

Es la ubicación donde los clientes que han solicitado el servicio esperan antes de recibirlo. Analizaremos sus dos características principales:

3.3.1 CAPACIDAD DE COLA

Es el número máximo de clientes que pueden estar haciendo cola. Esta capacidad puede ser finita o infinita, normalmente en los casos reales la cola es finita, pero el supuesto para el análisis normalmente es el de cola infinita. En la *Figura 2* se puede ver una clasificación básica de los tipos de colas según su capacidad.

Figura 2: *Clasificación de los tipos de colas según su capacidad.*



Fuente: Elaboración propia

3.3.2 DISCIPLINA DE COLA

(Ricardo Cao Abad, 2002 y Hillier & Lieberman, 2013) definen la disciplina de cola como el orden en el que sus miembros se seleccionan para recibir el servicio.

Normalmente, si no se establece otra forma la disciplina la disciplina utilizada será primero en entrar, primero en salir, (First In, First Out: FIFO).

Las principales formas son las siguientes:

- **FIFO (First In First Out)**: se le da servicio al primero que ha llegado, de forma que la cola esta ordenada según el orden de llegada de los usuarios.
- **LIFO (Last In First Out):** se le da servicio al último que ha llegado, de forma que la cola esta ordenada en orden inverso al de llegada de los usuarios.
- **RSS (Random Selection of Service) o SIRO (Service In Random Order):** se elige aleatoriamente a los clientes.
- **SJF (short job first)** se atienden primero a los clientes que demandan menor tiempo de servicio. En caso de existir varios clientes demandando el mismo tiempo de servicio, éstos serán atendidos según su orden de llegada.
- **PRI (Priority queue discipline):** establece determinadas prioridades a los diferentes usuarios según algunas de sus características, un ejemplo puede ser el tiempo de servicio que necesite, pudiendo dársele prioridad tanto al que más necesite como al que menos. Además, nos encontramos también el fenómeno de anticipación, que consiste en interrumpir un servicio por la llegada de otro prioritario.

El uso de las diferentes disciplinas de colas estará condicionado por el tipo de servicio que se realice, no todas las disciplinas son aplicables en los todos sistemas.
En sistemas finitos, en los que el número de usuarios en espera es limitado, es necesario establecer además qué sucede con aquellos usuarios que acceden al sistema cuando la cola de espera esta completa y son rechazados. Por otro lado, se debe tener en cuenta el cambio de cola o abandono de usuarios que, a la vista de la cola, renuncian a acceder al sistema.

3.4 MECANISMO DE SERVICIO

El mecanismo de servicio consiste en una o más estaciones de servicio, cada una de ellas con uno o varios canales de servicio paralelos, llamados servidores. De otra forma, el usuario tendrá que pasar por una secuencia de ellas cuando las estaciones se encuentran en serie.

Los modelos de colas deben especificar el número de servidores (canales paralelos). Los modelos más simples suponen una estación, ya sea con un servidor o con un número

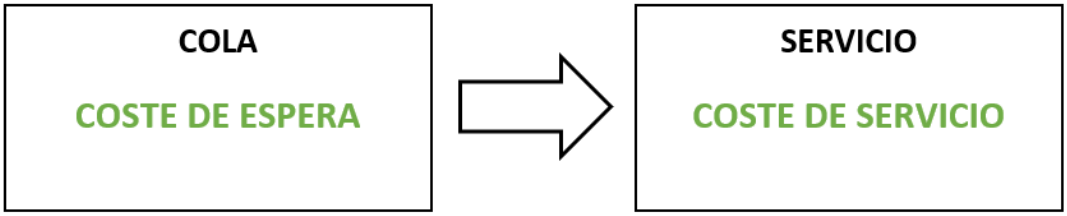
finito de servidores. En una estación de servicio única, el cliente entra por uno de estos canales y el servidor le presta el servicio completo.

Se llama tiempo de servicio o duración del servicio al tiempo que pasa desde el inicio del servicio hasta su salida de la estación.

Un modelo de un sistema de colas determinado debe especificar la distribución de probabilidad de los tiempos de servicio de cada servidor (y tal vez de los distintos tipos de clientes), normalmente se supone la misma distribución para todos los servidores. Para el tiempo de servicio, usaremos la distribución exponencial, ya que es la que más usada en la práctica, por ser más fácilmente manejable que cualquier otra. (Martín, 2003) (Hillier & Lieberman, 2013)

3.5 COSTES DE LOS SISTEMAS DE COLAS

Figura 3: *Costes de un sistema básico de colas*



Fuente: Elaboración propia

Como se explicó anteriormente la estructura más simple del sistema de colas se puede dividir en cuatro partes fundamentales: **llegadas,** unidades que acceden al sistema y demandan el servicio, **cola,** línea donde esperan a recibirlo (si no hay línea de espera se denomina cola vacía). A continuación de acuerdo con la disciplina de cola las unidades pasan al **servicio**, cuando este es completado los clientes abandonan el sistema como **salidas**.

En la *Figura 3* se pueden ver los costes asociados a algunos de los componentes del sistema, la cola y el servicio, los cuales se describen a continuación.

3.5.1 COSTE DE ESPERA

Es el coste del tiempo que se pierde al esperar, se considera este tiempo que pierden los usuarios de un sistema, como un activo que no está siendo utilizado y repercutirá negativamente, afectando a la imagen de la empresa o sistema, es por ello, que se considera de forma económica.

Se obtiene multiplicando C_w coste de espera por hora en unidades monetarias, por llegadas por unidad de tiempo y L longitud media de la línea de espera.

$$\text{Coste total de espera} = C_w L$$

3.5.2 COSTE DE SERVICIO

Este coste se integran el conjunto de los gastos de funcionamiento del servicio, algunos ejemplos son: coste de instalaciones, maquinaria, personal, mantenimiento, etc.

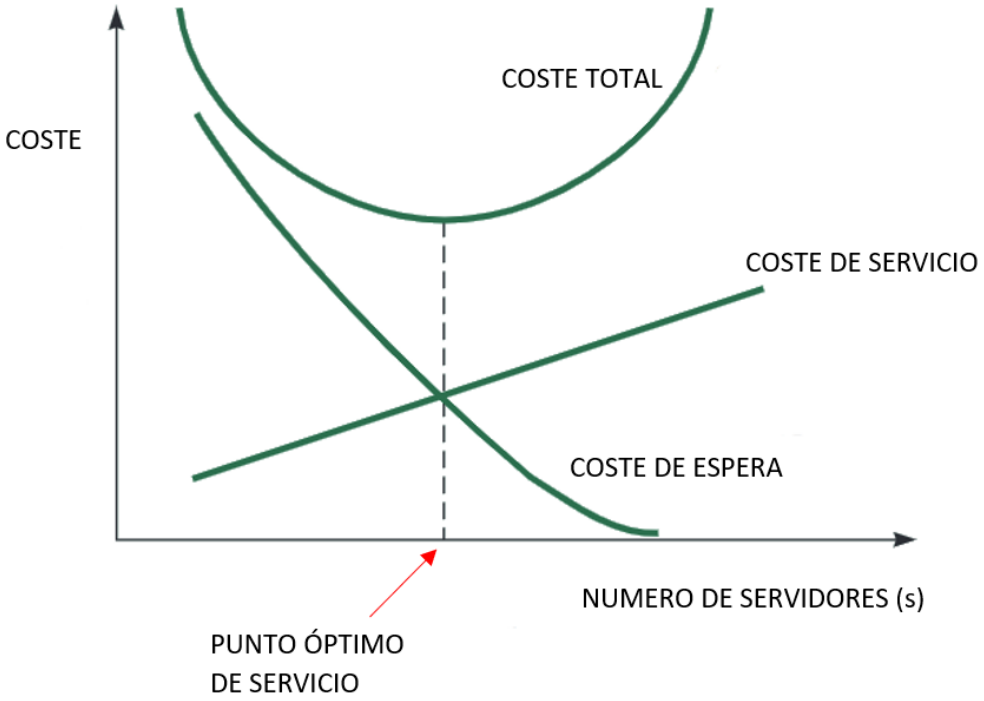
Se calcula multiplicando el coste de servicio por el número de servidores que componen el sistema:

$$\text{Coste total de servicio} = C_s\, s$$

3.5.3 EQUILIBRIO DE COSTES

Una de las finalidades de la Teoría de Colas, es buscar que estos costes sean los mínimos posibles, buscando un equilibrio adecuado entre los dos costes anteriores, ya que, si las tasas de servicio son muy bajas, los costes de espera serán muy altos, por el contrario, cuanto más altas son las tasas de servicio, más bajos son los costes de espera. La finalidad del estudio de costes es encontrar la combinación adecuada donde el coste total sea el mínimo. En la *Figura 4* se puede ver con claridad el equilibrio de costes y el comportamiento de estos.

Figura 4: *Gráfico equilibrio de costes.*



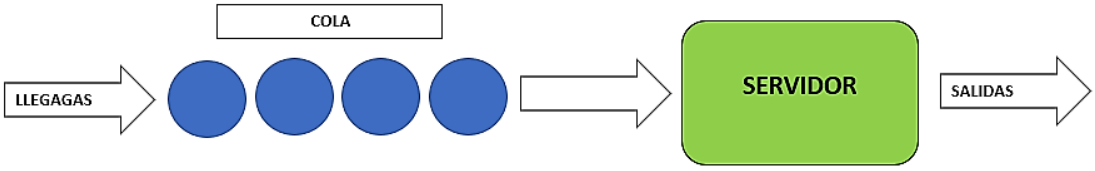
Fuente: Elaboración propia

Mas adelante, analizaremos los costes y su punto de equilibrio en algunos ejemplos de sistemas de colas.

