FACULTAD DE CIENCIAS EMPRESARIALES
Y TURISMO DE OURENSE

Trabajo de Final de Grado

Teoría de colas

Aplicación y comparativa de los modelos M/M/1 y M/M/S

Andrés López Fernández

Grado en Administración y Dirección de Empresas

Curso 2015-2016

FACULTAD DE CIENCIAS EMPRESARIALES
Y TURISMO DE OURENSE

Trabajo de Final de Grado

Teoría de colas

Aplicación y comparativa de los modelos M/M/1 y M/M/S

Autor:

Andrés López Fernández

Tutor:

Manuel Alfredo Mosquera Rodríguez

Grao en Administración y Dirección de Empresas

Curso 2015-2016

Agradecimientos

De una parte quiero agradecer y agradezco al equipo docente de la Facultad de Ciencias Empresariales y Turismo de Ourense, por haber contribuido con su trabajo diario a mi formación, habiéndome capacitado para afrontar este punto y seguido en mi vida académica.

A mi tutor, Don Manuel Alfredo Mosquera, gracias a sus innumerables consejos he podido finalizar de forma satisfactoria esta ardua labor.

Mi familia, por supuesto, no sin poca paciencia, que me ha apoyado durante la realización de este trabajo; desde las horas de trabajo de campo hasta las noches escribiendo sin fin, para borrarlo la mañana siguiente.

Por y para todos ustedes.

"*Afortunado es el hombre que tiene tiempo para esperar"* Pedro Calderón de Barca

Resumen

En este trabajo se presenta un estudio teórico de los principales modelos de colas. Bajo una perspectiva clara y sencilla, se realizará una explicación sobre elementos y características, así como las posibles aplicaciones de modelos simples (M/M/1), multicanal (M/M/S), avanzados y redes de colas.

Como complemento al marco teórico, posteriormente, se realizará un estudio empírico en un hipermercado. Bajo condiciones reales, modelizaremos el comportamiento de este sistema de colas, para hallar el consiguiente resultado y en su caso poder tomar las pertinentes decisiones.

Palabras clave:

Teoría de colas, modelizar, modelo de cola, tiempo de espera, servidor, red de colas, distribución Exponencial, distribución de Poisson.

Índice de contenido

Introducción 8
Objetivos 9
Antecedentes 10

Marco Teórico 11
1. Aplicaciones de la teoría de colas 11
2. Elementos y estrutura de un sistema de colas 11
3. Parámetros y medidas de eficiencia 13
4. Proceso de nacimiento y muerte 15
5. Notación de Kendall 17
6. Fórmulas de Little 18
7. Modelos elementales de teoría de colas 19
7.1 Modelo M/M/1 19
7.2. Modelo de múltiples canales y una sola fase M/M/S 23
8. Modelos avanzados de colas 25
8.1 Modelo M/M/1K 25
8.2 Modelo M/M/S/K 26
8.3 Modelo M/M/1/∞/H 26
9. Redes de colas 27
9.1 Redes de Jackson abiertas 28
9.2 Redes de Jackson cerradas 28

Estudio empírico 29
10.1 Ámbito de aplicación 29
10.2 Metodología y parámetros básicos 30
10.3 Medidas de eficiencia para el sistema M/M/10 36
10.4 Variaciones de eficiencia sobre el modelo M/M/S 39
10.5 Comparativa de eficiencia: M/M/10 y diez sistemas M/M/1 40

10.6 Análisis económico 42

10. Conclusiones 46

Bibliografía 47
Bibliografía adicional 48

Índice de Tablas

Tabla 1. Premuestra Tasa de llegadas 30
Tabla 2. Premuestra Tasa de Servicio 31
Tabla 3. Muestra tasa de llegadas 32
Tabla 4. Muestra tasa de servicio 33
Tabla 5. Ficha Técnica. Análisis de la tasa de llegadas 35
Tabla 6. Ficha Técnica. Análisis de la tasa de servicio 36
Tabla 7. Parámetros principales 37
Tabla 8. Resumen de los resultados del modelo 39
Tabla 9. Variaciones de eficiencia 40
Tabla 10. Comparativa de eficiencia 41
Tabla 11. Medidas eficiencia para Escenario B 44
Tabla 12. Medidas de eficiencia para Escenario C 45

