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: María Contreras García
Tutor: Ana María Martínez Rodríguez

Enero, 2021

RESUMEN

En la teoría de colas hay diferentes modelos, en este trabajo se explican teóricamente los modelos más generales que se hallan en nuestras vidas con más frecuencia.

Los modelos expuestos en este trabajo son explicados como procesos de nacimiento y muerte, donde nacimiento indica la entrada de un cliente al sistema y muerte la salida de este. Mi estudio se basará en explicar los elementos fundamentales del sistema de colas, así como los modelos donde el tiempo entre llegadas de clientes y el tiempo de servicio seguirán una distribución exponencial. Dicha distribución es la más utilizada debido a sus propiedades de función de densidad estrictamente decreciente y pérdida de memoria, donde las llegadas solo van a depender del momento en el que llegan al sistema. Además, explicaré los modelos más destacados con distribuciones no exponenciales.

Para poder comprender todo lo anterior, es necesario clasificar el sistema de colas previamente según su estructura y elementos principales, entre otros: fuente de entrada, cola y mecanismo de servicio. El objetivo de este estudio es mejorar los sistemas de servicio y encontrar un equilibrio entre los costes y el tiempo de espera.

ABSTRACT

In the theory of tails there are different models, in this work theoretically explain the more general models found in our lives more often.

The models exposed in this work are explained as birth and death processes, where birth indicates the entry of a client into the system and death the exit from it. My study wil be based on explaining the fundamental elements of the queue systems, as well as the models where the time between customer arrivals and service time will follow an exponential distribution. This distribution is the most commonly used due to its strictly decreasing density and memory loss function properties, where arrivals will only depend on when they arrive in the system. In addition, I will explain the most outstanding models with non-exponential distributions.

In order to understand all the above, it is necessary to classify the queue system previously according to its structure and main elements, among others: input source, queue and service mechanism. The aim of this study is to improve service systems and find a balance between costs and the amount of waiting.

ÍNDICE

1. INTRODUCCIÓN

Históricamente, la teoría de colas surge gracias al matemático A. K. Erlang. A principios del siglo veinte el matemático se encontraba trabajando en una compañía telefónica, por lo que entre 1903 y 1905, estudió los conflictos de congestión de tráfico que se presentaban en las comunicaciones telefónicas de forma científica. En 1909, Erlang publicó *La teoría de probabilidades y las conversaciones telefónicas.* Más tarde, esta teoría se ha aplicado a multitud de problemas de la vida real como clientes que esperan ser atendidos en la ventanilla de un banco, máquinas averiadas que esperan ser reparadas por un técnico, pacientes en lista de espera o automóviles esperando al semáforo en rojo. Todas estas situaciones y otras muchas más, tienen en común el incómodo fenómeno de la espera, como resultado de la aleatoriedad de las llegadas de clientes y/o de los tiempos de servicio que son necesarios para atenderlos. Como consecuencia, se formará una línea de espera cuando la demanda del servicio supere la capacidad que existe para proporcionarlo. Las líneas de espera son parte de nuestra vida diaria, el tiempo que perdemos es un factor importante tanto de la calidad de vida como de la eficiencia de su economía. Habrá que asumir altos costes para evitar que los clientes no tengan que soportar largas colas y encontrarse con una línea de espera lo más breve posible para que no abandone sin haber sido atendido. Una mala gestión también puede ocasionar largas líneas de espera lo que originan costes importantes no económicos como la pérdida de clientes y de prestigio.

La teoría de colas emplea los modelos de colas para representar los tipos de sistemas de líneas de espera. Por lo tanto, para saber cómo operar un sistema de colas de la forma más eficaz, estos modelos son muy útiles, puesto que ayudan a encontrar un balance adecuado entre el coste de servicio y la cantidad de espera.

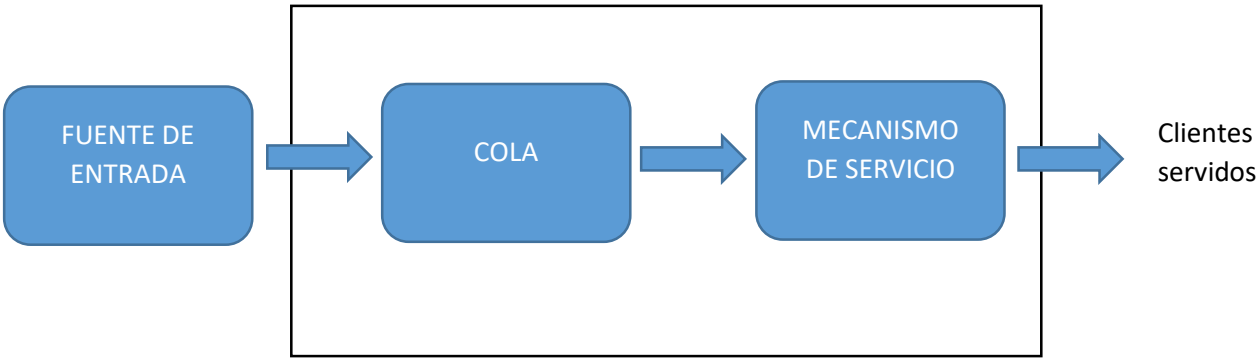
Este análisis es una de las herramientas más utilizadas para la mejora de los diferentes sistemas de producción o servicios. Para ello, a continuación, se definirán todos los elementos necesarios para comprender y entender mejor la teoría de colas.

2. ELEMENTOS DE UN SISTEMA DE COLAS

Un sistema de colas se produce gracias a la demanda de un servicio y su realización para que el cliente quede satisfecho. Para ello, los clientes que necesitan un servicio se generan en el tiempo en una fuente de entrada. Luego acceden a la instalación comprobando si se les puede atender inmediatamente, en caso de no ser así, se incorporan a la cola hasta ser seleccionados para proporcionarles dicho servicio mediante una disciplina de cola. Una vez llevado a cabo mediante un mecanismo de colas, el cliente abandona la cola, es decir, sale del sistema de colas.

Los tres componentes fundamentales que influyen en el proceso de un sistema de colas son fuente de entrada, cola y mecanismo de servicio (Figura 1).

FIGURA 1: ELEMENTOS DE UN SISTEMA DE COLAS



Fuente: Elaboración propia

2.1.-FUENTE DE ENTRADA

Se define como el proceso por el cual se generan clientes. La principal característica para definir la fuente de entrada es el tamaño, siendo este el total de clientes que solicitan el servicio. Podemos diferenciar entre un tamaño finito, en el que se limita la cantidad de personas que pueden entrar en el sistema[1], o infinito, en el que la llegada de clientes es abundante.

Por otro lado, otro factor importante es la llegada de los clientes, los cuales pueden acudir simultáneamente o independientes. Cuando ocurre el primer caso, existe la posibilidad de que la llegada de clientes sea de forma individual o en lotes (ej. Llegada de una familia a la instalación). Hay que tener en cuenta el tiempo de llegada entre ellos ya que se considera determinista si la duración se mantiene constante, pero si por el contrario varía lo llamaremos aleatorio. En la mayoría de estos sistemas los clientes llegan siguiendo un proceso de Poisson, es decir la llegada de clientes ocurre de manera aleatoria sin importar cuántos clientes hay, por

[1] *Sistema: lugar al que acceden los clientes para demandar un servicio y ser atendidos.*

lo que el tamaño de la fuente es infinito. La distribución de probabilidad entre dos llegadas consecutivas es exponencial, la cual expondré más adelante.

2.2.- COLA

Parte del sistema en el que los clientes esperan ser atendidos, se produce cuando la demanda de un servicio por parte del cliente excede la capacidad del servicio.

Una de las principales características es la capacidad de la cola, hay que diferenciar entre colas finitas e infinitas. Se denomina cola finita, cuando el número de clientes que puede esperar en la cola está limitado, es decir cuando se alcanza una longitud en la cola cualquier cliente que llegue puede ser rechazado hasta que se vuelva a disponer de sitio. Mientras que si la población tiene un gran tamaño es difícil limitar la capacidad, por tanto, cuando se da este caso se conoce como colas infinitas. Por otro lado, si el número de clientes que forman la cola es grande suele indicar que el servidor tiene baja eficiencia atendiendo.

Otra característica es la disciplina de la cola, los clientes deben seguir un orden para ser atendidos. Entre las disciplinas de colas más comunes según se encuentran las siguientes (Taha, 2012):

1.- FCFS (first-come-first-served): esta disciplina es la más común, los clientes son atendidos según en el orden en el que llegaron, por lo tanto, será atendido primero quien llegó primero. La prioridad la tiene quien llega primero.

2.- LCFS (Last-come First-served): el último cliente que llega es el que es atendido, es decir último en llegar, primero en ser atendido. Por lo tanto, para atender la cola sería en orden inverso a la llegada de los clientes.

3.- SIRO (Service-In Random-Order): denominada servicio en orden aleatorio, en este caso el usuario que accederá al servicio se sortea de forma aleatoria, es decir es irrelevante el orden en que se les proporcione servicio.

4.- Otras: existen otros métodos entre los que se selecciona al cliente en función de alguna característica, por ejemplo, por orden de prioridad, si algún cliente llega con una prioridad superior al cliente que está siendo atendido, este tendrá que retirarse para dejarlo pasar y esperar de nuevo su turno. Otros ejemplos que podemos encontrar son: atender primero el servicio que requiera menor tiempo, atender prioritariamente las emergencias, atender aquellos clientes que generen un beneficio mayor, etc.

Su última característica es el comportamiento en las colas, veremos si los clientes se unen a la cola o por el contrario deciden abandonarla. Se considera cliente paciente a aquel que se mantiene en la cola hasta ser atendiendo. Sin embargo, se califican como clientes impacientes a aquellos que se unen a la cola sólo cuando es corta (cuando creen que el tiempo de espera no es demasiado) o directamente ni se unen a ella porque hay muchos clientes. También se consideran impacientes a los que cuando se encuentran en una cola larga, se cambian a una corta para así reducir el tiempo de espera o abandonan por haber esperado demasiado.

2.3.- MECANISMO DE SERVICIO

En esta característica es importante destacar el número de servidores que nos van a realizar el servicio demandado y las estaciones de servicio, donde el cliente espera hasta ser atendido dando lugar a cada uno de los servidores. Es importante hacer referencia a la variable tiempo de servicio, tiempo que pasa desde que el cliente es atendido hasta que finaliza.

El sistema se colapsa y la cola se hace mayor si el tiempo que los usuarios tardan en salir del sistema es mayor que el intervalo entre llegadas. Por tanto, para evitar que eso ocurra es necesario tener en cuenta varias soluciones, entre ellas que el tiempo de servicio sea igual o menor que el intervalo entre llegadas, que el número de servidores se aumente o se limite la capacidad del sistema. También es importante conocer durante cuánto tiempo va a estar inactivo un servidor, sería recomendable que fuera el menor tiempo posible para así poder optimizar el rendimiento. La duración del servicio se considera aleatoria en la mayoría de los casos, aunque también puede ser constante, depende muchas veces del número de clientes que estén esperando en la cola, así se atenderán más rápido o más lento.