3.6 REPRESENTACIÓN GRÁFICA

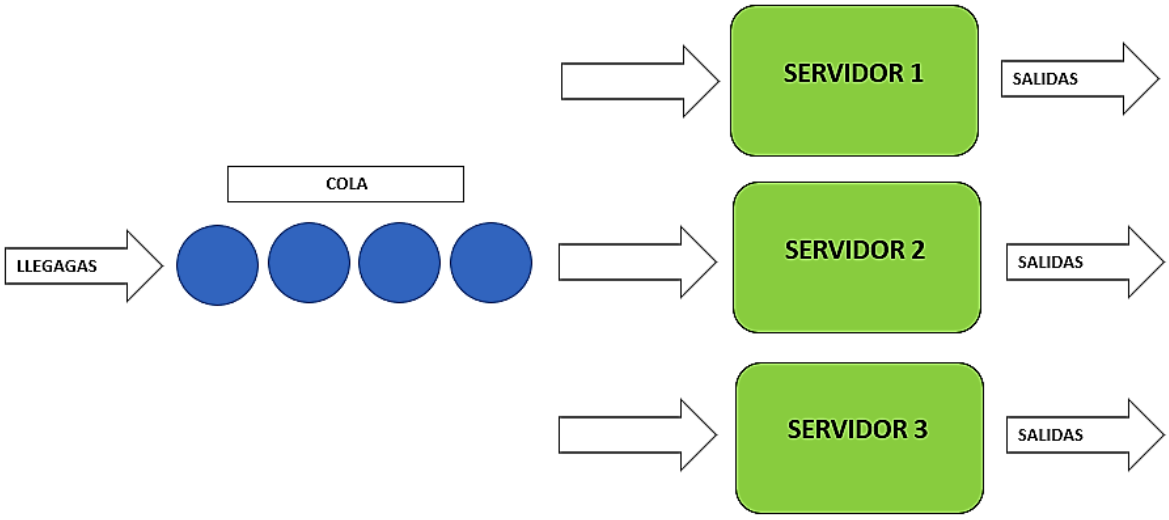
En las *Figuras 5 a 8* se muestran gráficamente algunos de los principales ejemplos de sistemas de colas.

Figura 5: *Una cola con un solo servidor.*



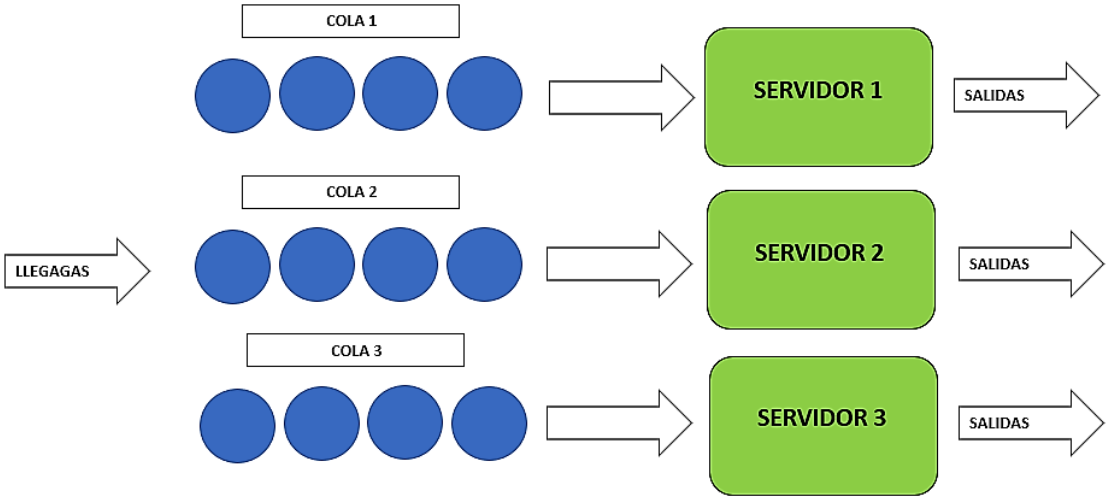
Fuente: Elaboración propia

Figura 6: *Una cola con varios servidores.*



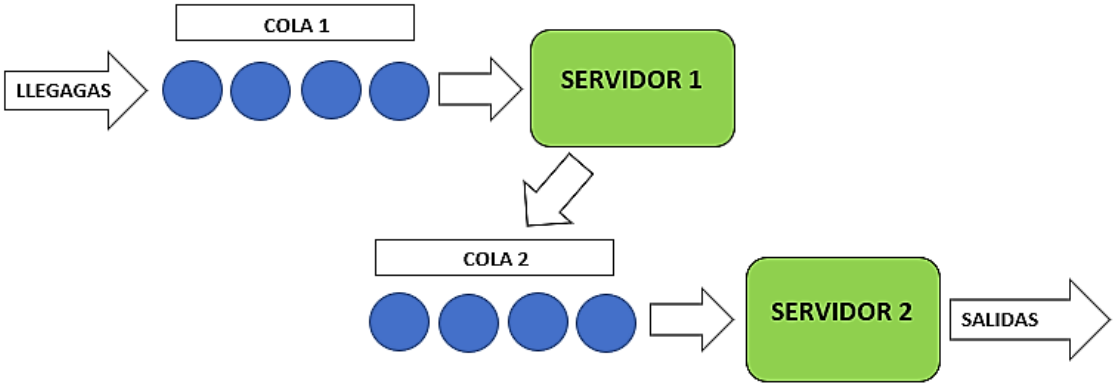
Fuente: Elaboración propia.

Figura 7: *Varias colas con varios servidores.*



Fuente: Elaboración propia.

Figura 8: *Modelo de servidores secuenciales.*



Fuente: Elaboración propia.

3.7 EJEMPLOS DE SISTEMAS DE COLAS REALES

Si nos paramos a observar nuestro día a día y analizamos situaciones cotidianas, nos percataremos de la presencia de esta teoría en una gran variedad de ámbitos: transportes, servicios, telecomunicaciones, industria, agricultura, e incluso en cada individuo que, tiene sus propias líneas de espera personales o tareas programadas como: estudiar, hacer deportes, libros que leer, etc.

Veamos ahora cómo los sistemas de colas se aplican con bastante frecuencia en una amplia variedad de contextos:

Comenzando desde las más básicas, en el ámbito comercial, encontramos panaderías o tiendas donde normalmente encontramos una cola con un mecanismo de servicio (el panadero o dependiente) este modelo corresponde a la *Figura 5* del apartado anterior.

Algo más complejos, como se observa en la *Figura 6*, son los sistemas de cola única implementados de forma exitosa por algunas cadenas de supermercados, donde observamos el modelo de una sola cola con varios servidores (cajeros/as) otros ejemplos comunes de este sistema, pueden ser las cooperativas agrícolas, oficinas bancarias o de correos.

Respecto a la *Figura 7*, varias colas con varios servidores, es común, en supermercados, gasolineras, (varias colas, varios surtidores), entre otros.

Sistemas más complejos, con servidores secuenciales *Figura 8*, existen gran variedad de ejemplos: hospitales, juzgados, cooperativas, fabricas, puertos de mercancías, aeropuertos**,** entre otros muchos. Cogiendo como ejemplo los aeropuertos, encontramos numerosos ejemplos de colas desde que entramos hasta que llegamos a nuestro asiento en el avión, puede variar según el aeropuerto, pero un modelo podría ser: primero pasaremos por una cola única con distintos servidores para la facturación del equipaje, a continuación nos tendremos que dirigir a otra cola única, donde nos derivan a varios mecanismos de servicio, en el control de seguridad, y para terminar tendríamos que pasar por la cola de embarque.

Los anteriores ejemplos son sistemas físicos o perceptibles en nuestro día a día, por otro lado, con igual o incluso mayor presencia de la Teoría de colas, se encuentran los sistemas de telecomunicaciones, de donde como ya vimos anteriormente, surgieron los primeros estudios y avances en esta teoría. Aunque estos sean menos perceptibles son de vital importancia en la conocida como *era digital*, donde las tecnologías de la información y comunicación (TIC) se han convertido en parte fundamental de nuestras vidas.

3.8 TERMINOLOGÍA BÁSICA Y FÓRMULA DE LITTLE

Para las fórmulas y cálculos de este trabajo, se utiliza la notación más común, basada en los trabajos de (Martín, 2003) y (Hillier & Lieberman, 2013). Es importante señalar que la teoría de colas normalmente basa sus análisis a la condición de **estado estable**, en parte porque el caso transitorio es analíticamente más difícil. La terminología será la siguiente:

- **L_q (t) = longitud de la cola,** número de clientes que están en la cola en el instante t, se puede calcular: estado del sistema menos el número de clientes que están siendo atendidos.
- **N(t)** = número de clientes en el sistema de colas en el tiempo t ($t \geq 0$).
- **$P_n(t)$** = probabilidad de que exactamente n clientes estén en el sistema en el tiempo t, dado el número en el tiempo 0.
- **s** = número de servidores (servicio en paralelo) en el sistema de
- colas.
- **λ_n** = tasa media de llegadas (número esperado de llegadas por unidad de
- tiempo) de nuevos clientes cuando hay n clientes en el sistema.
- **μ_n** = tasa media de servicio en todo el sistema (número esperado de clientes que completan su servicio por unidad de tiempo) cuando hay n clientes en el sistema. Nota: μ_n representa la tasa combinada a la que todos los servidores ocupados (aquellos que están sirviendo a un cliente) logran terminar sus servicios.
- **ρ** = factor de utilización del sistema $\rho = \frac{\lambda_n}{s\mu_n}$

El termino **estado estable** definido forma simple, quiere decir que el sistema no depende del instante de tiempo (t), (Hillier & Lieberman, 2013) nos explica, que estado estable es cuando un sistema se vuelve independiente del estado inicial y del tiempo trascurrido, esto ocurre una vez que ha trascurrido suficiente tiempo desde que se inicia el sistema, este periodo inicial se denomina condición transitoria.