Índice de Ilustraciones

Ilustración 1. Sistema de colas 12
Ilustración 2. Distribución de Poisson y Distribución Exponencial 15
Ilustración 3. Resprentación de un modelo M/M/S/K 26
Ilustración 4. Representación de una red de colas 27
Ilustración 5. Representación de un sistema M/M/10 39

Introducción

El estudio de los sistemas de colas no ha sido una decisión baladí. ¿Es más eficiente una sola cola en común o una para cada servicio? Es una pregunta que todos nos hemos hecho a la espera de recibir un servicio. Existen diversos sistemas para gestionar la espera de los clientes y efectivamente no todos manifiestan la misma eficiencia de cara al consumidor.

Desde que existe la prestación de servicios, ya sea realizar una llamada telefónica, comprar medicamentos en una farmacia, echar gasolina o comprar el pan, existe la probabilidad de que el servidor esté ocupado, y se tenga que esperar.

El crecimiento del sector servicios originó, amén de millones de puesto de trabajo, nuevos problemas a la hora de tomar decisiones. El cálculo de los servidores necesarios para la prestación de un servicio, evitando las indeseables colas, bajo una condición de calidad esperada, es uno de los principales problemas operativos de las empresas.

La investigación operativa trabaja bajo el paraguas de la estadística para solventar numerosos dilemas de la decisión diaria. Este es el cuadro donde aparecen numerosos ámbitos: programación lineal, teorías de juegos, modelos de transporte, teoría de colas…

Enmarcada, la teoría de colas ha sido la elección para este trabajo porque a pesar de su relevancia en la toma de decisiones de las empresas, no figura en el catálogo docente de nuestra facultad. Por lo tanto, y en primer lugar, la justificación para estudiar los diferentes ítems relacionados con esta especialidad, es la de ampliar mis conocimientos al respecto, de los que, en mi humilde opinión, son necesarios.

Este trabajo se estructurará en dos partes diferenciadas, la primera un marco teórico en el cual se abordarán los modelos necesarios para poder explicar lo relevante de la materia, y posteriormente una segunda parte con una aplicación eminentemente práctica con los resultados y conclusiones relevantes, sobre los cuales se podrán tomar decisiones al respecto.

Atendiendo a la metodología que se desempeñará. Después de leer a numerosos autores, así como otros trabajos de la misma índole que éste, se ha observado que se suelen realizar resúmenes de la materia en sí, obteniendo trabajos densos y estrictamente matemáticos, por lo tanto y con el beneplácito del tutor; este trabajo se realizará, para la mejor comprensión, bajo una perspectiva sencilla adaptada a nuestro perfil del empresario.

Para la ampliación de los conocimientos matemáticos, se pone a disposición bibliografía especializada en la materia, en la cual se podrá comprobar los recursos utilizados así como las fórmulas y teoremas que fundamentan los modelos.

Objetivos

Como futuro graduado en Administración y Dirección de Empresas, necesito enfocar este último trabajo en la universidad como complemento formativo. Cabe esperar que después de realizar un trabajo de esta índole, terminaré conociendo en profundidad las aplicaciones prácticas de esta disciplina. Por eso y sabiendo que mis ambiciones futuras pasan por dedicarme al sector servicios trazo así el primer objetivo:

Dominar los modelos principales de teoría de colas para poder aplicarlos en el futuro.

En segundo lugar, partiendo de la consecución de mi primer objetivo, se realizará este trabajo, en la medida de lo posible, desde el punto de vista del decisor. Considerando que cualquier desarrollo o cálculo matemático debe dar lugar a un resultado con el cual se puede tomar decisiones operativas. Con todo esto se fija el segundo objetivo:

Analizar los modelos de teoría de colas de una forma aplicable, sencilla y enfocada desde el punto de vista del gestor.