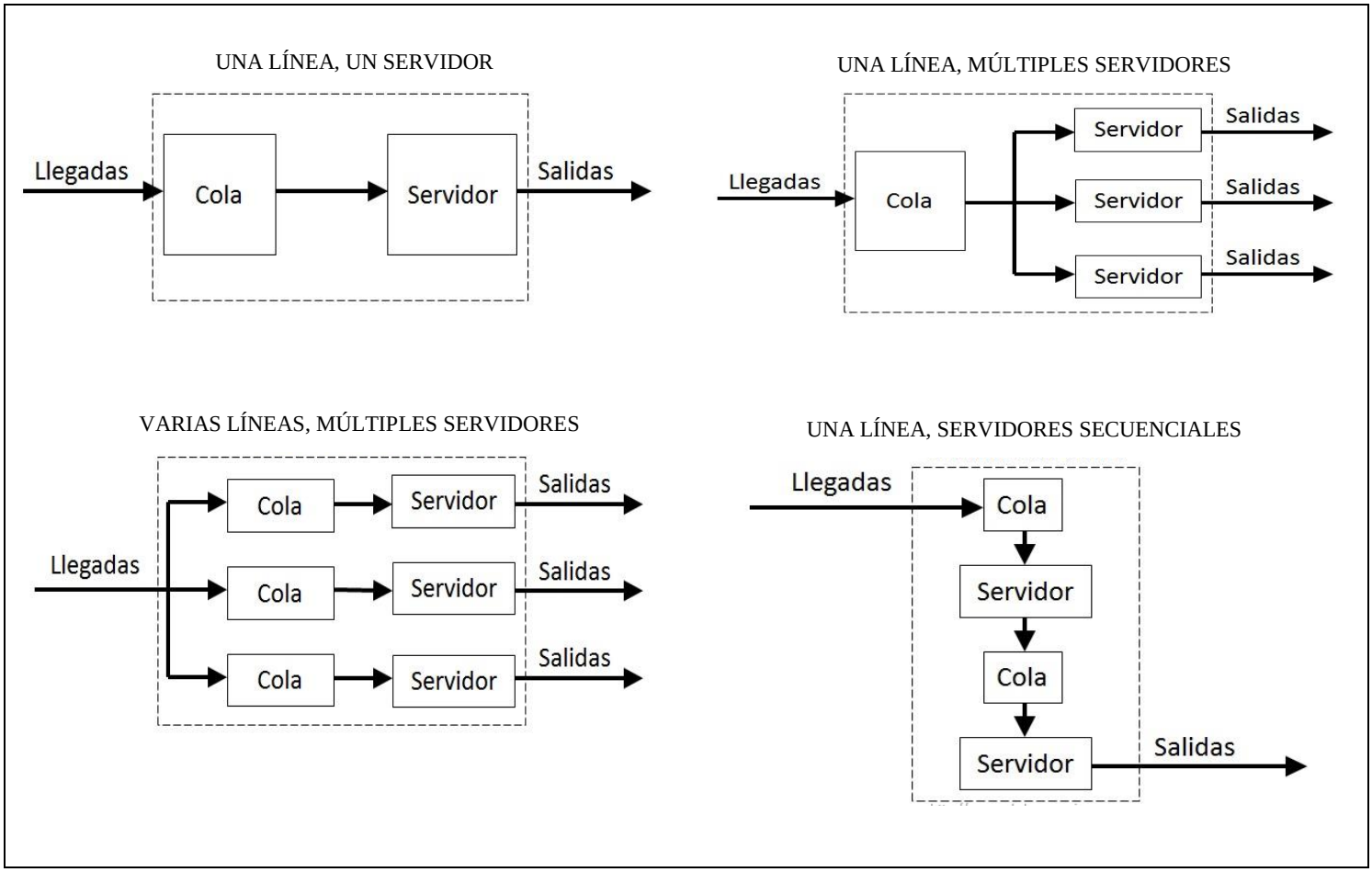
También es fundamental conocer la distribución de probabilidad de los tiempos de servicio de cada servidor, por lo general se usa la misma distribución para todos. Más adelante, expondré la distribución exponencial, que es la más utilizada. Otras distribuciones empleadas también son la distribución gamma[2] y la distribución Erlang.

De acuerdo a la disciplina de cola anteriormente explicada, los servidores van siendo ocupados por los clientes, para llegar a ellos se tiene que especificar tanto el número de servidores como el de estaciones de servicio. Para ello, veremos a continuación las diferentes estructuras de un sistema de colas según las cantidades de servidores y estaciones de

[2] *Distribución gamma: función que engloba el concepto de factorial a los números complejos. Creada entre los años 1730 y 1731 por Leonard Euler.*

servicios. Tal y como indica (Arciniega, 2019) (Echeverria, 2013) las diferentes estructuras de un sistema de colas se dividen principalmente en cuatro (Tabla 1).

TABLA 1: ESTRUCTURAS TIPICAS DEL SISTEMA DE COLAS



Fuente: Echeverria, Diego (2013). Modelos y simulación.

- Primera estructura ➡ sistema de un servidor y una línea. Los clientes se encuentran en una misma cola y pasan de uno en uno, dando prioridad a quien haya llegado primero, a la instalación del servicio, todas las necesidades del cliente son atendidas con el mismo servidor. Las llegadas se producen de manera aleatoria. Estos clientes se conocen como clientes pacientes, ya que al formar una única cola el tamaño de esta es infinita. Un ejemplo muy común de esta estructura sería los servicios de lavado automático.
- Segunda estructura ➡ se plasma una línea con múltiples servidores. Los clientes forman una única fila y cuando les toque su turno, escogen aquel servidor que esté disponible. En el caso de que todos los servidores estén ocupados, los clientes que llegan se van incorporando a la cola hasta que algún servidor quede libre. Los bancos es un ejemplo de esta estructura.

- Tercera estructura ➡ se muestra varias líneas y múltiples servidores. Cada servidor tiene su propia cola, los clientes esperan hasta que el servidor de su cola quede disponible y pueda atenderlos. Esta estructura es muy típica de los grandes supermercados.
- Cuarta estructura ➡ una línea y servidores secuenciales. Los clientes se encuentran en una única fila y avanzan de forma secuencial, pasando de un servidor al siguiente. Como ejemplo tenemos los restaurantes de comida rápida, la primera cola realizada es para tomar nota del pedido, una vez quedas atendido, pasas a la siguiente cola para recoger la comida y tomarla.

3. NOTACIÓN BÁSICA

En 1953, David Kendall insertó una notación para describir las colas y resumir las diferentes características del sistema para así clasificar por medio de iniciales los diferentes tipos de colas de espera. De esta manera, el sistema se expresa de la siguiente forma:

A/S/c/K/N/D

- A: parámetro de distribución de tiempos entre llegadas.
- S: parámetro de distribución de tiempos de servicio.
- c: número de servidores, también conocido por la inicial "s".
- K: número máximo de clientes que puede haber en el sistema (capacidad del sistema). En el caso de que sea infinito, este parámetro se puedo omitir.
- N: parámetro que indica la disciplina de cola. Se puede omitir en el caso de que siga una disciplina FCFS.
- D: tamaño de la fuente de entrada. Se puede omitir en caso de ser infinita.

Los valores que pueden tomar A y S son:

- D: tiempos determinísticos, con tiempo promedio constante.
- M: tiempos exponenciales (proceso de Poisson).
- G: tiempos generales (cualquier distribución).
- E_k: existe una distribución tipo Erlang.

4. TERMINOLOGÍA DE LAS MEDIDAS DE EFECTIVIDAD

A continuación, definimos las variables que van a influir en este sistema para así estudiar la efectividad del sistema. La terminología es la siguiente:

1. N: número de clientes en el sistema (estado del sistema)
2. N(t): se refiere al número de clientes en el instante t ($t \geq 0$).
3. s: número de servidores en el sistema de colas.
4. Longitud de la cola: número de clientes que esperan ser atendidos para el servicio. Esta longitud se calcula como el número de clientes en el sistema (N) menos los clientes que están siendo atendidos. Por lo tanto, longitud cola = N(t) – s decimos "s" porque los clientes que están siendo atendidos coinciden con el número de servidores.
5. $P_n(t)$: probabilidad de que justo en el instante t haya n clientes en el sistema, dado el número en el tiempo 0.
6. λ_n: indica la tasa media o número esperado de llegadas de nuevos clientes cuando en el sistema hay n clientes.
7. μ_n: indica el número esperado de clientes despachados o tasa media de servicio en todo el sistema por todos los servidores en su conjunto.

En el caso de que λ_n y μ_n sea constante, es decir que el número de clientes no afecte a estas variables se expresarían como λ y μ, respectivamente. Cuando tenemos varios servidores y todos se encuentran ocupados, se cumple la igualdad: $\mu_n = s \cdot \mu$. En estas circunstancias, la inversa de cada una de las tasas nos dará el tiempo esperado entre llegadas y el tiempo esperado de servicio, $1/\lambda$ y $1/\mu$, respectivamente. Como estas tasas representan los flujos de entrada y salida al sistema, el sistema de utilización de la instalación de servicio se define de la siguiente manera: $\rho = \lambda/(s * \mu)$, es decir tiempo esperado que los servidores están siendo ocupados por los clientes, puesto que $\lambda/(s\mu)$ va a representar la capacidad de servicio del sistema ($s\mu$) que utilizan los clientes que llegan (λ). **El factor de utilización** se encontrará por debajo de la unidad cuando la tasa de servicio sea superior a la tasa de llegada de clientes, mientras que si ocurre lo contrario el factor de utilización obtendrá un valor superior a 1, que nos indicará que la tasa de llegada de clientes es mayor que la tasa de servidores por lo que la cola irá creciendo. **La intensidad de tráfico** es otra medida de efectividad, se calcula entre el tiempo de llegada de los clientes al sistema y el tiempo medio de servicio de cada servidor, $I = \lambda/\mu$ por lo que el tiempo de atención de los servidores debe ser superior al de llegada para así conseguir un sistema capaz de atender todas las peticiones, por el contrario, se conseguiría un sistema congestionado y la cola aumentaría el tamaño.

Por otro lado, describimos los resultados de estado estable y sus anotaciones necesarias. Cuando un sistema de colas se encuentra en el inicio de su operación, afectado por las condiciones iniciales y por el tiempo que ha pasado desde el inicio, decimos que el sistema se

encuentra en **condición transitoria**. Sin embargo, cuando el sistema se vuelve independiente del tiempo que ha pasado y del estado inicial, se dice que ha alcanzado su **condición de estado contable o estable**. Cuando se encuentra en este estado último, la notación que se utiliza es la siguiente, las cuales mencionaré, pero más adelante explicaré de forma más detallada.

8. P_n: se refiere a la probabilidad que haya n clientes en el sistema.
9. L: número medio de clientes en el sistema.
10. L_q: longitud media de la cola, es decir solo los clientes que están esperando ya que no se tiene en cuenta aquellos que están recibiendo el servicio.
11. W: nos indica el tiempo medio total de espera en el sistema ya que se incluye el tiempo de servicio (1/μ). Se calcula mediante la fórmula de Little que describo más adelante.
12. W_q: indica el tiempo medio de espera en la cola.

5. FÓRMULA DE LITTLE

Analizando lo anterior, se pueden establecer relaciones entre las distintas medidas de eficiencia, las cuales se conocen con el nombre de fórmulas de Little. John D.C. Little (1961) la formuló para calcular el rendimiento en el sistema y mostrar la eficacia de las líneas de producción, mostró que independientemente del tiempo transcurrido, el número medio de clientes es igual a la tasa de llegadas por el tiempo de espera medio. Esta hipótesis se conoce como fórmula de Little, esto es: $L= \lambda * W$, relaciona los tiempos medios de espera en la cola con el número medio de unidades que hay en el sistema.

$W = W_q + \frac{1}{\mu}$. El tiempo de servicio se representa como $\frac{1}{\mu}$.

$L_q = \lambda W_q$. Si se supone que un cliente llega a la instalación, este entrará al servicio después de un tiempo W_q. Cuando se encuentra en la cola, decide contar los se encuentran detrás de él, lo que se conocería como L_q. Por tanto, cada uno de los L_q que han estado esperando en la cola, han tardado en llegar $1/\lambda$ con respecto al anterior. De esta forma, el tiempo que el cliente ha estado esperando en la cola ha sido L_q $(1/\lambda)$.

$L = \lambda W$. Esta expresión se conoce como fórmula de Little. Tanto esta condición como la anterior se verifican para un modelo de colas con una única entrada con llegadas exponenciales y disciplina FCFS, sin importar el tiempo de servicio.