- **Estado del sistema =** número de clientes que se encuentran en el sistema, también se puede definir como el valor que toma la variable N(t) cuando el estado del sistema es estable.
- **P_n** = probabilidad de que haya exactamente n clientes en el sistema.
- **L** = longitud o número medio de unidades en el sistema (incluyendo la cola).
- **L_q** = longitud o número medio de unidades en la cola (excluye los clientes que están en servicio).
- **W** = tiempo de espera en el sistema (incluye tiempo de servicio) para cada cliente.
- **W_q** = tiempo de espera en la cola (excluye tiempo de servicio) para cada cliente.

3.8.1 FORMULA DE LITTLE

Esta fórmula es fundamental en el estudio de colas, permitiendo determinar las cuatro medidas fundamentales: L, W, L_q y W_q determinando el valor de solo una de ellas. Se conoce como fórmula de Little ya que fue John D. C. Little en 1961, quien proporcionó la primera demostración rigurosa de la misma.

Cuando las tasas de llegada son constantes, λ_n es una constante λ para toda n. La fórmula de Little seria:

$$L= \lambda W,$$

También verifica que:

$$L_q = \lambda W_q.$$

Una forma intuitiva de entender la fórmula de Little es la siguiente: considerando que un cliente que llega al sistema justo ahora, trascurrido un tiempo, cuya media es W, ese cliente saldrá servido del sistema. Como el número medio de clientes que llegan al sistema por unidad de tiempo es λ, el número medio de clientes que habrán llegado desde que nuestro cliente en cuestión entró en el sistema hasta que salió de él es λW. La fórmula de Little no es válida si λ_n no es constaste en n, pero puede sustituirse por $\bar{\lambda}$.

Otra relación importante seria:

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

Esto significa que: el tiempo medio que un cliente está en el sistema W, coincide con la suma del tiempo medio en la cola W_q más el tiempo medio que tarda en ser servido $1/\mu$, ya que μ es el número medio de clientes que un servidor puede atender por unidad de tiempo (Ricardo Cao Abad, 2002).

3.9 NOTACIÓN KENDALL

Para especificar o nombrar el tipo de cola, se utiliza la notación Kendall**,** esta notación permite ver las características principales de una cola de forma rápida y sencilla.

Esta notación sigue el formato:

$$(A \mid B \mid c \mid m \mid d)$$

A: Representa la distribución que siguen los tiempos de **llegada.**

B: Representa la distribución que siguen los tiempos de **servicio.**

A y B pueden ser las siguientes distribuciones:

M = distribución exponencial.
D = Determinista
Er = Erlang.
G = General

c: Número de canales de servicio (s).

m: Número máximo de unidades permitidas en el sistema (finito o infinito).

d: Disciplina de cola. Puede ser cualquiera de las explicadas anteriormente *(Epígrafe 3.3.2.).*

Es importante apuntar que cuando los valores m y d, son infinito y FIFO, al ser los valores más habituales, se suelen omitir, por lo tanto, quedaría:

$$(A \mid B \mid c)$$

Por ejemplo, uno de los modelos que se analizaran más adelante el $(M \mid M \mid 1)$, en notación Kendall significa que, la distribución de llegada y servicio es exponencial, tiene un canal o estación de servicio, infinitas unidades permitidas en el sistema y disciplina de cola FIFO.

4. DISTRIBUCIÓN EXPONENCIAL

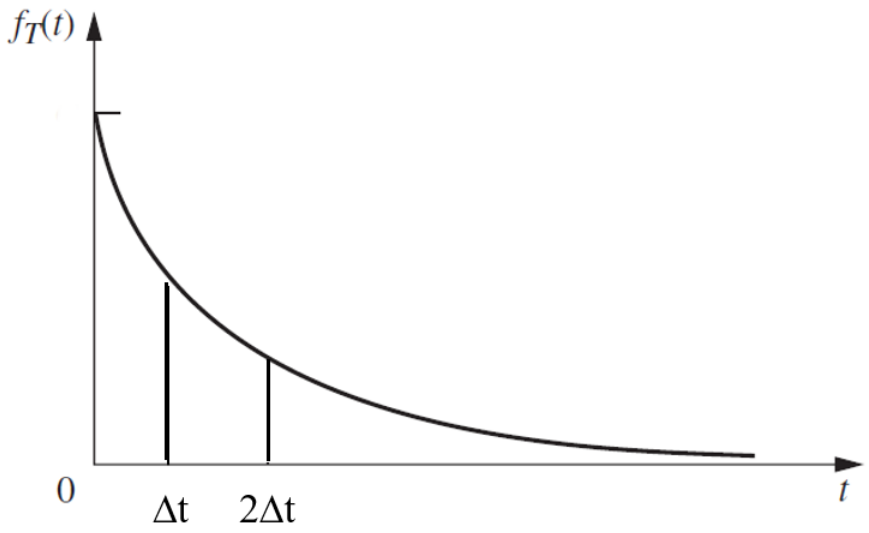
En la Teoría de Colas la principal distribución de probabilidad es la exponencial, se considera un intervalo grande de tiempo, y se marcan los intervalos en los que ocurre o un evento o llegada Poisson. El tiempo entre las llegadas sucesivas se representa con el símbolo t (siendo la variable t positiva aleatoria y continua). Se dice que esta variable aleatoria tiene una distribución exponencial con parámetro λ si su función de densidad de probabilidad es:

$$f\tau(t) = \lambda e^{-\lambda t},\ t \geq 0$$

Así pues, se analizarán las principales propiedades de la distribución exponencial.

Propiedad 1: f (t) es una función de t estrictamente decreciente de t (t ≥0).

Figura 9: *Distribución exponencial entre llegadas.*



Fuente: Elaboración propia.

Se verifica que: P (0 ≤T ≤Δt) > P (Δt ≤ T ≤ 2Δt).

Propiedad 2: Falta de memoria, el comportamiento de las llegadas ocurre de forma totalmente aleatoria, que ocurra un evento no se ve afectado por el tiempo trascurrido respecto al evento anterior, es igual a decir que la distribución de probabilidad del tiempo que falta hasta que ocurra un evento es igual, independientemente del tiempo que haya trascurrido. (Δt)

$$P(T > t + \Delta t \,|\, T > \Delta t) = P(T > t) = \frac{e^{-\lambda[t+\Delta t]}}{e^{-\lambda \Delta t}} = e^{-\lambda t}$$

5. PROCESOS ESTOCÁSTICOS

Normalmente se define un proceso estocástico como un conjunto de variables aleatorias {X(t)}, este subíndice *t* presenta valores de un conjunto T dado. Habitualmente se toma T como el conjunto de enteros no negativos, por otro lado, X representa una característica de interés evaluable en el tiempo t. Por ejemplo, el proceso estocástico, X(t) puede representar cada uno de los niveles de inventario semanales de un producto dado, es decir nivel de inventario de un producto X en la semana t. Es por esto, por lo que una de las principales aplicaciones de los procesos estocásticos, es el estudio del comportamiento de un sistema en periodos de tiempo $\{X(t)\} = \{X_1, X_2, X_3, \ldots\}$, siendo esto una representación matemática de la evolución de un sistema, en el tiempo.

Principales características:

- $\{X(t); t \in T\}$ es un proceso estocástico si X(t) es una variable aleatoria para cada t, normalmente t indica tiempo.

Los procesos estocásticos pueden ser:

Continuos: cuando t puede tomar cualquier valor dentro de un intervalo.

Discretos: cuando t toma valores dentro de un conjunto discreto de puntos.

Denotando X(t) = x. indica que el proceso aleatorio se encuentra en el estado x en el tiempo t.

- Sea un proceso estocástico $\{N(t); t\geq 0\}$. Decimos que $\{N(t); t\geq 0\}$ es un proceso de conteo si y solo si:

  1. N (0) = 0 c.s. (casi seguro)
  2. N(t) toma solo valores enteros y no negativos c.s.
  3. Si $s < t \rightarrow N(s) < N(t)$. c.s.
  4. N(t)- N(s) = Numero de sucesos ocurridos en el intervalo de tiempo (s; t].

- Un proceso $\{X(t); t\geq 0\}$ se dice que es de Markov si y solo si:

$P\{X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n \ldots, X(t_1) = x_1\} = P\{X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n\}$.

Esto, quiere decir que la probabilidad de que ocurra un suceso depende exclusivamente del suceso inmediatamente anterior y no de sucesos pasados.

6. PROCESO POISSON

(Martín, 2003) Señala el proceso Poisson como el más común en el diseño de los modelos de colas, este viene dado por la siguiente definición:

Sea {N(t); t≥0} un proceso de conteo. Se dice que {N(t); t≥0} es un proceso de Poisson de parámetro λ si y sólo si:

1. {N (t); t ≥ 0} es un proceso estocástico de incrementos independientes y estacionarios, es decir, que a lo largo del tiempo no cambia el modelo de llegadas (salidas) y que es independiente una llegada (salida) de otra llegada (salida).

2. P(N(h) = 1) = λh + o(h): Esto significa, que la probabilidad de que ocurra una llegada o salida en un periodo de tiempo de longitud h es proporcional a dicha longitud, a más longitud de tiempo, más probabilidad de que ocurra una llegada o salida.

3. P(N(h) ≥ 2) = o(h) En un periodo de tiempo h, si este es muy pequeño, la probabilidad de que ocurran 2 o más llegadas o salidas, es prácticamente nula, es decir no llegan ni salen clientes en grupo.

4. Las llegadas son procesos sin memoria: cada llegada en un intervalo de tiempo es independiente de las llegadas en intervalos previos o futuros.