Si cabe desgranar este objetivo en sub-objetivos, debo explicar que se realizará un análisis de la funcionalidad de la teoría de colas, así como un breve repaso de los antecedentes. Se detallará el origen de las fórmulas de aplicación de los modelos principales y se explicará con ejemplos la composición del resto de los modelos.

Hasta aquí los objetivos de la parte teórica. En cuanto a la segunda parte del trabajo, el estudio empírico. Se analizará el comportamiento de un modelo de colas real M/M/S, el hipermercado Carrefour de Ourense, y por consiguiente se marca como objetivo:

Poner en práctica el comportamiento del modelos M/M/1 y M/M/S bajo hipótesis reales.

Relacionado con este, se realizará un análisis económico sobre las posibilidades de aplicación de un modelo u otro, para conocer las posibles diferencias entre los modelos en cuanto a variables económicas y por lo tanto se marca como objetivo final:

Obtener la información necesaria para poder tomar decisiones al respecto.

Antecedentes

Agner Krarup Erlang (Dinamarca 1878 – 1929), es considerado el padre de lo que hoy conocemos como Teoría de Colas. La aplicación de sus estudios de ingeniería y estadística para resolver la congestión del tráfico telefónico, en las centralitas de recepción de llamadas, asentó la base para el desarrollo posterior de numerosos modelos de espera que hoy en día se explican y sobre todo ayudan a tomar importantes decisiones operativas.

Los estudios de Erlang se fundamentaron en el cálculo del número de centralitas necesarias para intercomunicar dos pueblos teniendo en cuenta los tiempos de espera. Cada llamada debía pasar por un operador que la enviara, en aquella época la compañía telefónica solo podía transferir las llamadas de una en una. Además no debemos olvidar que un tiempo elevado de una llamada incrementaban el tiempo de espera de las demás, entonces ¿Cuántos servidores necesitaba para intercomunicar cada localidad?

Tras dos años de estudio, Erlang publicó "*Telephone waiting times*", donde probaba que la llegada de las llamadas telefónicas era aleatoria y seguían una distribución de Poisson[1]. A partir de ese estudio, llegó su trabajo más importante "*Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges*", en el cual detallaba ya fórmulas para el cálculo de los tiempos de espera en base a una distribución exponencial[2].

E.C Molina (España) en 1927 avanzó sobre los estudios de Erlang, con *Aplicación de la teoría de la probabilidad a problemas de líneas telefónicas,* posteriormente fue T.C Fry con *Probabilidad y sus usos en ingeniería*. Durante la década de los 30, numerosos autores de diversas nacionalidades: A.N Kolmogorov y A.Y Khintchine (Rusia), C.D Crommelin (Francia), C.Palm (Suecia), realizaron aportaciones a las teorías de Erlang. (Cao, 2002)

A partir de los años 50, las aportaciones se multiplicaron debido a las nuevas aplicaciones posibles en el ámbito de la informática y de las telecomunicaciones.

[1] Distribución de Poisson: distribución de probabilidad discreta, utilizada para modelizar el número de sucesos que se pueden producir en un intervalo de tiempo, bajo supuestos de aleatoriedad. (Montero , 2007)

[2] Distribución Exponencial: distribución de probabilidad del tiempo de espera entre dos hechos que sigan un proceso de Poisson. (Montero , 2007)

Marco Teórico

1. Aplicaciones de la teoría de colas

El estudio de los modelos de espera nos ayudan a tomar decisiones con respecto a diferentes factores que influyen directamente a la atención al cliente, un tiempo de espera elevado por parte de un cliente producirá un alto deterioro en la calidad percibida del servicio, por lo tanto y haciendo la pertinente referencia a Cao (2002) los objetivos del estudio de colas son:

- Conocer el nivel óptimo de funcionamiento de un sistema de atención, para minimizar su coste global.
- Evaluar la repercusión de las distintas modificaciones a aplicar sobre el sistema.
- Fijar un balance equilibrado que permita una atención cualitativa aceptable sin una repercusión cuantitativa muy relevante en coste.