$L = L_q + \frac{\lambda}{\mu}$. Ocurre como consecuencia de las anteriores, en los modelos de colas en los que se confirme la fórmula de Little, esta también se verifica.

$W_q = \frac{1}{\mu}L$. Cuando un cliente accede al sistema espera encontrarse con L clientes delante suya. Para poder empezar a ser atendido, tendrá que esperar a que finalice el servicio de los L anteriores. Puesto que el tiempo medio de servicio es 1/μ, el tiempo promedio de espera en la cola es $\frac{1}{\mu}L$. Las restricciones para que se cumpla esta relación son más represivas que para la fórmula de Little, por lo que se cumple solo en modelos de colas M/M/1 u otros semejantes que más adelante explicaré.

A raíz de la anterior ecuación se deduce $L_q = \lambda * W_q$ y también $L = L_s + L_q$, nos indica que el número medio de clientes es equivalente al número medio de clientes que están siendo servidos más los que están esperando en la cola. Si suponemos que el tiempo medio de servicio es una constante 1/ μ para toda n, obtendremos que la suma del tiempo en la cola más el tiempo de servicio es igual al tiempo en el sistema, por lo tanto:

$$W = W_q + (1/\mu)$$

$$T_{sistema} = T_{servicio} + T_{cola}$$

Estas igualdades también se conocen como ley de Little, gracias a ellas se puede conocer el valor de cada una de ellas, aunque solo tengas el valor de una.

6. DISTRIBUCIONES HABITUALES EN EL SISTEMA DE COLAS

Las distribuciones más conocidas tanto para los tiempos de llegadas y los tiempos de servicio son los siguientes:

6.1.- LA DISTRIBUCIÓN EXPONENCIAL

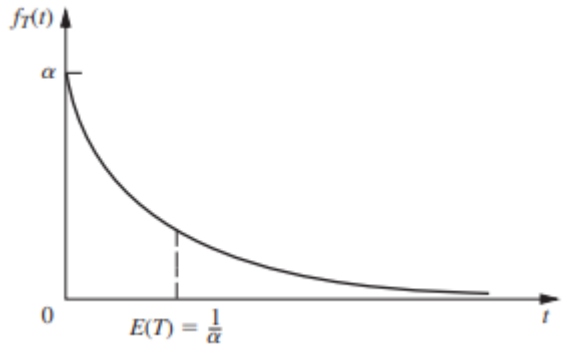
El estudio de los distintos modelos de colas está definido la mayor parte por la distribución de probabilidad de los tiempos entre llegadas y la distribución de los tiempos de servicio, las cuales, mientras no tomen ningún valor negativo, pueden tomar cualquier forma. Por lo tanto, para formular un modelo matemático que sea útil, es importante especificar la forma de cada una de las distribuciones, las cuales convienen que sean los suficientemente

realista para que el modelo ocasione predicciones razonables. En teoría de colas la distribución más importante es la distribución exponencial.

En la mayoría de los sistemas de colas, el tiempo entre la llegada de un cliente y el posterior sigue una distribución exponencial. Con esta distribución se supone que las llegadas son aleatorias, y que por lo tanto la última llegada no influye en la llegada siguiente. Lo mismo para los servicios, el tiempo de servicio sigue una distribución exponencial. Estos tiempos como ya he indicado son aleatorios y las variables se van repitiendo independientemente a lo largo del tiempo, así que se necesita de funciones que relacionen una medida de probabilidad a cada valor de las variables. El número de llegadas por unidad de tiempo y el número de servicios por unidad de tiempo sigue una distribución de Poisson, la cual es una distribución discreta. Para representar los tiempos entre llegadas o los de servicio suponemos una variable T, esta tiene una función exponencial con parámetro λ cuando su función de densidad es la siguiente (Figura 2):

$$f(t) = \begin{cases} \lambda * e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

FIGURA 2: FUNCIÓN DE DENSIDAD DE PROBABILIDAD DE LA DISTRIBUCIÓN EXPONENCIAL.



Fuente: "Introducción a la investigación de operaciones (2010)" Autores Frederick S. Hillier y Gerald J. Lieberman.

Las probabilidades acumuladas serán (para t≥0):

$$P[T \leq t] = 1 - e^{-\lambda t}$$

$$P[T > t] = e^{-\lambda t}$$

mientras que el valor esperado y la varianza de T están dadas por las siguientes expresiones respectivamente:

$$E(T) = 1/\lambda$$

$$Var(T) = 1/\lambda^2$$

Las variables aleatorias de la distribución exponencial tienen varias propiedades que las caracterizan (S.Hillier & J.Lieberman, 2010) mencionan seis propiedades, sin embargo, yo haré referencia a las tres más importantes y más conocidas.

- Propiedad 1: se refiere a una función de densidad de t estrictamente decreciente (para t≥0). La consecuencia de esta propiedad es que:

  $P\{0 \leq T \leq \Delta t\} > P\{t \leq T \leq t + \Delta t\}$

  Esta función nos indica que es probable que T tome valores pequeños. Puesto que se tiene que,

  $P(0 \leq T \leq \frac{1}{2\lambda}) = 0{,}393$ mientras que, $P(\frac{1}{2\lambda} \leq T \leq \frac{3}{2\lambda}) = 0{,}383$

  Lo que indica que es menos posible que T tome un valor cercano a su media, aunque el segundo intervalo tenga el doble de longitud que el primero, por lo tanto, lo más probable es que T tome un valor cercano a cero.

  Para razonar esta propiedad en un sistema de colas, tomaremos T como tiempos de servicio o de llegadas.

  - Si T modeliza al tiempo de servicio, tendremos que este será cercano al tiempo de servicio esperado si este servicio requiere igual duración para cada cliente y el servidor la misma cadena de operaciones El tiempo de servicio muy pequeño sería casi imposible (que quedaría por debajo de la media), puesto que se necesita de una cantidad mínima de tiempo para realizar el servicio, aunque el servidor consiga trabajar a la máxima velocidad. En esta situación, la distribución exponencial no proporciona una aproximación a la distribución de tiempos de servicio. Por otro lado, vamos a considerar que las tareas que tiene que realizar el servidor son diferentes de un cliente a otro, es decir que el tipo específico de servicio y la cantidad requerida difieren. En este tipo de situaciones, la distribución exponencial es muy posible.
  - Si T modeliza los tiempos entre llegadas, la propiedad anterior elimina situaciones en las que los clientes que llegan a la instalación tienden a posponer su entrada si observan que otro cliente ha entrado antes que ellos.

- Propiedad 2: falta de memoria.

  Se expresa como:

  $P\{T > t + \Delta t \mid T > \Delta t\} = P\{T > t\}$

  Hasta que suceda la llegada o la finalización del servicio, la distribución de probabilidad del tiempo que falta va a ser la misma, sin importar cuanto tiempo (Δt) haya pasado, es

decir, el organismo no recuerda que ha estado varias unidades de tiempo esperando. Esto ocurre con la distribución exponencial debido a

$$P\{T > t + \Delta t \,/\, T > \Delta t\} = \frac{P\{T > \Delta t\ ,\ \ T > t+\Delta t\}}{P\{T > \Delta t\}}$$

$$= \frac{P\{T > t+\Delta t\}}{P\{T > \Delta t\}}$$

$$= \frac{e^{-\lambda(t+\Delta t)}}{e^{-\lambda \Delta t}}$$

$$= e^{-\lambda t}$$

$$= P\{T > t\}.$$

- En los tiempos entre llegadas, el tiempo que transcurre hasta la llegada siguiente no está influenciado por el instante en que ocurrió la última llegada.
- En los tiempos de servicio, se presentan dos situaciones: cuando el servidor tiene que realizar la misma cadena fija de operaciones para cada cliente resulta difícil de interpretar porque se puede dar el caso de tener un servicio largo y lento. Mientras que, si las operaciones de servicio son diferentes entre clientes, puede darse que algún cliente requiera un servicio de mayor extensión que el resto.

❖ Propiedad 3: mínimo de variables aleatorias exponenciales. El mínimo de variables aleatorias exponenciales independientes sigue una distribución exponencial. Siendo estas variables T_1, T_2, …, Tn con parámetros λ_1, λ_2, … λ_n respectivamente. Por lo que U será la variable aleatoria cuyo valor es igual al mínimo de los valores que toman T_1, T_2, …, Tn, es decir, U = mín. {T_1, T_2, …, Tn}.

6.2.- LA DISTRIBUCIÓN ERLANG

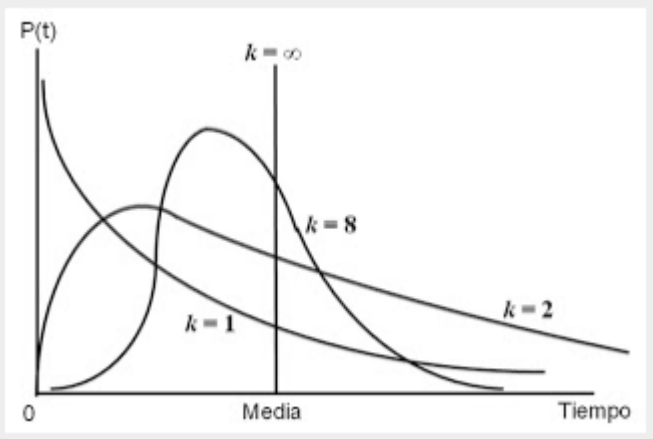
La distribución Erlang también conocida como distribución hipoexponencial, nos indica el tiempo que transcurre hasta que tiene lugar k eventos en un proceso de tipo Poisson de media λ. Cuenta con un parámetro k que determina su desviación típica: $\sigma=1/\sqrt{k}$

Si k es igual a 1, la distribución Erlang es de igual forma que la exponencial.

Si k es igual ∞, la distribución Erlang es igual a la distribución degenerada de tiempos constantes.

Por ello, la forma de esta distribución varía dependiendo del valor de k (figura 3).

FIGURA 3: REPRESENTACIÓN DE LA DISTRIBUCIÓN ERLANG SEGÚN EL VALOR DE K



Fuente: Sistema de colas (González, Sistema de colas, 2010)

7. PROCESO DE NACIMIENTO Y MUERTE

Como he expuesto anteriormente, en un sistema de colas las entradas y salidas de clientes representan a aquellos clientes que llegan y a los que se van, respectivamente. En este contexto, el término nacimiento hace referencia a la llegada de un cliente a la instalación mientras que muerte, se refiere a la salida del cliente una vez servido. El número de clientes, expresado como N(t), describe el estado del sistema en el tiempo t, siendo $t \geq 0$. Este proceso describe en términos probabilísticos como cambia N(t) cuando aumenta t. Para ello, explicaré las hipótesis del proceso de nacimiento y muerte según (S.Hillier & J.Lieberman, 2010):

- Hipótesis 1. Dado N(t) = n, la distribución de probabilidad del tiempo que falta para el próximo nacimiento es exponencial con parámetro λ_n (n = 0, 1, 2, …).
- Hipótesis 2. Dado N(t) = n, la distribución de probabilidad del tiempo que falta para la próxima muerte es exponencial con parámetro μ_n (n = 1, 2, …).
- Hipótesis 3. El tiempo que falta hasta el siguiente nacimiento (variable de la hipótesis 1) y el tiempo que falta hasta la próxima muerte (variable de la hipótesis 2) son independientes mutuamente. Esto indica que en un cambio al estado siguiente sólo ocurre un nacimiento $n \longrightarrow n + 1$ (un solo nacimiento) y en un cambio al estado anterior sólo una muerte $n \longrightarrow n - 1$ (una sola muerte).

También es importante señalar que tanto la llegada (λ_n) como el abandono de clientes (μ_n) pueden ser diferentes ante valores distintos de n, ya que podemos encontrar a clientes potenciales que rechacen entrar al sistema debido al gran tamaño de n, o que los clientes abandonen el sistema sin haber sido atendidos por el aumento del tamaño de la cola.