Teorema 1. Si se tiene un proceso de Poisson {N(t); t≥0} con parámetro λ, entonces ∀ t N(t) ≡ $_d$P (λ t), es decir,

$$P(N(t) = n)\frac{(\lambda\, t)^n e^{-\lambda\, t}}{n!}$$

Teorema 2. Sea {N(t); t≥0} un proceso de Poisson con parámetro λ. Sea T_i el instante de ocurrencia del suceso i-ésimo. Definamos t_i; =T_i- T_{i-1} como el tiempo transcurrido entre los sucesos i -1e i. Entonces:

1. t_1, t_2, ..., son v.a.i.i.d. Exp(λ)
2. ∀n ϵN, $T_n \equiv_d$Erlang(n, λ)

Teorema 3. (Reciproco del anterior) Sea{N(t); t≥0} un proceso de Poisson de conteo. Si t_1, t_2, ..., son v.a.i.i.d. Exp(λ), entonces {N(t); t≥0} es un proceso estocástico de Poisson con parámetro λ.

Teorema 4. Los tiempos entre llegadas son exponenciales con parámetro λ si y sólo si el número de llegadas que suceden en un intervalo de tiempo, t, sigue una distribución de Poisson con parámetro λ t.

Propiedad 3: Relación de la distribución exponencial con la distribución de la Poisson. Si tenemos un proceso de Poisson {X(t)}, definido en [0, t], con parámetro λt se verifica

$$P(X(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

X(t) tiene una distribución de Poisson con parámetro λ t. Por ejemplo, para n =0

$$P(X(t) = 0) = e^{-\lambda t} = P(T > t) = 1 - F(t)$$

que coincide con la probabilidad que se obtuvo a partir de la distribución exponencial para que ocurra el primer evento después de un tiempo t. La media de la distribución de Poisson es

$$\boldsymbol{E[X(t)]} = \lambda t \rightarrow \lambda = \frac{E[X(t)]}{t},$$

de manera que el número esperado de eventos por unidad de tiempo es λ. Por lo tanto, se dice que λ es la tasa media a la que ocurren los eventos es decir un proceso de conteo $\{X(t); t \geq 0\}$ es un proceso de Poisson con parámetro λ (la tasa media).

7. PROCESO DE NACIMIENTO Y MUERTE

Cuando un proceso de colas donde los tiempos entre llegadas son exponenciales (y por lo tanto el número de llegadas por unidad de tiempo Poisson) y los tiempos de servicios son exponenciales, es un proceso de nacimiento y muerte. Según (Hillier & Lieberman, 2013) la gran mayoría de los modelos elementales de colas suponen que las entradas (llegadas de clientes) y las salidas (clientes que salen) del sistema ocurren de acuerdo a un proceso de nacimiento y muerte.

Debido a su flexibilidad y utilidad, los modelos basados en este proceso son los más usados en la teoría de colas. En el contexto de la Teoría de Colas, el termino nacimiento se refiere a la llegada de un cliente al sistema de colas, mientras que el termino muerte se refiere a la salida del cliente servido.

Este proceso es un tipo especial de **Cadena de Markov**[1] con restricciones en cada paso, el estado de transición, que sólo puede ocurrir entre los estados más cercanos. El proceso de nacimiento y muerte describe en términos probabilísticos como cambia N(t) al aumentar t. En general, sostiene que los nacimientos y muertes individuales ocurren de manera aleatoria, y que sus tasas medias de ocurrencia dependen del estado del sistema actual. A continuación, definiremos los supuestos del proceso de nacimiento y muerte:

Supuesto 1. Dado N(t) = n, la distribución de probabilidad actual del tiempo que falta para el próximo nacimiento (llegada) es exponencial con parámetro λ_n (n = 0, 1, 2, . . .).

Supuesto 2. Dado N(t) = n, la distribución de probabilidad actual del tiempo que falta para la próxima muerte (finalización de servicio) es exponencial con parámetro μ_n (n = 1, 2, . . .).

Supuesto 3. La variable aleatoria del supuesto 1 (el tiempo que falta hasta el próximo nacimiento) y la variable aleatoria del supuesto 2 (el tiempo que falta hasta la siguiente muerte) son mutuamente independientes. La siguiente transición del estado del proceso es:

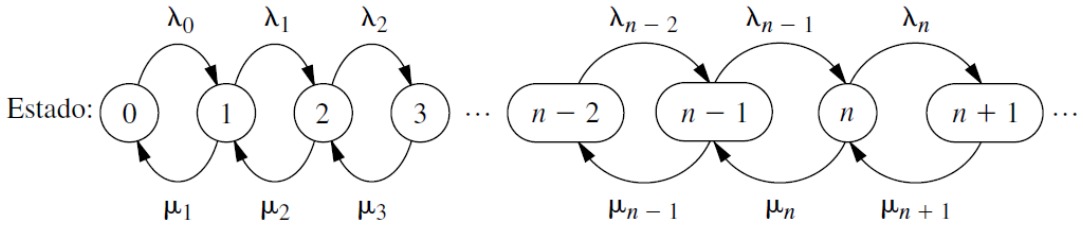
$n \rightarrow n+1$ (un solo nacimiento)

o

$n \rightarrow n-1$ (una sola muerte).

[1]Cadena de Markov: son una serie de sucesos, en los cuales la probabilidad de que ocurra un suceso depende únicamente del suceso inmediato anterior.

Figura 10: *Diagrama de las tasas del proceso de nacimiento y muerte.*



Fuente: (Hillier & Lieberman, 2013)

Como se expone en anteriores apartados, λ_n y μ_n representan, respectivamente, la tasa media de llegada y la tasa media de terminaciones de servicio, cuando hay n clientes en el sistema.

En algunos sistemas de colas, los valores de las λ_n serán las mismas para todos los valores de n, y las μ_n también serán las mismas para toda n excepto para aquella n tan pequeña que el servidor este desocupado n=0.

Esto puede ser diferente en algunos sistemas de colas, normalmente las variaciones pueden deberse a cantidades muy grandes tanto de clientes potenciales como de clientes en cola, por ejemplo, una de las formas en las que λ_n puede ser diferente para valores distintos de n es si los clientes que llegan se pueden perder (rechazar la entrada al sistema) con mayor probabilidad a medida que n aumenta. En cuanto a μ_n es diferente cuando hay más de un servidor, también suele ser diferente ante valores distintos de n debido a que existe una mayor probabilidad de que los clientes renuncien a medida que aumenta el tamaño de la cola, sistemas de colas impacientes (se vayan sin haber sido atendidos)

7.1 ANÁLISIS DEL PROCESO DE NACIMIENTO Y MUERTE

Para realizar un análisis, en primer lugar, se toma un estado n (n = 0,1, 2, ...) del sistema. Se supone que en el tiempo 0 se inicia el conteo del número de veces que el sistema entra a este estado y el número de veces que sale de él, como se denota seguidamente:

- $E_n(t)$ = número de veces que el proceso entra al estado n hasta el tiempo t.
- $L_n(t)$ = número de veces que el proceso sale del estado n hasta el tiempo t.

Como los dos tipos de eventos (entrar y salir) deben alternarse, estos dos números serán iguales o diferirán en solo 1, explicado de forma sencilla, no podrá haber dos

clientes en una estación de servicio al mismo tiempo, ni llegar dos clientes al mismo tiempo, la diferencia máxima será uno en los momentos de transición entre la salida de un cliente y la entrada del siguiente, matemáticamente seria:

$$\left| E_n(t) - L_n(t) \right| \leq 1.$$

Al dividir ambos lados entre t y después hacer el límite de t cuando tiende a infinito obtendríamos:

$$\left|\frac{E_n(t)}{t} - \frac{L_n(t)}{t}\right| \leq \frac{1}{t}, \text{ entonces } \lim_{t\to\infty}\left|\frac{E_n(t)}{t} - \frac{L_n(t)}{t}\right| = 0$$

Si se dividen En(t) y Ln(t) entre t se obtiene la tasa real (número de eventos por unidad de tiempo) a la que ocurren estos dos tipos de eventos, y cuando $t \to \infty$ se obtiene la tasa media (número esperado de eventos por unidad de tiempo):

$\lim_{t\to\infty} \frac{E_n(t)}{t}$ = tasa media a la que el proceso entra al estado n.

$\lim_{t\to\infty} \frac{L_n(t)}{t}$ = tasa media a la que el proceso sale al estado n.

Estos resultados conducen al principio de la ecuación de balance:

Tasa de llegada = tasa de salida o servicio.

Tasa media de entrada = tasa media de salida.

Se analizará entonces ecuación, en primer lugar, se supone el estado 0, el proceso solo puede entrar a dicho estado, desde el estado 1. Por lo tanto, la probabilidad de estado estable de encontrarse en el estado 1 (P_1) representa la proporción de tiempo que es posible que el proceso entre al estado 0. Dado que el proceso se encuentra en el estado 1, la tasa media de entrada al estado 0 es μ_1. (En otras palabras, para cada unidad acumulada de tiempo que el proceso pasa en el estado 1, el número esperado de veces que lo dejaría para entrar al estado 0 es μ_1.) Desde cualquier otro estado, esta tasa media es 0. Por lo tanto, la tasa media global a la que el proceso deja su estado actual para entrar al estado 0 (la tasa media de entrada) es:

$$\mu_1 P_1 + 0(1 - P1) = \mu_1 P_1$$

Realizando el mismo razonamiento, la tasa media de salida debe ser, $\lambda_0 P_0$, de manera que la ecuación de balance del estado 0 es:

$$\mu_1 P_1 = \lambda_0 P_0$$

Se pueden ver las ecuaciones de balance en la *Tabla 1*.

Tabla 1*: Ecuaciones de balance.*

Estado	Tasa de entrada = Tasa de salida.
0	$\mu_1 P_1 = \lambda_0 P_0$
1	$\lambda_0 P_0 + \mu_2 P_2 = (\lambda_1 + \mu_1) P_1$
2	$\lambda_1 P_1 + \mu_3 P_3 = (\lambda_2 + \mu_2) P_2$
⋮	⋮
n - 1	$\lambda_{n-2} P_{n-2} + \mu_n P_n = (\lambda_{n-1} + \mu_{n-1}) P_{n-1}$
n	$\lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} = (\lambda_n + \mu_n) P_n$
⋮	⋮

Fuente: Elaboración propia.