Se debe tener en cuenta que la modelización de un comportamiento social está afectada por la personalidad de cada individuo. La paciencia de los individuos que forman la cola influirá de forma positiva o negativa sobre la percepción de calidad de servicio en un mismo sistema.

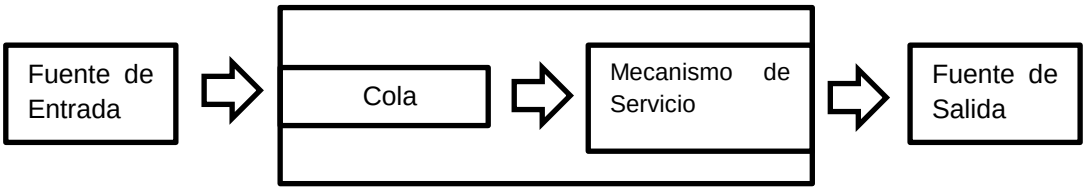
2. Elementos y estrutura de un sistema de colas

La llamada cola se produce cuando la demanda real de servicio es superior a la capacidad de atención. Asimismo debemos diferenciar entre la cola y el sistema de colas.

La cola, propiamente dicha, es un conjunto de clientes que están esperando para recibir el servicio, es decir que ya han solicitado atención, pero aún no han sido llamados al mecanismo de atención. Sin embargo el sistema de colas es el conjunto formado por la cola y el mecanismo de servicio (Cao, 2002).

Para explicar los sistemas de colas, comenzaremos por un sistema de colas básico compuesto por una fuente de entrada, una línea de espera y un servidor que sigue una disciplina de cola FIFO (first in, first out), se atenderá el primer cliente que haya llegado.

Ilustración 1. Sistema de colas



Fuente: Elaboración propia.

Como podemos observar en la ilustración 1, la fuente de entrada suministra individuos al sistema, los cuales esperan a ser atendidos si el mecanismo de servicio está ocupado. Una de las características más relevante de la fuente de entrada es su tamaño (población potencial). El considerarla finita o infinita modifica de forma sustancial el estudio del proceso. La otra característica que debemos tener en cuenta es en función de que parámetros obtenemos las entradas en el sistema. Atendiendo al tiempo entre llegadas (distribución de exponencial), o tasa de llegadas (distribución de Poisson).

En cuanto a la cola en sí, como ya mencionamos con anterioridad, es donde los clientes esperan a ser atendidos, una vez lo hayan solicitado. A parte de la disciplina de cola, debemos tener en cuenta la capacidad de la cola. De acuerdo con Carro (2009), solo supondremos la cola finita cuando su límite sea pequeño y se alcance con frecuencia, como por ejemplo una pequeña peluquería.

Para terminar de explicar el proceso. Los elementos de la cola pasarán a ser atendidos en el mecanismo de servicio, el cual puede estar compuesto por varias estaciones de atención, e incluso en fases sucesivas. El tiempo de atención será el factor determinante y la homogeneización de este fomentará el equilibrio del sistema.

Cómo ejemplo práctico, podemos observar todo este proceso en la caja de una sucursal bancaria. Los clientes entran en el banco y se sitúan detrás del último cliente que espera en la cola. El cajero atiende a cada cliente por orden de llegada, disciplina FIFO. El tiempo medio de atención, determinará junto con la tasa de llegada de los clientes, la eficiencia del sistema.

A continuación se explicarán los parámetros a utilizar y los consiguientes resultados que podemos obtener con ellos para analizar la eficacia de cualquier sistema de colas.