8. MODELO DE COLAS BASADOS EN EL PROCESO DE NACIMIENTO Y MUERTE

En este apartado expondré diferentes tipos de modelos de colas.

8.1.- MODELO DE COLAS M/M/1

Este modelo se caracteriza por que tanto el tiempo entre llegadas y el de servicio se distribuyen de manera exponencial, los clientes llegan según un proceso de Poisson, y el tiempo entre dos llegadas consecutivas y el tiempo de servicio son variables independientes. También se caracteriza por tener un único servidor. La forma ampliada de este modelo sería M/M/1/∞/∞/ FCFS, esto nos informa de que la capacidad del sistema es infinita, el tamaño de la población es infinito y la disciplina de cola seguida es "primero en llegar, primero en ser atendido". Este modelo se considera el más sencillo y, por tanto, cumple con las siguientes características:

- Tanto los tiempos entre llegadas como los de servicio siguen una función de densidad exponencial de parámetro $\lambda_n = \lambda$ y $\mu_n = \mu$, respectivamente.
- Un único servidor, es decir s = 1.

El factor de utilización del modelo M/M/1 nos mide el número medio de servidores ocupados, siendo este: $\rho = \frac{\lambda}{\mu}$ .Como tenemos un único servidor este factor coincide con el tiempo que espera un cliente nuevo en la cola hasta ser servido, este presenta un valor entre 0 y 1, depende de μ y λ. Para que el sistema no colapse o estalle, tiene que ser menor que uno, es decir, λ tiene que ser menor que μ, esto significa que el número medio de llegadas por unidad de tiempo es inferior al número medio de servicios por unidad de tiempo. Esta condición se tiene que cumplir para que exista distribución estacionaria.

Otros factores que son importantes mencionar en este sistema son:

- El número medio de clientes en el sistema (L) $L = \frac{\lambda}{\mu - \lambda}$
- El número medio de clientes en la cola (L_q) $L_q = \lambda^2 / \mu\,(\mu - \lambda)$
- El tiempo medio de cada cliente dentro del sistema (W) $W = \frac{L}{\lambda}$

  Este se descompone en la suma del tiempo medio de servicio (1/μ) y en el tiempo medio de espera en la cola (W_q), ya que para su cálculo se tiene en cuenta ambos.

  $W_q = L_q / \lambda$

En caso de inestabilidad del sistema, las fórmulas anteriores no serían aplicables, ya que estas están sujetas a una condición de estabilidad como consecuencia de que el factor de utilización debe de ser inferior a la unidad, como he explicado anteriormente. También se interpreta como $\lambda < \mu$, significa que el número medio de clientes que llegan al sistema por unidad de tiempo es inferior al número medio de clientes que pueden ser atendidos por el servidor por unidad de tiempo, en el caso de que éste estuviera todo el tiempo atendiendo a clientes (situación que es difícil que ocurra, ya que no ocurre siempre). Esto se explica por la existencia de una distribución estacionaria, ya que no se puede dar la existencia de una cola que tienda hacia el infinito. Por este motivo, podemos encontrar que en el sistema de colas no se encuentre ningún cliente, lo que se conoce como la probabilidad de que el sistema se encuentre en estado 0, siendo su ecuación $P_o = 1- (\lambda/ \mu)$.

Por último, a partir de la probabilidad del estado 0, podemos definir la probabilidad de que se encuentre n clientes en el sistema como $P_n = 1- (\lambda/ \mu)$.

8.1.1.- EJEMPLO

En una asesoría, nos encontramos con un único servidor, es decir el servicio es atendido por una persona. Las llegadas siguen un proceso de Poisson con una tasa de 8 clientes por hora durante los sábados. Los clientes son atendidos según la disciplina de cola FCFS (primero en llegar primero en ser atendido) y ninguno decide abandonar la cola sin haber sido atendido, ya que es la única asesoría que abre en fin de semana. El tiempo promedio para atender a un cliente sigue una distribución exponencial con un tiempo de 6 minutos. La solución es la siguiente,

Como se trata de un modelo M/M/1:

λ_n: tasa media de llegadas al sistema.

μ_n: tasa media de servicio.

ρ: factor de utilización del sistema: λ_n / μ_n

λ = número de clientes que llegan/unidad de tiempo = (8/60) clientes/minuto.

μ = número de servicios/unidad de tiempo = (1/6) servicio/minuto.

De esta manera, $\rho = \frac{(\frac{8}{60})}{1/6} = 0{,}8 < 1$ lo que se conoce como condición de estabilidad del sistema.

P (haya línea de espera) = 1 – P (no haya línea de espera) = 1- (P (nadie en la tienda) + P (1 cliente en la tienda) =1- (P_0 + P_1). Para que exista línea de espera debe de haber más de dos personas ya que solo hay un servidor, por tanto:

$P_0 = 1 - \rho = 1 - 0{,}8 = 0{,}2$

$P_1 = (1 - \rho)\,\rho = (1 - 0{,}8)\,0{,}8 = 0{,}16$

P (haya una línea de espera) = $1 - (P_0 + P_1) = 1 - (0{,}2 + 0{,}16) = 0{,}64$

Calculamos la longitud esperada media de la cola,

$$L_q = \frac{\lambda^2}{\mu\,(\mu - \lambda)} = \frac{(\frac{8}{60})^2}{\frac{1}{6}(\frac{1}{6} - \frac{8}{60})} = 3{,}2 \approx 3 \text{ clientes.}$$

El tiempo medio que un cliente espera en la cola sería,

$$W_q = L_q / \lambda = \frac{\lambda}{\mu\,(\mu - \lambda)} = \frac{\frac{8}{60}}{\frac{1}{6}(\frac{1}{6} - \frac{8}{60})} = 24 \text{ minutos.}$$

8.2.- MODELOS DE COLAS M/M/s

Este modelo aparece de forma abreviada, por lo que nos muestra como en el caso anterior que la capacidad de cola es infinita, el tamaño de la población es infinito y la disciplina que sigue es FCFS. La diferencia que presenta con respecto al modelo anterior es en el número de servidores ya que en este caso puede ser cualquier número natural mayor o igual a uno. Por lo tanto, se caracteriza por:

- Los tiempos entre llegadas independientes y distribuidos según la ley exponencial de parámetro $\lambda_n = \lambda$.
- El tiempo de servicio independiente y distribuido según la ley exponencial de parámetro μ.
- Varios servidores $s > 1$.

La tasa de llegadas por unidad de tiempo y la tasa de salidas por unidad de tiempo y servidor ocupado son constantes, por lo que

$\lambda_n = \lambda \quad n = 0, 1, 2.$

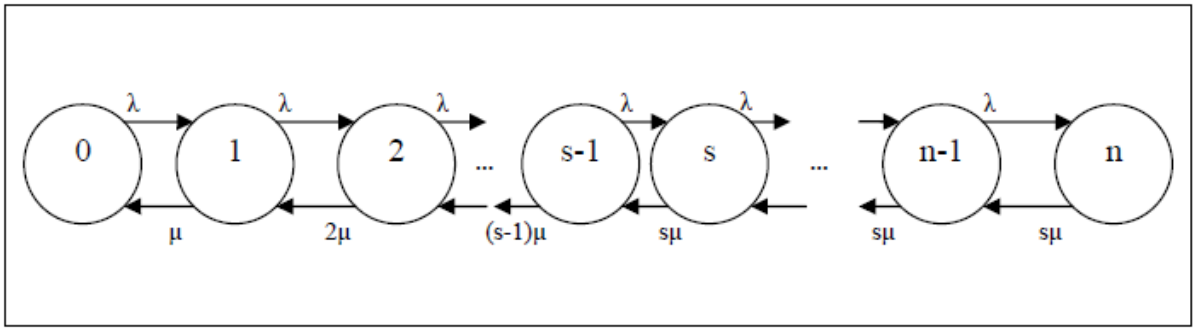
$\mu_n = n * \mu$, cuando $n \leq 0$

$\mu_n = s * \mu$, cuando $n \geq 0$

donde μ indica la tasa media de servicio de todos los servidores y μ * s, la tasa máxima de servicio para s servidores.

Como observamos en la figura 3 podemos ver las tasas de transición en el modelo M/M/s. Las tasas de llegadas se ven afectadas por la tasa media de servicio y no por el estado en el que se encuentre el sistema, esto quiere decir que según el tiempo requerido para atender al servicio van a acudir al sistema más o menos clientes, ya

FIGURA 4: DIAGRAMA DE TASAS DE TRANSICIÓN EL MODELO M/M/s



Fuente: "Introducción a la investigación de operaciones (2010)" Autores Frederick S. Hillier y Gerald J. Lieberman.

En este caso, el factor de utilización se calcula de la siguiente forma teniendo este un valor inferior a la unidad, $\rho = \lambda / (s * \mu)$.

Las expresiones de este sistema son las siguientes, (S.Hillier & J.Lieberman, 2010):

- Al igual que en el caso anterior, en este también existe la probabilidad de que ningún cliente se encuentre en el sistema de colas, siendo su ecuación la siguiente:

$$P_o = \frac{1}{\sum_{n=0}^{s-1} \frac{(\frac{\lambda}{\mu})^n}{n!} + \frac{(\frac{\lambda}{\mu})^s}{s!} * \left(\frac{s*\mu}{s*\mu - \lambda}\right)}$$

- Número medio de clientes en la cola

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

- Número medio de clientes en el sistema $L = L_q + \frac{\lambda}{\mu}$
- Tiempo medio de espera en la cola $W_q = \frac{L_q}{\lambda}$
- Tiempo medio de estancia en el sistema $W = W_q + \frac{1}{\mu}$

8.2.1.- EJEMPLO

Los trabajadores de la fábrica de Valeo, cuando realizan su trabajo tienen que llevar el producto al departamento del control de calidad, antes de que sea terminado. Esta empresa cuenta con un gran número de empleados y las llegadas son aproximadamente de 20 por hora. El tiempo medio para inspeccionar una pieza es de 4 minutos siguiendo una distribución exponencial. Vamos a calcular el número medio de trabajadores en el departamento de calidad si hay 2 inspectores de trabajo.

λ_n: tasa media de llegada.

μ_n: tasa media de servicio.

ρ: factor de utilización del sistema, $\rho = \lambda / (s * \mu)$.

Como se trata de un modelo M/M/s, el número de servidores es 2.

λ = número de trabajadores que llegan/unidad de tiempo = 1/3 trabajadores/minutos.

μ = número de servicio/unidad de tiempo = 1/4 servicio/minuto.

$$P_o = \frac{1}{\sum_{n=0}^{s-1} \frac{\frac{\lambda}{\mu}^n}{n!} + \frac{\frac{\lambda}{\mu}^n}{s!} \frac{1}{1-\lambda(s\mu)}} = 0{,}2$$

Por lo tanto, $L_q = \frac{\left(\frac{\lambda}{\mu}\right)^s \lambda\mu}{(s-1)!*(s\mu-\lambda)^2} * P_0 = 1{,}06 \approx 1$. El número de trabajadores en el departamento de calidad es 1 trabajador.

8.3.- MODELO M/M/1/K

Se trata de un modelo como el M/M/1, ya explicado anteriormente, pero con limitación en “k” para el número máximo de clientes en la cola, que coincide con la suma del número de servidores y el tamaño de la cola. Este sistema tiene cola finita, por ello la capacidad de esta sería k-s, ya que el número de clientes que pueden estar en la cola es como máximo k. Significa que se puede dar el caso en el que la cola se encuentre llena y el cliente no consiga entrar al

sistema, por lo que tiene que abandonarlo. Otra situación que se puede dar, es que la longitud de la cola sea tan larga, que el cliente no esté dispuesto a soportarla y abandone el sistema para irse a otra parte. Por lo tanto, la única diferencia que presenta con el modelo anterior es que M/M/1/k presenta cola finita mientras que el modelo M/M/s "k = ∞".