Se encuentran dos transiciones posibles, para todos los demás estados: hacia adentro y hacia afuera. Por lo tanto, cada lado de las ecuaciones de balance de estos estados representa la suma de las tasas medias de las dos transiciones incluidas. Por lo demás, el razonamiento es igual que para el estado 0.

Se observa como la primera ecuación de balance contiene dos variables (P_0 y P_1), las primeras dos ecuaciones contienen tres variables (P0, P1 y P2), y así sucesivamente, de manera que siempre se tiene una variable "adicional". Por lo tanto, el procedimiento para resolver estas ecuaciones es despejar todas las variables en términos de una de ellas, entre las cuales la más conveniente es P_0.

En la *Tabla 1* se puede observar como la primera ecuación se usa para despejar P_1 en términos de P_0; después se usa este resultado y la segunda ecuación para obtener P_2 en términos de P_0, etc. Al final, el requisito de que la suma de todas las probabilidades debe ser igual a 1 se puede usar para evaluar P_0.

Aplicando este procedimiento se obtienen los resultados en la *Tabla 2.*

Tabla 2: *Resultados de la ecuación de balance.*

Estado:	
0	$P_1 = \frac{\lambda_0}{\mu_1} P_0$
1	$P_2 = \frac{\lambda_1}{\mu_2} P_1 + \frac{1}{\mu_2}(\mu_1 P_1 - \lambda_0 P_0) = \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} P_0$
2	$P_3 = \frac{\lambda_2}{\mu_3} P_2 + \frac{1}{\mu_3}(\mu_2 P_2 - \lambda_1 P_1) = \frac{\lambda_2}{\mu_3} P_2 = \frac{\lambda_2 \lambda_1 \lambda_0}{\mu_3 \mu_2 \mu_1} P_0$
⋮	⋮
n - 1	$P_n = \frac{\lambda_{n-1}}{\mu_n} P_{n-1} + \frac{1}{\mu_n}(\mu_{n-1} P_{n-1} - \lambda_{n-2} P_{n-2}) = \frac{\lambda_{n-1}}{\mu_n} P_{n-1} = \frac{\lambda_{n-1} \lambda_{n-2} \dots \lambda_0}{\mu_n \mu_{n-1} \dots \mu_1} P_0$
n	$P_{n+1} = \frac{\lambda_n}{\mu_{n+1}} P_n + \frac{1}{\mu_{n+1}}(\mu_n P_n - \lambda_{n-1} P_{n-1}) = \frac{\lambda_n}{\mu_{n+1}} P_n = \frac{\lambda_n \lambda_{n-1} \dots \lambda_0}{\mu_{n+1} \mu_n \dots \mu_1} P_0$
⋮	⋮

Fuente: Elaboración propia.

Simplificando:

$$c_n = \frac{\lambda_{n-1} \lambda_{n-2} \dots \lambda_0}{\mu_n \mu_{n-1} \dots \mu_1}, \text{ para n= 1, 2, ...,}$$

A continuación, se define c_n=1 para n=0. Por lo tanto, las probabilidades de estado estable son:

$$P_n = c_n P_0, \text{ para n = 1, 2, …,}$$

El requisito:

$$\sum_{n=0}^{\infty} P_n = 1$$

Implica que

$$\left(\sum_{n=0}^{\infty} c_n \right) P_0 = 1$$

Así, la probabilidad de que no se encuentren clientes en el sistema seria:

$$P_0 = \left(\sum_{n=0}^{\infty} c_n \right)^{-1}$$

Por lo tanto, para que exista dicha distribución la serie tiene que ser convergente. Cuando un modelo de líneas de espera se basa en el proceso de nacimiento y muerte, de manera que el estado del sistema n representa el número de clientes en el sistema de colas, las medidas clave de desempeño del sistema (L, L_q, W y W_q) se pueden obtener de inmediato después de calcular las P_n mediante las fórmulas anteriores.

Longitud o número medio de unidades en el sistema:

$$L = \sum_{n=0}^{\infty} nP_n$$

Y longitud o número medio de unidades en la cola:

$$Lq = \sum_{n=s}^{\infty} (n - s)P_n$$

También se extrae a través de la fórmula de Little:

$$W = \frac{L}{\bar{\lambda}} \qquad W_q = \frac{L_q}{\bar{\lambda}}$$

Donde $\bar{\lambda}$ es la tasa de llegadas promedio a largo plazo. Como λ_n es la tasa media de llegadas cuando el sistema se encuentra en el estado n (n = 0, 1, 2, . . .) y P_n es la proporción de tiempo que el sistema está en este estado, se obtiene como:

$$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n$$

7.2 MODELOS DE COLAS INFINITAS

7.2.1 MODELO (M | M | 1)

Se observa en la notación de este modelo que tanto los tiempos de llegadas como los tiempos de servicio siguen una distribución exponencial y sólo hay un servidor. Como ya se expuso en el apartado de la notación Kendall, si los valores que señalan el número máximo de unidades permitidas en el sistema y la disciplina de cola son omitidos, se supondrá una población infinita y disciplina de cola FIFO.

Por lo tanto:

- La tasa de llegada seria: $\lambda_n = \lambda$. Las llegadas se suponen aleatorias, vienen de una población infinita y llegan de una en una.
- La tasa de servicio: $\mu_n = \mu$ para $n \geq 1$.
- Y el número de servidores s = 1.

- El factor de utilización será $\rho < 1$, trabajaremos en régimen estable.

El objetivo será determinar los valores: c_n, P_0, P_n, L, L_q, W y W_q.

En el apartado anterior se obtuvo la expresión para c_n

$$c_n = \frac{\lambda_{n-1}\lambda_{n-2}\dots\lambda_0}{\mu_n\mu_{n-1}\dots\mu_1}, = \frac{\lambda^n}{\mu^n} = \rho^n, \quad \text{para n= 0,1, 2, ...,}$$

La probabilidad de que la cola se encuentre vacía se obtiene de la condición de normalización, es decir, del hecho de que la suma de las probabilidades para todos los estados es la unidad. Cuando $\lambda \geq \mu$, la tasa media de llegadas excede la tasa media de servicio, la serie para calcular P_0 no sería convergente, por lo tanto, no existiría la distribución estacionaria y se dice que el sistema explota.

$$P_0 = \left(\sum_{n=0}^{\infty} c_n\right)^{-1} = \left(\sum_{n=0}^{\infty} \rho_n\right)^{-1} =$$

$$= \left(\frac{\mathbf{1}}{\mathbf{1}-\boldsymbol{\rho}}\right)^{\mathbf{-1}}$$

$$\mathbf{= 1 - \rho}$$

Por lo tanto, la probabilidad de que haya n clientes en el sistema viene dada por:

$$P_n = (1-\rho)\,\rho^n, \quad \text{para n= 0,1,2…}$$

Para obtener la longitud esperada de clientes en el sistema y en la cola, calculando su valor esperado, al tratarse de una serie convergente, se puede intercambiar el sumatorio por la derivada, para calcular la longitud esperada media del sistema:

$$L = \sum_{n=0}^{\infty} n\,P_n = \sum_{n=0}^{\infty} n\,(1-\rho)\rho^n = (1-\rho)\rho \sum_{n=0}^{\infty} n\,\rho^{n-1} = (1-\rho)\rho \sum_{n=0}^{\infty} \frac{d}{d\rho}(\rho^n) =$$

$$= (1-\rho)\rho \frac{d}{d\rho} \sum_{n=0}^{\infty} (\rho^n) = (1-\rho)\rho \frac{d}{d\rho} \frac{1}{1-\rho} = (1-\rho)\rho \frac{1}{(1-\rho)^2}$$

$$L = \frac{\rho}{1-\rho} = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda}$$

Y para la longitud media espera en la cola:

$$L_q = \sum_{n=1}^{\infty}(n-1)\,P_n = \sum_{n=1}^{\infty} n\,P_n - \sum_{n=1}^{\infty} P_n = L - (\sum_{n=1}^{\infty} P_n - P_0) = L - (1 - P_0) =$$

$$= \frac{\lambda}{\mu - \lambda} - \left(1 - \left(1 - \frac{\lambda}{\mu}\right)\right) = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = \frac{\lambda\mu - (\lambda\mu - \lambda^2)}{\mu(\mu - \lambda)} = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

Aplicando la fórmula de Little se obtiene el tiempo medio de espera en el sistema y el tiempo medio de espera en cola, respectivamente W y W_q:

$$W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda}$$

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

7.2.2 MODELO (M | M | s)

Es una generalización del modelo anterior a varios servidores (s). Por lo tanto, un sistema en la que la distribución del tiempo entre llegadas como la distribución del tiempo de servicio son exponenciales y hay s servidores. Como en el apartado anterior, la población y la capacidad de la cola serán infinitas y la disciplina de la cola FIFO.

Tendríamos:

s= 2,3,4…

La tasa de llegadas:

$\lambda n = \lambda$ para n=0,1,2…

La tasa de servicio:

$$\mu_n = \begin{cases} n\mu, & n = 0,1,2,\dots,s \\ s\mu, & n \geq s \end{cases}$$

El factor de utilización para este modelo viene dado por:

$$\rho = \frac{\lambda}{\mu s} < 1$$

Teniendo que ser $\rho < 1$ para que exista la distribución estacionaria.

La expresión para c_n es:

$$c_n = \frac{\lambda_{n-1}\lambda_{n-2}\dots\lambda_0}{\mu_n\mu_{n-1}\dots\mu_1} = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n} & n = 1,2,\dots,s \\ \dfrac{\lambda^s}{s!\,s^s}\left(\dfrac{\lambda}{s\mu}\right)^{n-s} & n \geq s \end{cases}$$

En este modelo el sistema será estacionario, siempre que ρ<1, o lo que es lo mismo, siempre que λ< sμ.