3. Parámetros y medidas de eficiencia

Al igual que en cualquier otro modelo matemático o estadístico la terminología es fundamental para comprender los diferentes parámetros. Siguiendo a Taha (2012) se indicará la simbología básica, necesaria para la comprensión de los diferentes modelos que detallaremos a lo largo del trabajo.

Los parámetros que determinan las características de todo sistema son solamente dos, aunque podemos expresarlos en función de las unidades medias (clientes) o en función del tiempo que transcurre entre que entra una unidad y otra, expresándose de la siguiente forma:

- Parámetros principales (en función de las unidades medias)

λ = **tasa promedio de llegada**: número medio de unidades que entran en el sistema en un intervalo de tiempo

μ = **Velocidad media del servicio**: número medio de unidades que se atienden en el servidor en un intervalo de tiempo

- Parámetros transformados (en función del tiempo medio que transcurre)

$1/\lambda$ = **Tiempo promedio entre llegadas**: tiempo medio que transcurre entre una y otra llegada al sistema.

$1/\mu$ = **Tasa media de servicio**: tiempo medio que utiliza el servidor para atender una unidad

Para obtener estos parámetros, tenemos dos opciones: contar las unidades que entran en el sistema y se atienden en el servidor para conocer el promedio, o bien medir el tiempo que transcurre en promedio desde la entrada de una unidad a otra. Obteniendo uno podemos hallar el otro, y viceversa.

Una vez obtenidos los parámetros, los cuales se obtienen de la misma forma para cualquiera que sea el sistema, podremos calcular las llamadas medidas de eficiencia de un sistema de colas. El cálculo de estas sí que varía de forma sustancial en función de las características del modelo que estemos estudiando.

- Medidas de eficiencia

Se entiende como medidas de eficiencia todos aquellos resultados que podamos obtener en base a los parámetros y que sirvan para tomar decisiones. En sí mismas, estas medidas pueden ser tenidas en cuenta no solo por el gestor del sistema, sino que

también por los clientes. El tiempo promedio de espera en la cola, por ejemplo, es un factor de interés para ambos. (Cao, 2002)

Por lo tanto se distinguen las medidas como enfocadas a la gestión (factor de utilización y probabilidad de sistema vacío) y medidas enfocadas al cliente (número promedio de unidades en cola, y en el sistema, tiempo promedio de espera en cola y en el sistema).

ρ = **Factor de utilización del servidor**: se refiere al tiempo que realmente trabaja el servidor en atención al cliente.

P_0 = **Probabilidad de sistema vacío**: este parámetro se refiere a la probabilidad que cero unidades se encuentren en el sistema en determinado período. Se podrá hallar la probabilidad que se encuentren cualquier número de individuos dentro del sistema

L_q = **Número promedio de unidades en la cola**: se refiere al número de unidades, por término medio, que se encuentran esperando recibir servicio.

L_s = **Número promedio de unidades en el sistema**: se refiere a las unidades que se encuentran en el sistema, en promedio, tiene en cuenta las que esperan en la cola y las que están siendo atendidas.

W_q = **Tiempo promedio que espera en la cola**: se refiere al tiempo que trascurre desde que ingresa el cliente al sistema, hasta el momento en que es atendido por el servidor.

W_s = **Tiempo promedio en el sistema**: tiempo total medio, desde el momento en que entra una unidad al sistema, espera en la cola y es atendido.

Bajo la perspectiva de la gestión, el factor de utilización del sistema parece ser el factor a optimizar a priori, para que los servidores no estén ociosos, aunque como veremos en el estudio empírico, la calidad de atención de un servicio estará mejor enfocada bajo medidas de eficiencia dirigidas a los usuarios, como por ejemplo el tiempo medio de espera para ser atendidos.

Conocida la simbología se declara:

"*La condición uniforme en Teoría de Colas, se logra solamente cuando μ es mayor que λ, es decir que la velocidad de servicio debe ser superior a la velocidad de llegadas para que se presente la condición uniforme. Por el contrario cuando μ es menor o igual que λ el sistema de colas es inestable, ya que la línea puede acumularse potencialmente al infinito, debido a que las unidades llegan con mayor rapidez, en comparación a la prestación del servicio*." Schroeder; 167.