En resumen, las características de este modelo son:

- Tiempo entre llegadas es independiente y distribuidas según la distribución exponencial, $\lambda_n = \lambda$.
- Tiempo de servicio independiente y distribuido según la distribución exponencial $\mu_n = \mu$.
- Único servidor, s = 1.
- El número de clientes en el sistema no puede superar k.

La tasa de nacimiento y muerte de este modelo depende del número de clientes que se encuentren en ese momento allí, de esta manera se afirma que la tasa de llegadas por unidad de tiempo es constante hasta alcanzar la capacidad del sistema y la tasa de salidas por unidad de tiempo también es constante, es decir:

$$\lambda = \begin{cases} \lambda & \text{si } n = 0, 1, \ldots. K\text{-}1 \\ 0 & \text{si } n \geq K \end{cases}$$

$$\mu_n = \begin{cases} \mu & \text{si } n = 1, 2, \ldots. K \\ 0 & \text{si } n > K \end{cases}$$

Cuando $\lambda_n = 0$, el sistema de colas alcanzará la condición de estado contable en algún momento, aun cuando $p = \lambda / (\mu * s) \geq 1$.

$$C_n = \begin{cases} \left(\frac{\lambda}{\mu}\right)^n = \rho^n & \text{para } n = 0,1,2\ldots., K \\ 0 & \text{para } n > K \end{cases}$$

Entonces, para $\rho \neq 1$.

$$P_0 = \frac{1}{\sum_{n=0}^{K} \left(\frac{\lambda}{\mu}\right)^n}$$

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

$$= \frac{1-\rho}{1-\rho^{K+1}}$$

de manera que,

$$P_n = \frac{1-\rho}{1-\rho^{K+1}} * \rho^n \qquad \text{para } n = 0,1,2,\ldots,K.$$

Entonces,

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

$$= \frac{1-\rho}{1-\rho^{K+1}} * \rho \sum_{n=0}^{K} \frac{d}{d\rho}(\rho^n)$$

$$= \frac{1-\rho}{1-\rho^{K+1}} * \rho \frac{d}{d\rho} \sum_{n=0}^{K} \rho^n$$

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

$$= \rho * \frac{-(K+1)\rho^K + K\rho^{K+1} + 1}{(1-\rho^{K+1})(1-\rho)}$$

$$= \frac{\rho}{1-\rho} - \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}}$$

Como es usual (cuando s=1)

$$L_q = L - (1 - P_0)$$

Estos resultados no exigen que $\lambda < \mu$.

Cuando $\rho < 1$, se puede afirmar que el segundo término de la expresión de L coincide hacia 0 cuando K tiende a ∞, por lo que todos los resultados anteriores coinciden con los que se obtuvieron en el caso anterior M/M/1.

Los tiempos de espera esperados de los clientes se calculan como:

$W = L / \bar{\lambda}$ $W_q = L_q / \bar{\lambda}$ donde,

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

8.4.-MODELO M/M/s/K

Este modelo a diferencia del anterior, cuenta con más de un servidor. No permite más de k clientes en el sistema por lo que el número máximo de servidores que se puede tener es k.

El modelo M/M/s/K cuenta con las siguientes características:

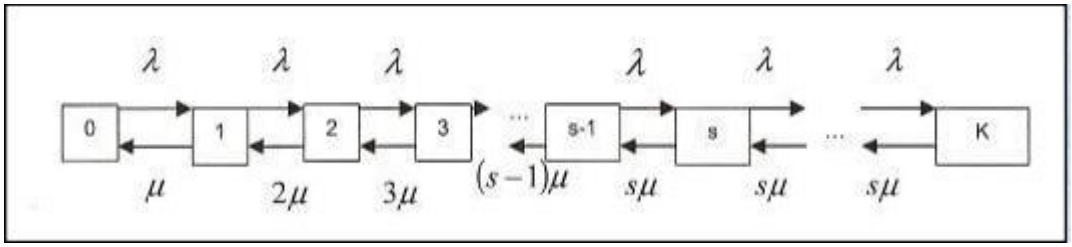
- Los tiempos entre llegadas independientes siguen una distribución exponencial de parámetro $\lambda_n = \lambda$.
- El tiempo de servicio por servidor independiente está distribuido según la distribución exponencial de parámetro μ.
- Número de servidores, s > 1.
- El número de clientes en el sistema se limita al valor positivo K. Se supone k ≥ s.

Tanto la tasa de llegadas, la tasa de salidas por unidad de tiempo y servidor ocupado es constante hasta alcanzar el límite de capacidad, es decir:

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

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

FIGURA 5: DIAGRAMA DE TASAS DE TRANSICIÓN EN EL MODELO M/M/s/K



Fuente: "Introducción a la investigación de operaciones (2010)" Frederick S. Hillier y Gerald J. Lieberman.

$$C_n = \begin{cases} \frac{(\frac{\lambda}{\mu})^n}{n!} & \text{para } n = 0,1,2,\ldots,s \\ \frac{(\frac{\lambda}{\mu})^s}{s!}(\frac{\lambda}{s\mu})^{n-s} = \frac{(\frac{\lambda}{\mu})^n}{s!s^{n-s}} & \text{para } n = s,s+1,\ldots,K \\ 0 & \text{para } n > K \end{cases}$$

Así,

$$P_n = \begin{cases} \frac{(\frac{\lambda}{\mu})^n}{n!} P_0 & \text{para } n = 0,1,2,\ldots,s \\ \frac{(\frac{\lambda}{\mu})^n}{s!s^{n-s}} P_0 & \text{para } n = s,s+1,\ldots,K \\ 0 & \text{para } n > K \end{cases}$$

Donde

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

Si se acomoda a este caso la notación L_q del modelo M/M/s, se llega a

$$L_q = \frac{P_0\,(\lambda/\mu)^s\,\rho}{s!\,(1-\rho)^2}\,[1-\rho^{K-s}-(K-s)\rho^{K-s}(1-p)]$$

donde $\rho = \lambda/(\mu*s)$. Todas las fórmulas anteriores se cumplen cuando $\rho > 0$.

Para que este sistema alcance la condición de estado estable es que $n \geq K$ aun cuando $\rho \geq 1$, así que el número de clientes en el sistema no puede crecer de forma indefinida.

W y W_q se calcula de igual forma que en el modelo anterior.

8.4.1.- EJEMPLO

Un taller mecánico tiene una zona específica para realizar las reparaciones, pero además cuenta con dos zonas acondicionadas para dejar los elementos que aún no está reparando. Cada día a este taller vienen un promedio de 2 clientes según una distribución de Poisson. Cuenta con 2 trabajadores, los cuales tardan 1 día en realizar una reparación según una distribución exponencial. Además, cuando se tiene varios arreglos para realizar, se contrata a un equipo de reparación móvil cuyo coste es de 20.000 euros dependiendo del tiempo que trabaje. El beneficio obtenido por cada reparación es de 30.000 euros.

Para representar esta situación, el modelo más adecuado es un modelo para una cola finita con más de un servidor, donde:

λ = número de reparaciones que llegan = 2 reparaciones/día.

μ = número de servicios = 1 reparación/día.

s = 3 servidores.

ρ = λ/μs = 2/3 < 1 (condición de estabilidad del sistema)

A continuación, calcularemos el tanto por ciento de tiempo que los trabajadores están sin realizar su trabajo.

$P_0 = \frac{1}{1+\sum_{n=1}^{3} C_n}$, por lo que $\begin{cases} C_1 = \lambda_0/\mu_n = 2/1 = 2 \\ C_2 = (\lambda_0 \lambda_1)/(\mu_1 \mu_2) = (2*2)/(1*1) = 4 \\ C_3 = (\lambda_0 \lambda_1 \lambda_2)/(\mu_1 \mu_2 \mu_3) = (2*2*2)/(1*1*2) = 4 \end{cases}$

$P_0 = \frac{1}{1+\sum_{n=1}^{3} C_n} = \frac{1}{1+2+4+4} = \frac{1}{11} = 0{,}0909$, significa que el 9,09% del tiempo están estos trabajadores sin trabajar.

Por último, el beneficio medio del taller es el siguiente:

Beneficio = Beneficio por reparación * λ – (Coste del equipo móvil) * P_3 siendo,

$\lambda = \lambda * (1-P_M)$ con M = longitud máxima de la cola, es decir M=3.

$P_n = \frac{\lambda_0 \ldots \lambda_{n-1}}{\mu_1 \ldots \mu_n} = P_0 = C_n * \rho_0$, así

$P_3 = C_3 * P_0 = 4 * \frac{1}{11} = \frac{4}{11} = 0{,}3636$

$\lambda = \lambda * (1\text{-}P_M) = 2 * (1 - \frac{4}{11}) = \frac{14}{11} = 1{,}2727$

Beneficio = (30.000 * 1,2727) - (20.000*0,3636) = 38.181 – 7.272 = 30.909 euros/día.

9. MODELOS DE COLAS CON DISTRIBUCIONES NO EXPONENCIALES

Los modelos de colas explicados anteriormente se basan como bien he dicho en el proceso de nacimiento y muerte, lo que significa que tanto los tiempos entre llegadas como los tiempos de servicio siguen una distribución exponencial. Cuando los tiempos de llegadas siguen una distribución exponencial quiere decir que se producen las diferentes llegadas al azar, pero no siempre es aplicable a todas las situaciones ya que algunas son de manera programada o reguladas. De igual forma ocurre con los tiempos de servicios, cuando los requerimientos de servicio son parecidos. Por ello, es necesario disponer de otros modelos de colas que lleven a cabo otra distribución de probabilidad, los cuales explicaré a continuación (S.Hillier & J.Lieberman, 2010).

9.1.- MODELO M/G/1

Este modelo dispone un único servidor y un proceso de entradas de Poisson (distribución exponencial entre tiempos de llegadas) con una tasa de llegadas fija λ. Los tiempos de servicios de los clientes son independientes con igual distribución de probabilidad, pero no se ponen restricciones sobre cuál debe ser la distribución, ya que solo es necesario conocer la media 1/μ y la varianza σ^2 de esta.

Cualquier sistema de colas de este tipo podrá alcanzar el estado estable cuando $\rho = \lambda / \mu < 1$. Los resultados de estado estable de este modelo son los siguientes:

$$P_o = 1 - \rho$$

$$L_q = \frac{\lambda^2 \sigma^2 + \rho^2}{2\,(1-\rho)}$$

$$L = \rho + L_q$$

$$W_q = L_q / \lambda$$

$$W = W_q + (1/\mu)$$

La fórmula de L_q se ha podido obtener de manera sencilla, siendo muy fácil de aplicar debido a la complejidad que representa el análisis de este modelo, ya que permite cualquier distribución de tiempos de servicio. Esta recibe el nombre Pollaczek-Khintchine, la cual establece una relación entre la longitud de la cola y el tiempo de servicio, publicada por primera vez en 1930. Se observa que cuando ocurre un aumento de σ^2, el tiempo de servicio esperado fijo 1/μ, L_q, L, W_q, W se incrementan. Este resultado nos indica la congruencia del servidor tiene gran importancia en el desempeño de la instalación de servicio, no solo en su velocidad promedio.

9.2.- MODELO M/D/s

En este modelo según (S.Hillier & J.Lieberman, 2010), el servidor realiza para todos los clientes que se encuentren en el sistema el mismo servicio consistiendo este en la misma tarea rutinaria, por lo tanto, hay poca variación en el tiempo de servicio que se requiere. La mayoría de las veces, este modelo nos proporciona que todos los tiempos de servicio son iguales a una constante fija y que tiene un proceso de entradas de Poisson con tasa media de llegadas fija λ. Cuando se dispone de un único servidor, el modelo M/D/1 es un caso especial del modelo M/G/1, donde $\sigma^2 = 0$ con lo que la fórmula de Pollaczek-Khintchine, descrita anteriormente, se reduce a $L_q = \rho^2 / 2 (1-\rho)$, a partir de esta se puede obtener L, W_q y W. Se puede observar que el valor de L_q y W_q vale la mitad que en el modelo M/M/1 que tenía tiempos de servicios exponenciales, en el que $\sigma^2 = 1/\mu^2$, entonces al decrecer σ^2 pueden mejorar las medidas de desempeño de un sistema de colas. Para el caso de más de un servidor como es este modelo se dispone de un método para obtener la distribución de probabilidad de estado estable del número de clientes en el sistema y su media.