P_0 será:

$$P_0 = \frac{1}{\sum_{n=0}^{s-1}\frac{\lambda^n}{n!\,\mu^n} + \sum_{n=s}^{\infty}\frac{\lambda^n}{s!\,s^s}\left(\frac{\lambda}{s\mu}\right)^{n-s}} = \frac{1}{\sum_{n=0}^{s-1}\frac{\lambda^n}{n!\,\mu^n} + \frac{(\lambda/\mu)^s}{s!}\frac{1}{1-(\lambda/s\mu)}}$$

Y la posibilidad de que haya n clientes en el sistema P_n:

$$P_n = c_n P_0 = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n}P_0 & n = 0,1,2,\dots,s \\ \dfrac{\lambda^n}{s!\,s^s}\left(\dfrac{\lambda}{s\mu}\right)^{n-s}P_0 & n \geq s \end{cases}$$

Cuando existen varios servidores, resulta más sencillo calcular primero la longitud esperada de la cola y con ella la longitud esperada en el sistema:

$$L_q = \sum_{n=s}^{\infty}(n-s)P_n$$

Se hacen los cambios, j= n-s y n=s+j, y quedaría:

$$L_q = \sum_{j=0}^{\infty} jP_{s+j} = \sum_{j=0}^{\infty}\frac{\lambda^s}{s!\,\mu^s}\left(\frac{\lambda}{s\mu}\right)^j P_0 j =$$

$$= \sum_{j=0}^{\infty} j\frac{\lambda^s}{s!\,\mu^s}\rho^j P_0 = \frac{\lambda^s}{s!\,\mu^s}P_0\sum_{j=0}^{\infty} j\rho^j = \frac{\lambda^s}{s!\,\mu^s}\rho P_0\sum_{j=0}^{\infty} = \frac{d}{d\rho}\rho^j$$

$$= \frac{\lambda^s}{s!\,\mu^s}\rho P_0\frac{d}{d\rho}\sum_{j=0}^{\infty}\rho^j = \frac{\lambda^s}{s!\,\mu^s}\rho P_0\frac{d}{d\rho}\frac{1}{1-\rho} = \frac{\lambda^s}{s!\,\mu^s}\rho P_0\frac{1}{(1-\rho)^2}$$

$$L_q = P_0\frac{\lambda^s}{s!\,\mu^s}\,\frac{\lambda}{s\mu}\,\frac{1}{\left(1-\frac{\lambda}{s\mu}\right)^2}$$

Aplicando las formula de Little se extraen los tiempos de espera media en cola y en el sistema:

$$W_q = \frac{L_q}{\lambda}, \qquad W = W_q + \frac{1}{\mu}$$

Usando el tiempo de espera medio en el sistema, se calcula la longitud media en el sistema:

$$L = \lambda W = \lambda\left(W_q + \frac{1}{\mu}\right) = L_q + \frac{\lambda}{\mu}$$

La expresión completa sería:

$$L = P_0 \frac{\lambda^s}{s!\,\mu^s} \frac{\lambda}{s\mu} \frac{1}{\left(1 - \frac{\lambda}{s\mu}\right)^2} + \frac{\lambda}{\mu}$$

7.2.3 EJEMPLO MODELOS (M | M | 1) Y (M | M | s)

Para la realización de los ejemplos de este proyecto, se utilizará el software POM-QM v3, un programa para las ciencias de la decisión. Se usa en los ámbitos de la producción y administración de operaciones, métodos cuantitativos, o investigación operativa. El módulo utilizado para el proyecto de colas será Waiting Lines.

Se analiza el ejemplo de un sistema de cola con el caso de un taller de reparación de lunas para coches. El mecánico es capaz de cambiar 3 lunas por hora, es decir μ=3. Mientras que al taller llegan 2 clientes por hora solicitando el servicio, esto quiere decir que λ=2. Suponiendo que el taller como un modelo (M | M | 1), estudiaremos sus principales tasas. Realizaremos el estudio de forma matemática, con la base teórica del trabajo y mediante el software POM-QM.

λ=2 coches por hora solicitan el servicio.

μ=3 coches por hora son atendidos.

Puesto que $\rho < 1$ existe distribución estacionaria. Se puede determinar la probabilidad de que se encuentre n coches en el sistema con la fórmula:

$P_n = (1-\rho)\,\rho^n$

$\rho = \frac{\lambda}{\mu} = \frac{2}{3} = 0.667$ es la probabilidad de que la estación de servicio este ocupada.

$P_0 = 1 - \rho = 1 - \frac{\lambda}{\mu} = 1 - \frac{2}{3} = 0.333$ la probabilidad de que no haya coches en el sistema.

$L = \frac{\lambda}{\mu - \lambda} = \frac{2}{3-2} = 2$ coches de media del sistema.

$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{2^2}{3(3-2)} = \frac{4}{3} = 1.333$ coches de media esperando en la cola.

$W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda} = \frac{1}{3-2} = 1$ hora es la media que pasan los coches en el sistema.

$W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{2}{3(3-2)} = \frac{2}{3}$ de una hora, es decir 40 min es el tiempo de espera medio por coche.

Ejemplo:

$P_1 = (1-0.667) \cdot 0.667^1 = 0.222$ esto quiere decir que hay un 22.2% de posibilidades de que haya 1 coche en el sistema.

Introduciendo los parámetros en el programa POM-QM se extraen los siguientes resultados *Tabla 3.*

Tabla 3: *Parámetros POM-QM, Taller de lunas.*

Parameter	Value
Average server utilization	0,67
Average number in the queue(Lq)	1,33
Average number in the system(Ls)	2
Average time in the queue(Wq)	0,67
Average time in the system(Ws)	1

Fuente: Elaboración propia POM-QM v3.

Se puede ver cómo coinciden con los calculados mediante las fórmulas.

Obsérvese también la *Tabla 4* con las distintas probabilidades de n coches en el sistema (la notación del software sustituye la n por k):

Tabla 4*: Distribución de probabilidades taller de lunas.*

k	Prob (num in sys = k)	Prob (num in sys <= k)	Prob (num in sys >k)
0	0,333	0,333	0,667
1	0,222	0,556	0,444
2	0,148	0,704	0,296
3	0,099	0,802	0,198
4	0,066	0,868	0,132
5	0,044	0,912	0,088
6	0,029	0,941	0,059
7	0,02	0,961	0,039
8	0,013	0,974	0,026
9	0,009	0,983	0,017
10	0,006	0,988	0,012
11	0,004	0,992	0,008
12	0,003	0,995	0,005
13	0,002	0,997	0,003
14	0,001	0,998	0,002
15	0	0,998	0,002
16	0	0,999	0,001
17	0	1	0
18	0	1	0

Fuente: Elaboración propia POM-QM v3.

La tercera columna muestra la probabilidad de que el número de coches en el sistema sea mayor estricto que la k de cada fila, mientras que en las otras dos columnas se muestra la probabilidad de que sea igual a k(n) y menor o igual a k(n) respectivamente. Se puede comprobar como el valor de la primera columna para k=1 es igual al calculado con la formula $P_{n,}\, P_1 = 0.222$.

Se añadirán ahora los cálculos de costes, el taller estima que tiene un coste de servicio de 15 u.m y un coste de espera de 50 u.m. ambos por hora.

Por lo tanto, el coste total por hora

- Coste total de espera = $C_W L = 50 \cdot 2 = 100$ u.m.
- Coste total del servicio = $C_s\, s = 15 \cdot 1 = 15$ u.m.
- Coste total teniendo en cuenta solo la espera en cola = $C_w L_q + C_s\, s$ $= 50 \cdot 333 + 15 \cdot 1 = 81.67$ u.m.
- Coste total del sistema por hora = $C_w L + C_s\, s = 115$ u.m.

Introduciendo estos parámetros al software, muestra los siguientes resultados:

Tabla 5: *Costes totales taller POM-QM*

Parameter	Value
Cost (Labor + # waiting*wait cost)	81,67
Cost (Labor + # in system*wait cost)	115

Fuente: Elaboración propia POM-QM v3.

Observando la *Tabla 5,* se puede ver como los valores coinciden con los calculados anteriormente. Utilizando el software, se obtendrán los costes para cada número de servidores, proporcionando el punto óptimo de servicio, en la *Figura 4, Epígrafe 3.5.3* se exponía gráficamente el punto donde los costes totales se minimizan.

Como se puede ver en la *Tabla 6,* el servicio del taller alcanzaría su punto óptimo de servicio con dos servidores, con un coste total de 67.5 u.m., Por otro lado, se observa como la opción de un servidor es la menos eficiente ya que aun siendo los costes superiores a los del punto óptimo, con 3, 4 y 5 servidores el coste total seguiría siendo inferior al caso s=1.

Tabla 6*: Costes totales por número de servidores.*

Nº of servers	Total cost based on system
1	115
2	67,5
3	78,8
4	93,38
5	108,34

Fuente: Elaboración propia POM-QM v3.

Suponiendo ahora un incremento muy grande de la demanda de lunas, pasando de $\lambda=2$ a $\lambda=15$, con un solo empleado en el taller era de $\mu=3$. Aplicando la teoría de los procesos de nacimiento y muerte, sabemos que para que el sistema sea convergente y no explote, λ no puede ser mayor o igual a $s\mu$. Para satisfacer esta demanda el taller decide, aumentar su plantilla y su inversión en maquinaria, aumentado esta inversión la capacidad de

servicio a μ=4, por lo tanto, para que $\rho = \frac{\lambda}{\mu s} < 1$, *s* tiene que ser mayor que 3, siendo el número mínimo de servidores necesarios de 4.