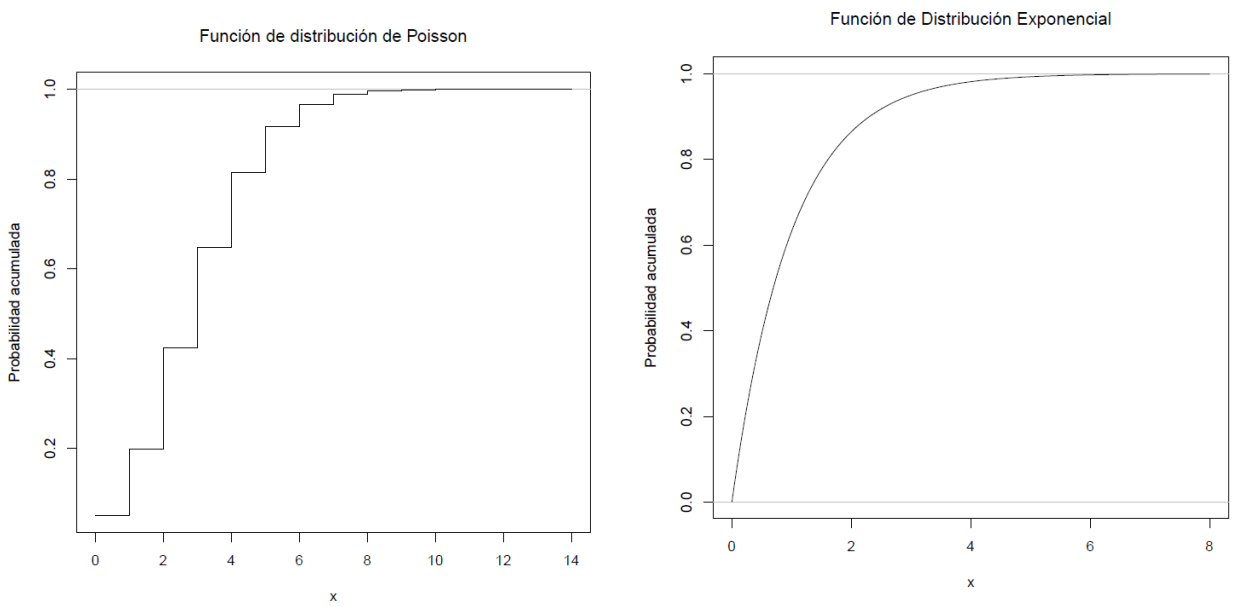
El llegar a la estabilidad o condición uniforme, como determina Schroeder, es una meta necesaria para todo gestor, evitar el colapso del sistema. Es decir, que los clientes no puedan acceder al sistema para hacer cola y esperar ser atendidos.

4. Proceso de nacimiento y muerte

En primer lugar cabe explicar la relación existente entran la distribución de Poisson y la distribución Exponencial.

Una distribución de Poisson es una distribución discreta, como se puede observar en la Ilustración 2, su función de distribución acumula la probabilidad por intervalos de tiempo. Es decir, obedece a los sucesos que ocurren en un determinado momento del tiempo. Sin embargo la distribución Exponencial, su función es continua, obedece al tiempo que ocurre entre los sucesos.

Ilustración 2. Distribución de Poisson y Distribución Exponencial



Fuente: Elaboración propia

Podríamos modelizar un sistema de colas en función del número de llegadas, en base a una distribución de Poisson:

"*Se entiende como proceso de Poisson el utilizado para contabilizar el número de ocurrencias de un fenómeno hasta un instante t*" Cao; 122.

Sin embargo la mayoría de los modelos actuales están fundamentados en procesos de nacimiento y muerte. Proceso que aplica el tiempo entre sucesos para modelizar en base a una distribución exponencial, por ser más práctico de obtener los parámetros.