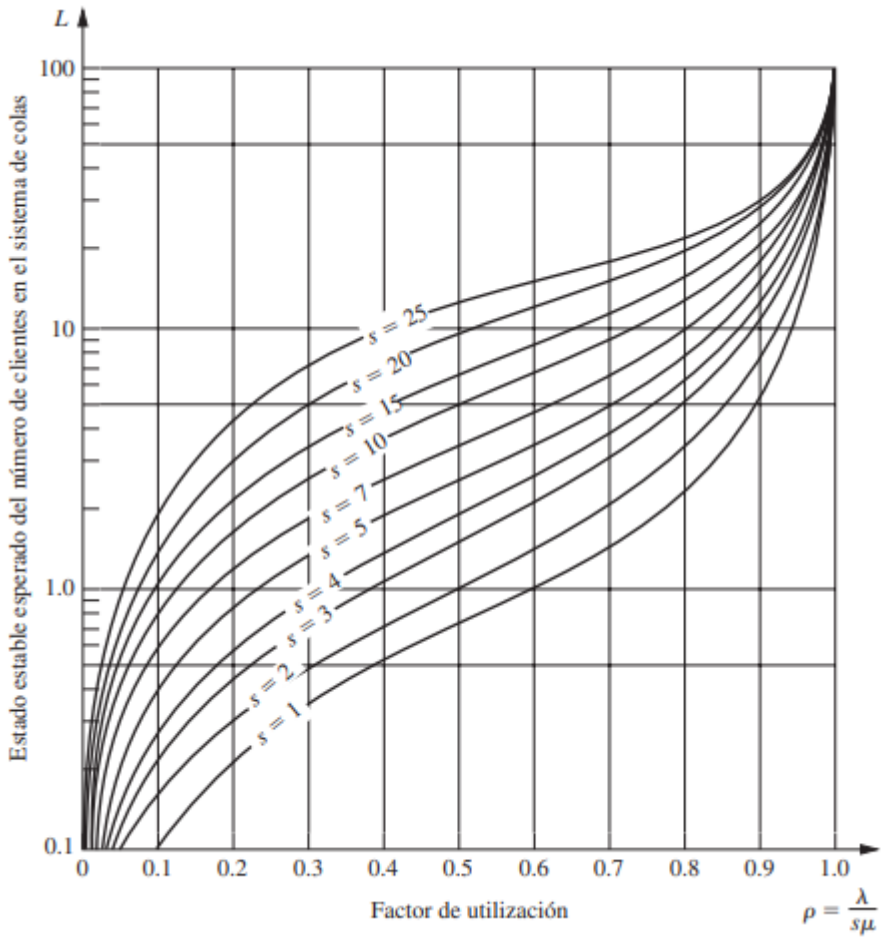
9.3.- MODELO M/E_k/s

Según (S.Hillier & J.Lieberman, 2010) "El modelo anterior M/D/s supone una alteración cero en los tiempos de servicio ($\sigma = 0$), mientras que la distribución exponencial de tiempos de servicio supone una variación muy grande ($\sigma = 1/\mu$). Entre estos dos casos extremos hay un gran intervalo ($0 < \sigma < 1/\mu$), donde decaen la mayor parte de las distribuciones de tiempos de servicio reales. Otro tipo de distribución teórica de tiempos de servicio que concuerda con este espacio intermedio es la de Erlang, la cual es llamada así en honor del fundador de la teoría de colas.

La función de densidad de probabilidad de la distribución de Erlang es la siguiente:

$$f(t) = \frac{(\mu k)^k}{(k-1)!} t^{k-1} e^{-k\mu t} \qquad \text{para } t \geq 0$$

FIGURA 6: VALORES DE L DEL MODELO M/D/s



Fuente: "Introducción a la investigación de operaciones (2010)" Frederick S. Hillier y Gerald J. Lieberman.

donde μ y k son parámetros positivos, estando k restringido a valores enteros. Su media y desviación estándar son:

$$\text{Media} = \frac{1}{\mu}$$

$$\text{Desviación estándar} = \frac{1}{\sqrt{k}}\frac{1}{\mu}$$

En este caso, como podemos ver según (S.Hillier & J.Lieberman, 2010) "k es el parámetro que indica el grado de variabilidad de los tiempos de servicio en relación con la media. La distribución de Erlang es esencial en la teoría de colas por dos motivos: el primero de ellos supongamos que T_1, T_2, . . ., T_k son k variables aleatorias independientes con una distribución exponencial semejante, cuya media es 1/k. Entonces, su suma, $T = T_1 + T_2 + \cdots + T_k$. presenta una distribución de Erlang con parámetros μ y k. La presentación de la

distribución exponencial en el punto 6.1 propone que el tiempo requerido para realizar algunas tareas podría tener una distribución exponencial. Sin embargo, el servicio total solicitado por un cliente puede incluir una secuencia de k tareas, y no sólo una, realizadas por el servidor. Si las tareas respectivas tienen una distribución exponencial idéntica de su duración, el tiempo total de servicio tendrá una distribución de Erlang; éste sería el caso, por ejemplo, si el servidor debiera realizar la misma tarea exponencial k veces para cada cliente. La distribución de Erlang también es útil debido a que es una gran familia (dos parámetros) de distribuciones que permiten sólo valores no negativos. Así, por lo general se puede obtener una aproximación razonable de la distribución empírica de los tiempos de servicio si se usa una distribución de Erlang. En realidad, tanto la distribución exponencial como la degenerada (constante) son casos especiales de distribución de Erlang con k = 1 y k = ∞, respectivamente. Los valores intermedios de k proporcionan distribuciones intermedias con media = 1/ μ, moda = (k – 1) / (kμ) y varianza = l / (kμ^2), como lo sugiere la siguiente figura. Por lo tanto, después de estimar la media y la varianza de una distribución de servicio empírica, estas fórmulas de la media y la varianza se pueden usar para elegir el valor de k que se ajuste a estas estimaciones de manera más cercana. Ahora considere el modelo M/Ek /1, que es el caso especial del modelo M/G/1 donde los tiempos de servicio tienen una distribución de Erlang con parámetro de forma = k. Cuando se aplica la fórmula de Pollaczek-Khintchine con σ^2= 1/(kμ^2) y los resultados correspondientes dados por (M/G/1) se obtiene las siguientes fórmulas" (S.Hillier & J.Lieberman, 2010):

$$L_q = \frac{(\lambda^2\ (k\mu^2)+p^2}{2(1-p)}$$

$$W_q = \frac{1+k}{2k}\frac{\lambda}{\mu(\mu-\lambda)}$$

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

$$L = \lambda W$$

Con varios servidores (M/Ek/s) se puede aprovechar la relación de la distribución de Erlang con la distribución exponencial que se acaba de describir para formular un proceso de nacimiento y muerte modificado en términos de las fases del servicio exponencial individual (k por cliente) y no en términos de los clientes.

Sin embargo, no ha sido posible derivar una solución general de estado estable [cuando $\rho = \lambda/(s\mu) < 1$] para la distribución de probabilidad del número de clientes en el sistema. Más bien se necesitaría una teoría avanzada para resolver en forma numérica los casos Individuales".

10. COSTES ASOCIADOS AL SISTEMA DE COLAS

El objetivo principal de este sistema es conseguir la minimización de los tiempos de espera y el número de personas en la cola, lo cual conlleva grandes inversiones de capital por parte de la empresa. Tanto la contratación de personal como la adecuación de las instalaciones son fundamentales para reducir el tamaño de las colas y mejorar los niveles de servicio. Estos costes por regla general son registrados contablemente de manera mensual. Los costes que voy a explicar son los siguientes:

1. Costes de mano de obra.

La mano de obra es el recurso más importante para llevar a cabo la prestación del servicio, pero también el más costoso. Dentro del coste unitario de fabricar un bien, el 80% representa la mano de obra y la materia prima. Tener contratado personal es fundamental, pero puede llevar a la quiebra al negocio si se posee exceso de trabajadores. La mano de obra de un trabajador como he dicho anteriormente es muy costosa, no solo por la compensación salarial sino además por la aportación a la seguridad social y demás costes que debe soportar. En el sistema de colas, la mano de obra se considera homogénea y sin diferencias en la compensación salarial de los empleados o servidores. El coste total de esta es: $C_{mo} = C_s * s$, donde Cs representa la compensación salarial de los servidores y S el total de estos en el sistema.

2. Costes de utilización del sistema.

Cuando se presta un servicio, uno de los costes más importantes a tener en cuenta son las instalaciones físicas. La ubicación del establecimiento es un factor de éxito para el negocio. Junto a este se suman otros más que consigue que el lugar cuente con las condiciones necesarias para desarrollar la actividad. Este coste se toma de manera mensual, en la mayoría de negocios se define de la siguiente forma: $C_{utotal} = C_{ar} + C_{adm} + C_{mant} + C_{rh} + C_{lim} + C_{imp} + C_{sp,}$ donde,

C_{ar}: costo de arrendamiento del lugar

C_{adm}: costo de administración

C_{man}: costo de mantenimiento

C_{rh}: costo de seguridad

C_{limp}: coste de limpieza

C_{im}: impuestos

C_{sp}: coste de servicios públicos.

El cual varía según el número de personas dentro del sistema (S servidores) y la capacidad de clientes (K). Por lo que para conocer el coste por persona sería $C_u = C_{utotal} / (s+k)$.

3. Costes de espera.

La espera de clientes en el sistema se refiere al tiempo que pierden estos mientras esperan ser atendidos. Este coste es un elemento clave para conocer la eficiencia de un negocio. Aunque proporcione información a la empresa no todas tienen este coste ya que resulta muy difícil calcular dicho valor debido a su naturaleza intangible. Las empresas que tengan este coste va a depender de la actividad a la que se dedique y del tamaño de espera que ha tenido que soportar el cliente. El caso más común que se ofrece a los clientes que soportan grandes colas para que estos no abandonen y permanezca en la cola es la aplicación de cupones de descuentos en su próxima compra.

4. Costes de puestos de trabajo.

Los servidores contratados para que consigan un correcto funcionamiento de sus funciones debe tener un apropiado puesto de trabajo, cuyo aumento supone un coste y mantenimiento mensual, referido en recursos físicos, preparación y elementos de oficina, entre otros. Hoy en día otro de los costes a tener en cuenta es la tecnología, siendo este: $C_{tec} = C_{pt} * s$, donde C_{pt} representa el coste mensual de tener un servidor en el puesto de trabajo, el cual incluye el valor de los equipos, licencias de software, papelería, y demás artículos de oficina como el escritorio y silla.

5. Costes de oportunidad.

Cuando se produce la entrada al sistema de clientes impacientes o con limitación de tiempo lo más probable es que este abandone el sistema, lo que representa un coste de imagen, calidad y oportunidad al no poder vender ni satisfacer las necesidades de ese cliente.

Por último, también es importante tener en cuenta costes indirectos como multas, desgastes, daños materiales, etc…

11. SOLUCIONES TECNOLÓGICAS PARA SISTEMA DE COLAS

Existen diferentes softwares que solucionan el problema del sistema de colas, los programas matemáticos son los más eficaces para ello debido al cálculo simple de sus elementos. A continuación, mencionaré algunos:

- Siempre ha sido conocido como WinQSB, pero hasta hace un tiempo esta versión dejó de funcionar y quedó obsoleta, por lo que actualmente se conoce como WinQSB 2.0 (Luis, 2012): es un sistema que ayuda a tomar decisiones para resolver los distintos problemas que surgen en el campo de la investigación operativa, gracias a las herramientas útiles que contiene y a su sencillo manejo. Está formado por diferentes módulos, uno para cada tipo de problema que se presente. En total presenta 19 módulos,

entre los cuales destaco 2, que son los que están relacionados con la teoría de colas. El primero se conoce como Queuing Analysis-QA, permite el cálculo y resolución de problemas dependiendo de su modelo, así como resolver el rendimiento de sistemas de colas de etapa simple. El segundo relacionado se denomina Queuing Analysis Simulation–(QSS) reproduce sistemas de colas simples y multietapas con componentes, incluyendo servidores, clientes y colas.