Esta mejora del servicio propicia que los costes aumenten y pasen a ser de 18 u.m. por estación de servicio y hora, manteniendo los mismos costes de espera estimados. Se analizará cómo podría minimizar los costes con los nuevos datos *Tabla7*:

Tabla 7: *Costes totales Taller de lunas.*

Number of servers	Total cost based on waiting	Total cost based on system
1 (insufficient capacity)	n.a.	n.a.
2 (insufficient capacity)	n.a.	n.a.
3 (insufficient capacity)	n.a.	n.a.
4	720,77	908,27
5	159,27	346,77
6	126,95	314,45
7	131,94	319,44
8	145,88	333,38

Fuente: Elaboración propia POM-QM v3.

La *Tabla 7*, muestra que los costes totales se optimizan con 6 estaciones de servicio. Es reseñable el elevado coste total que tendría el servicio con 4 servidores. Se complementa con la *Tabla 8* donde el valor de espera en cola W_q para 4 servidores sería 0.87, (52.2 minutos de espera media) y la longitud media de la cola L_q = 12.98 coches, esto repercute en unos costes de espera muy altos, mientras que para 5 servidores estos valores se reducen considerablemente: W_q=0.09 (5.4 min) L_q=1.39.

Como se vio en *Epígrafe 3.5.3 Equilibrio de costes*, cuando pasamos el punto de equilibrio los tiempos de servicio siguen bajando, reduciendo los costes de espera y aumentando los de servicio. De este modo para 7 servidores, aunque los costes de espera sean menores, prácticamente la cola es nula W_q= 0.008, los costes totales son superiores.

Tabla 8: *Valores para (s) servidores.*

4 SERVIDORES	Value	**5 SERVIDORES**	Value
Average server utilization	0,94	Average server utilization	0,75
Average number in the queue(Lq)	12,98	Average number in the queue(Lq)	1,39
Average number in the system(Ls)	16,73	Average number in the system(Ls)	5,14
Average time in the queue(Wq)	0,87	Average time in the queue(Wq)	0,09
Average time in the system(Ws)	1,12	Average time in the system(Ws)	0,34
Cost (Labor + # waiting*wait cost)	720,77	Cost (Labor + # waiting*wait cost)	159,27
Cost (Labor + # in system*wait cost)	908,27	Cost (Labor + # in system*wait cost)	346,77
6 SERVIDORES	Value	**7 SERVIDORES**	Value
Average server utilization	0,63	Average server utilization	0,54
Average number in the queue(Lq)	0,38	Average number in the queue(Lq)	0,12
Average number in the system(Ls)	4,13	Average number in the system(Ls)	3,87
Average time in the queue(Wq)	0,03	Average time in the queue(Wq)	0.008
Average time in the system(Ws)	0,28	Average time in the system(Ws)	0,26
Cost (Labor + # waiting*wait cost)	126,95	Cost (Labor + # waiting*wait cost)	131,94
Cost (Labor + # in system*wait cost)	314,45	Cost (Labor + # in system*wait cost)	319,44

Fuente: Elaboración propia POM-QM v3.

7.3 MODELOS DE COLAS FINITAS

La característica principal de los siguientes modelos es como su nombre indica la capacidad limitada del sistema (K), no se permitirá un número de clientes en cola mayor que (K- s), cuando el sistema se encuentre completo, no se permite la entrada a ningún cliente al sistema. Por ello, se debe sustituir la tasa media de llegadas λ por 0 cuando n sea mayor o igual a K.

$$\lambda_n = \begin{cases} \lambda, & n = 0,1,2,\dots,K-1 \\ 0, & n \geq K \end{cases}$$

Seguidamente, se estudiarán los sistemas (M | M | 1) y (M | M | s) analizados ya en los anteriores epígrafes, pero con una capacidad del sistema limitada. Por lo tanto, en notación Kendall: (M | M | 1 | K) y (M | M | s | K).

Cabe señalar, que en este tipo de sistemas la distribución estacionaria existe siempre puesto que es una distribución finita.

7.3.1 MODELO (M | M | 1 | K)

Se determinarán los valores de c_n, P_0, P_n, L, L_q, W y W_q para este sistema. Los datos del sistema de colas serán:

- Numero de servidores: s=1
- Tasa de servicio: $\mu_n = \mu$
- El factor de utilización: $\rho = \frac{\lambda}{s\mu}$

La expresión de c_n, será la siguiente:

$$c_n = \begin{cases} \left(\frac{\lambda}{\mu}\right)^n = \rho^n, & n = 0,1,2,\dots,K \\ 0, & n > K \end{cases}$$

La probabilidad de que no haya clientes en el sistema viene dada por:

$$P_0 = \frac{1}{1+\sum_{n=1}^{K} c_n} = \frac{1}{1+\sum_{n=0}^{K} c_n} = \frac{1}{1+\sum_{n=1}^{K} \rho^n} = \frac{1}{\frac{1-\rho^{k+1}}{1-\rho}} = \frac{1-\rho}{1-\rho^{k+1}}$$

$$= \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{K+1}}$$

La probabilidad de que haya n clientes en el sistema:

$$P_n = c_n P_0 = \frac{(1-\rho)\rho^n}{1-\rho^{k+1}}, \qquad n = 0,1,2,\dots,K$$

La longitud o número medio de unidades en el sistema se obtiene mediante:

$$L = \sum_{n=0}^{K} n\, P_n = \sum_{n=0}^{K} n\, \frac{(1-\rho)\rho^n}{1-\rho^{k+1}} = \frac{(1-\rho)\rho}{1-\rho^{k+1}} \sum_{n=0}^{K} \frac{d}{d\rho}(\rho^n) =$$

$$= \frac{(1-\rho)\rho}{1-\rho^{k+1}} \frac{d}{d\rho}\left(\frac{1-\rho^{K+1}}{1-\rho}\right) = \frac{\rho}{1-\rho}\left(\frac{-(K+1)\rho^K(1-\rho)+(1-\rho^{K+1})}{1-\rho^{k+1}}\right)$$

Aplicando la formula $L_q = L - (1-P_0)$, válida para s=1, se extrae el número de clientes que hay en la cola.

$$L_q = \frac{\rho}{1-\rho}\left(\frac{-(K+1)\rho^K(1-\rho)+(1-\rho^{K+1})}{1-\rho^{k+1}}\right) - (1-P_0)$$

Los tiempos de espera se pueden obtener de forma directa con la fórmula de Little, siendo:

$$W_q = \frac{L_q}{\bar{\lambda}}, \qquad W = \frac{L}{\bar{\lambda}},$$

Donde la tasa media de entrada $\bar{\lambda}$, también llamada tasa efectiva de llegadas es el número medio de clientes que son atendidos en el sistema. Será menor que λ al tener en cuenta los clientes rechazados cuando el sistema está completo.

$$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n = \sum_{n=0}^{K-1} \lambda\, P_n = \lambda(\sum_{n=0}^{K} P_n - P_K) = \lambda\,(1 - P_K)$$

El número de clientes por unidad de tiempo que llegan al sistema cuando está completo y son rechazados se calcula:

Numero de rechazos/t = $P_K \cdot \lambda$

7.3.2 MODELO (M | M | s | K)

Generalizando el modelo a varios servidores (s), siguiendo el mismo procedimiento, se calculan los valores de: c_n, P_0, P_n, L, L_q, W y W_q. extrayendo que:

La tasa de servicio:

$$\mu_n = \begin{cases} n\mu, & n = 0,1,2,\dots,s \\ s\mu, & n = s, s+1, \dots, K \end{cases}$$

El factor de utilización:

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

La expresión c_n:

$$c_n = \frac{\lambda_{n-1}\lambda_{n-2}\dots\lambda_0}{\mu_n\mu_{n-1}\cdots\mu_1} = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n} & n = 1,2,\dots,s \\ \dfrac{\lambda^s}{s!\,s^s}\left(\dfrac{\lambda}{s\mu}\right)^{n-s} & n = s+1, s+2, \dots, K \\ 0 & n > K \end{cases}$$

La probabilidad de que el sistema se encuentre en P_0 será:

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{\lambda^n}{n!\,\mu^n} + \frac{(\lambda/\mu)^s}{s!} \sum_{n=s}^{\infty} \left(\frac{\lambda}{s\mu}\right)^{n-s}}$$

Así, P_n:

$$P_n = c_n P_0 = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n} P_0 & n = 0,1,2,\dots,s \\ \dfrac{\lambda^n}{s!\,s^s}\left(\dfrac{\lambda}{s\mu}\right)^{n-s} P_0 & n = s+1, s+2, \dots, K \\ 0 & n > K \end{cases}$$

Para el cálculo de los valores L, L_q, W y W_q, se calculará primero L_q y a partir de las formula de Little se extraen el resto.

$$L_q = \sum_{n=s}^{K} (n-s) P_n,$$

Haciendo los cambios, j= n-s y n=s+j, quedaría:

$$L_q = \sum_{j=0}^{K-s} j\, P_{j+s} = \sum_{j=0}^{K-s} j \left(\frac{\lambda}{\mu}\right)^s \frac{1}{s!} \rho^j P_0 =$$

$$= P_0 \left(\frac{\lambda}{\mu}\right)^s \frac{1}{s!} \sum_{j=0}^{K-s} j\rho^j = \frac{P_0}{s!}\left(\frac{\lambda}{\mu}\right)^s \rho \sum_{j=0}^{K-s} \frac{d}{d\rho}\rho^j = \frac{P_0}{s!}\left(\frac{\lambda}{\mu}\right)^s \rho \frac{d}{d\rho} \sum_{j=0}^{K-s} \rho^j =$$

$$= \frac{P_0}{s!}\left(\frac{\lambda}{\mu}\right)^s \rho \frac{d}{d\rho}\ \frac{1-\rho^{K-s+1}}{1-\rho}$$

$$L_q = \frac{P_0}{s!}\left(\frac{\lambda}{\mu}\right)^s \rho \frac{-(K-s+1)\rho^{K-s}(1-\rho) + (1-\rho)^{K-s+1}}{(1-\rho)^2}$$

La longitud del sistema y los tiempos de espera son:

$$W_q = \frac{L_q}{\bar{\lambda}}, \qquad W = W_q + \frac{1}{\mu}, \qquad L = \bar{\lambda} W$$

Donde la tasa media de llegadas $\bar{\lambda}$:

$$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n \mathrm{P}_n = \sum_{n=0}^{\mathrm{K-1}} \lambda\, \mathrm{P}_n = \lambda(\sum_{n=0}^{\mathrm{K}} \mathrm{P}_n - \mathrm{P}_K) = \lambda\,(1 - \mathrm{P}_K)$$

Es interesante señalar el supuesto (M | M | s | s) donde K=s, la capacidad de la cola por lo tanto K-s=0 en este caso cuando los clientes llegan y los servidores están ocupados abandonan el sistema, no hay cola. Esto ocurrirá, por ejemplo, en una red telefónica con s líneas de manera que cuando todas estas líneas están ocupadas, quien llama obtiene una señal de ocupado y cuelga. Este tipo de sistema, un "sistema de colas" sin cola, fue

estudiado por primera vez por el fundador de la Teoría de colas A. K. Erlang y se conoce como *sistema de pérdidas Erlang* o fórmula Erlang-B. Al no tener cola los valores L_q y W_q serán 0, mientras que el tiempo en el sistema es igual al tiempo de servicio $W=\frac{1}{\mu}$.