Cabe señalar la relación entra las variables, si medimos por ejemplo el tiempo promedio de atención de un servidor, el cual se define:

$$\frac{1}{\mu}$$

Obtenemos el número de clientes que se atienden por minuto:

$$\mu$$

Y de forma análoga, el tiempo que trascurre entre cada llegada al sistema, el cual se define:

$$\frac{1}{\lambda}$$

Obtenemos el número de clientes que se llegan en promedio al sistema:

$$\lambda$$

Para mostrarlo de una forma práctica; para obtener la tasa de llegadas de un sistema se podría contar las unidades que llegan en momentos determinados de tiempo al sistema. Por ejemplo en un banco se contaría la gente que entran durante un intervalo de tiempo determinado, para obtener un promedio. De igual modo, la obtención de la tasa de atención, se contaría la cantidad de clientes que se atienden en un intervalo de tiempo.

Para obtener las tasas de llegada y atención obedeciendo a un proceso de nacimiento y muerte, exponencial, en el mismo caso, bastaría con medir el tiempo que transcurre en promedio entre cada llegada, en el caso de la tasa de llegadas, y el tiempo de atención para calcular la tasa de atención.

Entendiendo nacimiento como llegadas al sistema, este proceso explica que las llegadas se producen en base a una distribución exponencial. Es decir en un sistema con N individuos en un momento t, N (t), es más probable que haya una llegada o una salida del sistema en un periodo corto de tiempo. Además permite que las tasas de entradas y salidas dependan de la cantidad de individuos de la población (Cao, 2002)

5. Notación de Kendall

A partir de la teoría de este proceso, detallaremos los diferentes modelos que componen la teoría de colas como parte de la investigación operativa. Para ello debemos explicar la notación de Kendall[3] (notación *A/B/s*)

Siguiendo a Winston (2004), la notación de Kendall indica la composición y estructura del modelo a estudiar de la siguiente forma.

El primer término, generalizado como *A,* hace referencia a la distribución probabilística del tiempo entre una llegada y la siguiente. Bajo este índice, las distribuciones más relevantes hacen referencia a exponencial (M), determinística[4] (D), uniforme[5] (U) o genérica[6] (G).

En segundo lugar, el término *B,* corresponde a la distribución que explica el tiempo de servicio, con las mismas variantes que el anterior.

En cuanto al indicador *s,* se utiliza para referenciar el número de servidores que atienden las demandas de servicio.

Posteriormente la notación *A/Bs* fue ampliada para dar cabida a nuevos factores determinantes como la capacidad de la cola (K), el tamaño potencial de la población (H), y la disciplina de cola (Z). Con respecto a esta última se tiene en cuenta si la disciplina implica prioridad (FIFO o LIFO) o bien la atención se baraja en una disciplina general aleatoria según demanda, en este caso (GD).

Con todo esto los modelos de teoría de colas se indicaran siguiendo la notación ampliada, y para que figure referenciado se detallan 2 ejemplos.

1. M/M/1/10/∞/FIFO

Las llegadas al sistema se suceden siguiendo una distribución exponencial, al igual que el tiempo de servicio. En cuanto al número de servidores en este caso es uno y el máximo de unidades en cola son diez. Bajo la condición de población potencial infinita y con un régimen o disciplina de cola de FIFO.

Para visualizar el modelo de una forma práctica podemos enmarcar en estas condiciones de sistema a cualquier comercio minorista de una ciudad, una carnicería,

[3] David G. Kendall formuló en 1953 la notación que se utiliza para determinar la composición de los modelos de colas, notación *A/B/s.*
[4] Determinístico: constante, igual para cada unidad. (Montero, 2007)
[5] Uniforme: de intervalos iguales, y rango igualmente probable (Montero, 2007)
[6] Genérica: no se conoce nada sobre la distribución (Montero, 2007)

pescadería o panadería en la cual sirva un solo dependiente y no quepan en espera más de 10 clientes.