- POM-QM: es un programa que se utiliza para ciencias de la decisión, como métodos cuantitativos, investigación de operaciones, gestión de las operaciones y de la producción. Este programa funciona como el anterior, por módulos. Waiting Lines será el módulo que se utiliza para el estudio de teorías de colas. Para poder utilizarlo se necesita licencia, al cual podemos tener acceso en la página web de la Universidad de Jaén ya que esta tiene licencia como software virtualizado. Su principal ventaja es la facilidad para usarlo a partir de unas simples indicaciones. También nos ofrece información complementaria acerca del análisis de sensibilidad, gráficas de dos variables, etc.
- Programación en R: "R es un entorno de Software libre de computación estadística y gráfica, que significa que incluye de manera cohesionada y coherente una serie de facilidades para la manipulación de datos, realización de cálculos estadísticos y la visualización gráfica" (Jiménez, 2016). Existe un paquete que estudia problemas relacionados con colas y líneas de espera, así como también sirve para calcularlas. Se conoce con el nombre de Queuing y proporciona herramientas para el análisis de modelos de colas basados en el proceso de nacimiento y muerte. A diferencia de los demás programas, este destaca por su libre disposición para su uso lo que supone una gran ventaja. Además de resolver problemas relacionados con la investigación operativa también lo hace desde el punto de vista de estadística y econometría.

Actualmente la tecnología forma parte de nuestro día a día, podemos observar su éxito en los sistemas de colas y como ayudan en varios puntos de venta. Algunos ejemplos son McDonald y los supermercados Kroger Co, Wal-Mart Stores Inc (Counterest, 2017) . Cuando vamos a McDonalds nos encontramos con un sistema rápido de pedido en el que elegimos lo que queremos consumir mediante un dispositivo con pantalla táctil, de esta manera mientras el cliente realiza el pago, el trabajador va preparando el pedido. Otro caso se da en los supermercados Kroger Co que han instalado cámaras térmicas junto a un software, el cual estudia la información que han recopilado las cámaras. De esta manera, han conseguido reducir

el tiempo de espera en las colas. También otro ejemplo es el de Wal-Mart Stores Inc, escanea los productos que deseamos adquirir para acelerar el proceso de pago con su sistema Scan&Go.

Otros lugares más conocidos que han instaurado un sistema de colas con una única cola para varios servidores, que usualmente la mayoría de personas visitan y son más conocidos que los anteriores son Zara, Carrefour, H&M, Primark,etc.

Por último, mencionar el sistema de cuenta personas que determina el número de cajas a habilitar. Este tiene un coste mínimo, pero consigue reducir la cola y agilizar el proceso, además de la satisfacción y fidelización del cliente.

12. IMPACTO DEL CORONAVIRUS

Actualmente, estamos viviendo una pandemia mundial surgida de un nuevo virus llamado Covid-19. Tanto el virus como la enfermedad eran desconocidos antes de que estallara el brote en China en el año 2019. Además del número de personas fallecidas e infectadas que deja este virus, también hay que destacar su influencia en la situación económica y laboral, ya que algunos negocios se han visto tan afectados hasta tener que cerrar para siempre. El 14 de marzo se estableció el estado de alarma, por lo que la actividad laboral que no fuera esencial quedaba parada. El sector servicios se vio totalmente afectado, las tiendas, a excepción de las alimentarias, cerraron y solo podían seguir su actividad de manera online y otras empresas como bancos, asesorías y demás instituciones vieron reducidos sus horarios de atención al público.

En los últimos meses, se consiguió la reactivación de negocios, pero con algunas restricciones, entre ellas la reducción de aforo, teniendo de esta manera aforo limitado, disminución de horario, lo que hace que todas las personas se aglomeren a las mismas horas en los locales, y con distancia social entre los usuarios. En este apartado, aunque el coronavirus ha afectado a todo tipo de empresas, yo destacaré el impacto en los comercios. Las restricciones implantadas actualmente son las siguientes, afectando así a los elementos principales del sistema de colas:

- Reducción de aforo al 30% del total: el tamaño de la fuente de entrada sigue siendo infinito, pero al verse reducido el aforo del local se limita la cantidad de personas que pueden entrar, es decir la capacidad se limita. Si llegas al comercio y el aforo está completo, debes esperar fuera hasta que alguien abandone el establecimiento, por lo que podemos encontrar una larga cola en la calle para poder acceder al establecimiento.

- Reducción del horario: debido a este motivo el tiempo entre llegadas de clientes es superior, ya que, al tener menos franja horaria para realizar las compras, todos los clientes coinciden en los mismos instantes. Se produce un aumento de λ.
- La cola se encuentra reducida: el número de personas que pueden esperar hasta ser atendidos está limitado, por lo que se tendría una cola finita, la cual depende de la superficie del establecimiento. Como he dicho anteriormente, una de las recomendaciones a tener en cuenta es el distanciamiento social, el cual es muy importante tenerlo cuando te encuentras esperando en una cola, siendo este de 2 metros entre cada persona. La disciplina de cola llevada a cabo actualmente es siempre FCFS, debido a las largas colas que se generan por el distanciamiento, la persona que prevalece para ser atendida es la primera en llegar. Por último y más importante, mencionar el comportamiento en dichas colas, la mayoría de clientes al ver el gran tamaño de la cola no se unen a esta y deciden abandonar sin haber satisfecho su necesidad.
- La cantidad de servidores es menor: dada la situación, la mayoría de comercios se han visto obligados a reducir personal, por lo que cuentan con un menor número de servidores. Esto es otro de los motivos al que se debe las largas colas, el sistema se colapsa y el tiempo de servicio es superior.

Para conocer más de cerca esta situación, he realizado una encuesta, de la cual he recolectado la opinión de 18 personas cercanas a mi según sus experiencias en los diferentes comercios. He enfocado la encuesta principalmente a tiendas, ya que considero que son lugares a los que todos tenemos que acudir en nuestro día a día para comprar nuestros productos esenciales.

El enlace para acceder al formulario es el siguiente: https://forms.gle/vNAzeuBKutiQDHzZA. Las preguntas realizadas (véase anexo 1) han conseguido las siguientes conclusiones:

- El 83,3 % de los encuestados siguen acudiendo a los comercios de forma física para realizar sus compras. El 62,5% afirman que las colas se realizan tanto fuera como dentro del establecimiento, pero a pesar de estas largas colas que ven, la mitad de ellos deciden incorporarse a la línea de espera hasta ser atendidos y la otra mitad eligen por salir del comercio sin ser atendido. La mayoría se han encontrado con no poder entrar debido a que el aforo es limitado y se encuentra completo.

- El 11,1 % prefieren hacerlas por internet. El principal motivo que les lleva a ello es por las largas colas que deben esperar hasta ser atendidos. El 100% admiten que cuando la web funciona de manera lenta, abandonan la página y regresan más tarde, lo que me lleva a afirmar que estos clientes que prefieren comprar online son clientes impacientes, ya que no quieren soportar largas colas, pero también abandonan la web si se colapsa y va lenta.

Por último, estas son algunas de las soluciones que ofrecen los usuarios a los que le he realizado la encuesta para evitar las largas líneas de espera:

Con la situación que nos encontramos, creo que no se puede hacer nada
Una administración en cadena del establecimiento y un aumento del número de personal
O bien, hacer la compra por internet, ya que es lo mas facil y comodo, puesto que ni ponemos en riesgo nuestra salud ni la de las personas que trabajan en los comercios, o, por otro lado, organizarse de tal manera que la persona que quiera asistir al comercio tuviese que notificarlo por ejemplo el dia de antes, de manera que la tienda pudiese controlar cuanta gente acudiria a comprar al dia de despues y en que rango horario y asi la tienda controlar que nadie mas entre para que ni se supere el aforo ni se formen colas.
Poner cajas automáticas, así pagamos sin la necesidad de tener que esperar a ser atendidos por el dependiente
Creo que ves algo difícil pueesto que cada comercio es distinto y si no hay más superficie pues no se puede hacer nada puesto que no podremos poner más cajas o más personal
Contratar más personal

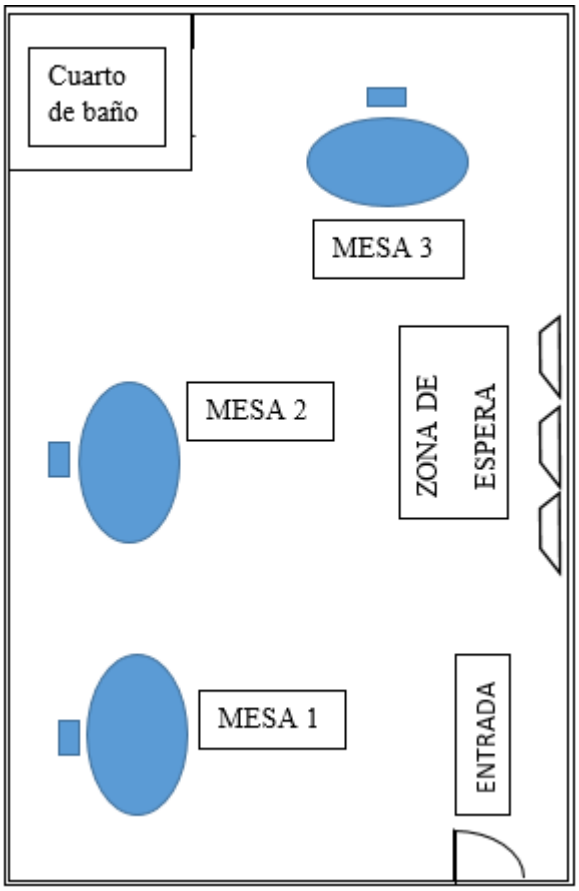
Esta es la situación que estamos viviendo hoy por hoy. Ir a los comercios o tiendas se considera una actividad de riesgo y recomiendan no pasar más de 15 minutos en el interior. Contratar más personal es la solución más mencionada para no tener que soportar tiempos de espera tan largos, así como que el tiempo de servicio sea más eficaz.

13. EJEMPLO REAL

En este apartado realizaré un estudio del sistema de colas en una asesoría situada en Martos, donde he tenido el placer de realizar mis prácticas curriculares. A continuación, aplicaré los conceptos explicados anteriormente para exponer posteriormente mi caso práctico.

Para realizar el análisis del sistema de colas lo primero que voy a exponer es una descripción general de la empresa, así como los elementos que intervienen en ella y la infraestructura donde se lleva a cabo el servicio, la cual se describe en la siguiente imagen.

FIGURA 7: MAPA ASESORÍA



Fuente: Elaboración propia.

La empresa lleva en funcionamiento más de 9 años, contando actualmente con tres profesionales que desarrollan tareas relacionadas principalmente con asesoramiento contable, fiscal y laboral entre muchas otras. Realizan de forma óptima la gestión contable, permitiendo rentabilizar los recursos de sus clientes y planificando estrategias fiscales para así alcanzar sus objetivos.

Apreciando el esquema anterior de sus instalaciones podemos observar que la empresa dispone de entrada principal, contando con espacio suficiente para la espera de los clientes. Como he dicho anteriormente, cuenta con tres trabajadores los cuales están a disposición de los clientes en los horarios siguientes:

TABLA 2: HORARIOS DE SERVICIO AL CLIENTE

DÍA DE LA SEMANA	HORARIO DE ATENCIÓN
LUNES-JUEVES	DE 9:00-14:00 DE 17:00-20:00
VIERNES	DE 9:00-14:00

Fuente: Elaboración propia.