7.3.3 EJEMPLO MODELOS (M | M | 1 | K) Y (M | M | s | K)

Suponiendo de nuevo el mismo taller de lunas. Las características del sistema serán las siguientes, un mecánico *s*=1 con una tasa de servicio de 3 lunas por hora μ=3 y una talla de llegadas de λ=2. Como el tamaño del taller es limitado, la capacidad máxima del servicio es K=3, es decir un cliente siendo atendido y un máximo de dos en espera. Los clientes que llegan al sistema cuando está completo no pueden acceder a él y la disciplina de cola será FIFO. Con lo cual quedaría un sistema (M | M | 1 | 3).

Tabla 9*: Parámetros POM-QM, Taller de lunas.*

Parameter	Value
M/M/1 with a Finite System Size	
Arrival rate(lambda)	2
Service rate(mu)	3
Number of servers	1
Maximum system size	3

Fuente: Elaboración propia POM-QM v3.

Introduciendo los parámetros, *Tabla 9* en el software POM-QM se obtienen los siguientes resultados:

Tabla 10: *Resultados Taller de lunas(M | M | 1 | 3).*

Parameter	Value
Average server utilization	0,58
Average number in the queue(Lq)	0,43
Average number in the system(Ls)	1,02
Average time in the queue(Wq)	0,25
Average time in the system(Ws)	0,58
Effective Arrival Rate	1,75
Probability that system is full	0,12

Fuente: Elaboración propia POM-QM v3.

Analizando principalmente los dos últimos datos de la *Tabla 10*, siendo estos de vital importancia para comprobar viabilidad del sistema de colas, En primer lugar, el último dato muestra la probabilidad de que el sistema esté completo o P_K siendo esta de un 12%, esto propicia que el número de coches que llegan al sistema y no pueden acceder sea bajo,

siendo esto importante para el cálculo de costes que se realizarán posteriormente. En segundo lugar, la penúltima fila muestra la tasa efectiva de llegadas o tasa media de entradas, este valor expone la efectividad del sistema, mayor será cuanto más cercano sea a λ. Los resultados obtenidos en la *Tabla 10* se pueden considerar buenos ya que no muestran una congestión muy alta del sistema.

Para comparar, se puede ver la *Tabla 11* donde se supone un sistema con idénticos parámetros, sustituyendo λ=2 por λ=12, aquí se aprecia como la probabilidad de que el sistema este completo es del 75%, y la tasa efectiva de entradas es de 2.96 muy lejos del valor de $\lambda = 12$.

Aplicando la fórmula: $P_K \cdot \lambda = 0.75 \cdot 12 = 9$ clientes rechazados por hora.

Tabla 11: *Resultados Taller de lunas para λ=12 (M│M│1│3).*

Parameter	Value
Average server utilization	0,99
Average number in the queue(Lq)	1,69
Average number in the system(Ls)	2,68
Average time in the queue(Wq)	0,57
Average time in the system(Ws)	0,9
Effective Arrival Rate	2,96
Probability that system is full	0,75

Fuente: Elaboración propia POM-QM v3.

Seguidamente, se pretende mejorar la eficiencia del sistema. Suponiendo la contratación de dos nuevos mecánicos, pasando a ser un sistema de colas (M │M │3 │3), es decir un *sistema de pérdidas Erlang*, (M │M │s │s).

La *Tabla 12*, muestra los resultados para el sistema con estos parámetros, como se vio en el epígrafe anterior este sistema no tiene cola, se puede corroborar ya que los valores L_q y W_q son 0.

En cuanto a la eficiencia del sistema, ha sufrido una leve mejora, siendo la probabilidad de que el cliente encuentre el sistema completo, un 45% frente al 75% anterior y la tasa de efectividad 6.59, aún lejos de la tasa de llegadas.

Tabla 12: *Resultados Taller de lunas para λ=12 (M│M│3│3).*

Parameter	Value
Average server utilization	0,73
Average number in the queue(Lq)	0
Average number in the system(Ls)	2,2
Average time in the queue(Wq)	0
Average time in the system(Ws)	0,33
Effective Arrival Rate	6,59
Probability that system is full	0,45

Fuente: Elaboración propia POM-QM v3.

Por lo tanto, lo aconsejable para la mejora de la eficiencia, sería el aumento de la capacidad del taller, ya que se ha comprobado que, para la máxima capacidad de servidores, sigue teniendo una pérdida de clientes considerable.

Para finalizar, se realizará el análisis de costes para un modelo (M│M│1│K), utilizando los mismos los datos del ejemplo para servidores con cola infinita, un coste de servicio C_s de 15 u.m. por hora y un coste de espera C_w de 50 u.m. por hora. En este apartado, se considerará el coste de la pérdida de clientes que se produce cuando el taller esta completo, englobando este la pérdida del beneficio del servicio más el daño a la imagen de la empresa, se supone un coste de pérdida de clientes C_p de 40 u.m. por cliente.

Siendo el coste de pérdida de clientes por hora: $C_p\ P_K.\lambda$.

El coste total se calcula como:

$$C_T = (C_s\ s) + (C_w\ L) + (C_p\ P_K.\lambda).$$

La *Tabla 13* muestra los resultados con costes para este caso.

Tabla 13: *Resultados Taller de lunas (M│M│1│3) con costes.*

Parameter	Value
Average server utilization	0,58
Average number in the queue(Lq)	0,43
Average number in the system(Ls)	1,02
Average time in the queue(Wq)	0,25
Average time in the system(Ws)	0,58
Effective Arrival Rate	1,75
Probability that system is full	0,12
Cost (Labor + # waiting*wait cost)	36,54
Cost (Labor + # in system*wait cost)	65,77

Fuente: Elaboración propia POM-QM v3.

Coste total sin la pérdida de clientes = 65.77 u.m.

Coste de pérdida de clientes por hora = $50 \cdot 0.12 \cdot 2 = 12$ u.m.

Coste total = 65.77+12 = 77.77 u.m.

El coste de pérdida de clientes es relativamente bajo, ya que la probabilidad de que el sistema este completo es baja y prácticamente no se rechazan clientes. Calculando el número de clientes rechazados seria:

$$P_K \lambda = 0.12 \cdot 2 = 0.24 \text{ clientes rechazados por hora.}$$

Por lo tanto, la eficiencia del sistema se puede considerar aceptable.

8. CONCLUSIONES

Tras el análisis de la Teoría de Colas y varios de sus modelos se exponen las siguientes conclusiones:

- Las líneas de espera están presentes en el día a día, se pueden considerar que son inevitables, el tiempo de espera en estas depende del grado de optimización.
- Existen una gran variedad de modelos de colas, con diferentes características y aplicaciones.
- Para las empresas es importante conseguir un balance adecuando entre los costes de espera y los costes de servicio.
- La pérdida de memoria es una característica fundamental. La llegada de un cliente no depende del cliente anterior, son dos sucesos independientes.
- Generalmente los sistemas de colas se estudian como modelos de nacimiento y muerte, donde los tiempos de llegada y servicio siguen una distribución exponencial. Donde nacimiento es la llegada de un cliente al sistema y muerte la salida del cliente servido.
- El factor de utilización generalmente ha de ser menor que uno, si la tasa media de llegadas excede a la tasa media de servicio, la cola crece sin límite y el sistema no sería estacionario "explota".
- Los datos obtenidos para el ejemplo del Taller de Lunas muestran la gran cantidad de información que arroga el estudio de un sistema de colas, medidas de eficiencia y económicas de gran utilidad para el análisis y diseño de sistemas de servicio.

9. BIBLIOGRAFÍA

Delgado, R. (2009). Recordando a Erlang : Un breve paseo (sin esperas) por la Teoría de Colas. *MATerials MATemàtics*, *20*(5), 1–33. https://ddd.uab.cat/record/97914

Hillier, F. S., & Lieberman, G. J. (2013). *Introducción a la Investigación de operaciones.* (McGRAW-HILL (ed.); 9º).

Martín, Q. M. (2003). *INVESTIGACION OPERATIVA* (Hespérides (ed.)).

Render, B., Stair, Ralph M., C., Hanna, Michael E., C., Hale, Trevor S., C., & Murrieta Murrieta, Jesús Elmer, T. (2016). *Métodos cuantitativos para los negocios* (Pearson (ed.); 12ª).

Ricardo Cao Abad. (2002). *Introducción a la Simulación y a la Teoría de Colas* (Netbiblo (ed.)).

Rodríguez, R., & Gámez, A. (2002). *Investigación operativa. Teoría, ejercicios y prácticas con ordenador* (UCA).