Otro de los aspectos fundamentales a considerar son los componentes, los cuales hacen que sea posible la realización de la tarea o servicio, y son básicamente lo que expliqué en el punto 2. Ahora bien, para el caso actual de la asesoría obtenemos los siguientes elementos:

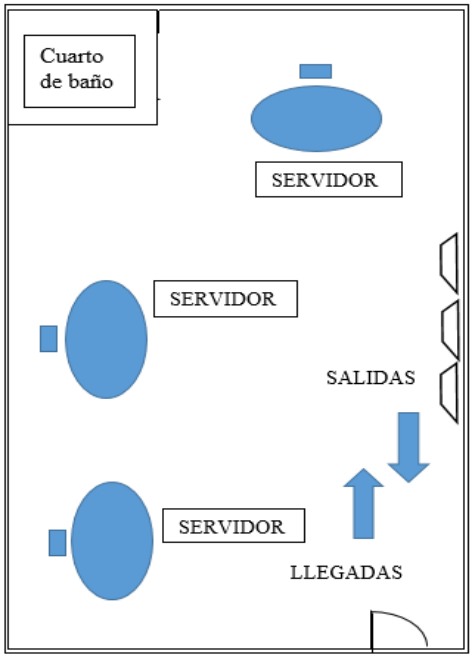
- Fuente de entrada: el total de clientes que solicitan estos servicios provienen de un tamaño infinito, ya que la cantidad de usuarios que entran no está limitada. La llegada de clientes es abundante, tanto los clientes habituales como los usuarios en general tienen acceso a ella.
- Tiempo de llegada: el tiempo de llegadas entre clientes es aleatorio, la duración entre ellas no son constantes. También se producen llegadas en masa, la mayoría de veces se da este caso cuando entran personas mayores, las cuales vienen acompañados de algún familiar.
- Comportamiento en la cola: la mayoría de clientes cuando acuden a la asesoría es para resolver algún problema urgente, por lo que aunque todos los administrativos estén ocupados y la cola sea larga, estos esperan ser atendidos. Por el contrario, solo algunos llegan y si ven una cola demasiado larga abandonan diciendo que vuelven en otro momento o más tarde, por lo que no son representativos.
- Cola: en este caso los clientes deben hacer una única cola para poder ser atendidos, si no hay ningún profesional disponible, la asesoría dispone de un espacio equipado con sillas para que los clientes esperen ahí hasta ser atendidos.
- Capacidad de cola: se considera infinita, no se limita en ningún momento el número de clientes en la cola.
- Disciplina de la cola: el orden que llevan siempre para atender es la disciplina FCFS, los usuarios se atienden según en el orden en el que llegaron, por lo que la prioridad la tiene quien llega primero.
- Mecanismo de servicio: se cuenta con 3 administrativos a disposición de los clientes que se van generando en la cola. El tiempo de servicio varía para cliente ya que depende del servicio que demanden. En horario de tarde, hay que destacar las pocas

entradas que se producen, por lo que los profesionales aprovechan para realizar trabajos atrasados o acumulados que tengan.

A continuación, hablaremos de la estructura que caracteriza a esta empresa. Se pueden destacar dos estructuras según el servicio que se demande, siendo estas: una línea, varios servidores y una línea, varios servicios secuenciales.

Esta estructura es la más usada, el cliente entra al establecimiento, si algún servidor se encuentra disponible se dirige a su mesa para ser atendido, por el contrario si todos los servidores están ocupados espera en la cola a su turno. Una vez que el servicio demandado por el cliente es satisfecho, este abandona el sistema. Podemos ver este proceso en la siguiente imagen:

FIGURA 8: ESTRUCTURA UNA LÍNEA, VARIOS SERVIDORES



Fuente: Elaboración propia.

A continuación, los tiempos de llegadas y de servicio, los recogí durante una mañana con ayuda de un cronómetro, para así obtener datos más exactos. El número de clientes que acudieron a la empresa fueron 10 personas, para conocer la distribución que siguieron habría que aplicar un test de bondad de ajuste. La muestra obtenida es: 2,2,1,3,2 y con esto aplicaría el test de bondad de ajuste, pero habría sido necesario contar con más observaciones.

TABLA 3: RECOGIDA DE TIEMPOS

CLIENTES	HORA DE ENTRADA	TIEMPO DE SERVICIO
1	9:05	35’
2	9:47	54’
3	10:02	96’
4	10:40	6’
5	11:10	15’
	DESCANSO	30’
6	12:01	3’
7	12:28	14’
8	12:45	21’
9	13:15	19’
10	13:42	7’

Fuente: Elaboración propia.

Con los datos anteriores, considero que el tiempo entre la llegada de un cliente y el posterior sigue una distribución exponencial, estas llegadas son aleatorias al no influir la última llegada en la siguiente. Como podemos observar en la tabla anterior, los servidores se encuentran inactivos durante 30 minutos, por lo que ningún cliente podrá demandar un servicio en ese tramo horario.

Lo aconsejable para resolver este caso con más profundidad, sería simplificar esta situación a un modelo que esté estudiado e implementado en un software, como podría ser el POM-QM. Habría que tener en cuenta varias consideraciones como si la afluencia no es similar en todos los horarios habría que hacer un modelo para cada horario en el que el comportamiento sea similar, o solo para uno de ellos.

Aunque no haya podido concluir el ejemplo real aplicado a un modelo exacto, creo que ha servido para conocer los elementos de colas, estructura y tiempos de llegadas de dicha asesoría en referencia al sistema de colas.

14. CONCLUSIÓN

La realización de mi trabajo de fin de grado, me ha servido para entender sobre un tema que desconocía por completo a pesar de tenerlo presente en mi día a día. Las líneas de espera son muy usuales en nuestra vida cotidiana y aunque no lo creamos, tienen una gran repercusión en la actividad de las empresas. A lo largo de este trabajo, he explicado que las colas son formadas cuando la demanda de un servicio excede de la oferta efectiva, por ello las empresas deben de tomar decisiones acerca de la cantidad de servicios que son capaces de ofrecer. Sin embargo, en la mayoría de los casos es muy difícil predecir cuándo llegarán los clientes y cuánto tiempo será necesarios emplear para satisfacer el servicio. Estar preparado para ofrecer todo tipo de servicio en cualquier momento, conlleva unos costes excesivos, pero no contar con el personal suficiente genera largas colas que los clientes tienen que pagar con su tiempo. Las largas líneas de espera no solo producen costes económicos para la empresa sino también pérdida de prestigio y pérdida de clientes. El espacio del que dispone cada empresa, así como su distribución, influye en la estructura que va a utilizar para ofrecer el mejor servicio, las cuales fueron vistas anteriormente en la tabla 1. La elección de un tipo de estructura aporta al empresario determinar un modelo de colas u otro para así optimizar clientes y su beneficio.

Como última conclusión, mencionar la importancia de un software en la empresa para que solucione el problema del sistema de colas, alcanzando gran eficiencia y evitando la aglomeración de clientes, consiguiendo también de esta manera la satisfacción de ellos.

Por último, a partir de los resultados obtenidos en el formulario realizado, podemos observar que los comercios se han visto afectados actualmente por el Covid-19, no solo por las restricciones implantadas, si no por las opiniones de los clientes, ya que son demasiados los que abandonan o ni entran a realizar sus compras por tal de no soportar las grandes colas ocasionadas.

15. BIBLIOGRAFÍA

Arciniega, P. (septiembre de 2019). *Estructuras Típicas de sistemas de Colas*. Obtenido de https://sites.google.com/site/pearciniega357/home/estructuras-tipicas-de-sistemas-de-colas

Counterest. (23 de octubre de 2017). *Innovación para reducir tiempos de espera en colas*. Obtenido de https://counterest.net/reduciendo-tiempo-espera-colas/

Echeverria, D. G. (27 de mayo de 2013). *Modelos y simulación*. Obtenido de https://sites.google.com/site/dgecheverria1992/estructuras-tipicas-de-sistemas-de-colas

González, G. (2010). *Sistema de colas*. Obtenido de http://investigaciondeoperacionesunilibre3.blogspot.com/2010/10/sistemas-de-colas-modelo-basico-un.html

Jiménez, P. C. (julio de 2016). *Análisis estadísticos con R.* Obtenido de https://masteres.ugr.es/moea/pages/curso201516/tfm1516/caaadillajimenez_tfm/!

Luis, M. I. (2012). *Modelo de líneas de espera con WinQSB*. Obtenido de Investigación operativa 2012: https://es.slideshare.net/ivanmorales7393/lineas-de-espera-investigacin-operativa-2012

S.Hillier, F., & J.Lieberman, G. (2010). *Introducción a la investigación de operaciones.* Mc Graw Hill.

Siqueiros, G. F. (febrero de 2013). *Introducción a la teoría de colas y su simulación.* Obtenido de https://lic.mat.uson.mx/tesis/044_Gerardo_FabianPS.pdf

Taha, H. A. (2012). *Investigación de operaciones.* Pearson.

16. ANEXOS

Anexo 1: Preguntas realizadas en el cuestionario.

La pregunta principal de este cuestionario que llevará a contestar unas u otras preguntas es la siguiente.

Actualmente, ¿acude a los distintos comercios para realizar la compra?

- Si.
- No, prefiero hacer todas mis compras online.
- Tal vez.

Si la opción anteriormente es "Si" o "Tal vez", las preguntas a contestar son:

1. ¿Dónde se encuentran las personas realizando la cola?
   - Dentro del establecimiento.
   - En la calle.
   - En ambos sitios, tanto fuera como dentro del establecimiento.
2. ¿Se ha encontrado con algún establecimiento que tenga el aforo limitado y no lo dejen entrar?
   - A elegir entre las opciones "Si" o "No".
3. Al llegar al establecimiento, se ha encontrado una larga cola. ¿Cuál ha sido su decisión?
   - A elegir entre las opciones "Realizar mi compra y esperar hasta ser atendido" o "No realizar mi compra y abandonar el sistema"
4. ¿Cuánto tiempo ha estado esperando en la cola hasta ser atendido?
   - Respuesta corta.
5. Una vez que está siendo atendido, ¿ha notado que el tiempo de servicio sea mayor o sigue siendo el mismo de siempre?
   - Mi servicio a diferencia de antes tarda en realizarse más de 5 minutos.
   - Mi servicio a diferencia de antes tarda en realizarse más de 10 minuto.
   - Mi servicio una vez que está siendo atendido la duración es igual que antes.
6. ¿Qué solución daría para que no hubiera largas colas y el sistema no se colapsara?
   - Respuesta corta.

Por el contrario, si la opción a la pregunta principal es "No, prefiero hacer todas mis compras online", las preguntas a contestar son:

1. Actualmente con la pandemia realizo mis compras online
    - debido a las largas colas que se generan en las tiendas físicas
    - debido a la frustración de llegar a la tienda y no poder entrar porque el aforo se limita
    - siempre las he realizado por internet por comodidad
2. Con el Covid-19 he notado la web de las tiendas online colapsada y su funcionamiento más lento
    - Si.
    - No.
    - Tal vez.
3. Cuando la web funciona de manera lenta
    - prefiero esperar y seguir realizando mi pedido.
    - abandono la página y regreso más tarde.
4. ¿Se cumple el tiempo estimado de entrega o suele venir más tarde de lo esperado?
    - A elegir entre "Mi pedido siempre está en casa con puntualidad cumpliendo la fecha de entrega" o "Mi pedido llega más tarde de lo esperado".
5. ¿Cree que han aumentado las compras por internet?
    - Si.
    - No.

Las respuestas recogidas fueron 19, la mayoría de personas encuestadas realizan sus compras acudiendo al establecimiento, en el siguiente gráfico podemos ver las proporciones.