2. M/D/2/∞/∞/FIFO

Indica, como la anterior, que las llegadas se suceden son tiempos entre llegadas exponenciales, sin embargo el tiempo de atención del servidor es determinístico (constante para todas las unidades), existen dos servidores y tanto el tamaño de la cola y la población se considera infinita, en cuanto a la disciplina, por orden de llegada o solicitud. En cuanto a /∞/∞/FIFO referentes a la capacidad de la cola, tamaño potencial de la población y disciplina, son los indicadores que consideramos por defecto, por lo tanto todos los modelos que cumplan estos requisitos podríamos simplificarlo en M/D/2, en este caso. (Cao, 2002).

Como ejemplo de un sistema con estas características, en la que el tiempo de atención sea constante, encontramos los servicios automatizados, la entrada a un parking, por ejemplo. Los coches hacen cola con tasas de llegada exponenciales pero la tasa de atención, lo que tarda en abrir la barrera desde que pulsas el botón es para todos igual. Entonces la tasa de servicio es constante; así, un parking con dos entradas podemos modelizarlo como un sistema M/D/2.

6. Fórmulas de Little

John Little[7] en 1961 verificó que en sistemas de colas modelizados bajo una distribución de tiempo de servicio y llegada exponencial, la relación entre los números medios de clientes, con el tiempo medio de un cliente dentro del sistema se manifestaba de la siguiente forma.

Siendo las tasas de llegadas constantes Little estableció la que hoy se conoce como primera fórmula de Little.

$$L = \lambda W$$

Esta fórmula manifiesta que si el tiempo promedio en el sistema es W, y λ, como ya se indicó, el número medio de unidades que entran, L representa el número de unidades medias que están en el sistema en cualquier momento.

De igual razonamiento obtenemos la conocida como segunda fórmula de Little, que según Cao (2002), realmente no fue formulada por Little, pero se le reconoce por ser análoga a la anterior.

$$L_q = \lambda W_q$$

[7] John Little: Físico estadounidense

L_q hace referencia al número promedio de unidades en cola, como ya se indicó en las medidas de eficiencia.

Ambas fórmulas se fundamentan, como se explica, en tasas de llegadas constantes, pero no siempre es válido este razonamiento. En cambio se puede generalizar mediante:

$$L = \bar{\lambda}\, W$$

$$L = \bar{\lambda}\, W_q$$

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

Entendido p_n como la probabilidad de que haya un número n de unidades en el sistema.

Será $\bar{\lambda}$, tasa de llegada determinada, teniendo en cuenta su probabilidad de ocurrencia.

Si atendemos al tiempo, otra relación relevante es la del tiempo total de un cliente en un sistema, como la suma de tiempo en cola más tiempo servicio. (Cao, 2002).

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

7. Modelos elementales de teoría de colas

Como hemos visto en el apartado anterior, cada sistema es diferente y sus combinaciones muy numerosas. Sin embargo la base de todos es ellos es el modelo M/M/1, tasa de llegadas y atención bajo una distribución aleatoria exponencial y con un solo servidor.

A continuación pormenorizaremos las características del modelo simple M/M/1 y de múltiple canal M/M/S, además se darán indicaciones básicas sobre modelos avanzados y redes de colas.

7.1 Modelo M/M/1

"*El modelo simple, se conoce como M/M/1, es decir, que es un modelo de espera con llegadas aleatorias, distribución de servicio aleatorio y un sólo canal de servicio*." De la Fuente; 586.

El modelo deberá cumplir las siguientes condiciones:

"*1ª Las llegadas son atendidas sobre la base primero en entrar, primero en salir, y cada una de las llegadas espera el servicio, haciendo caso omiso de la longitud de la cola.*

2ª Cada entrada es independiente de la anterior.

3ª Las llegadas son descritas por una distribución de probabilidad de Poisson y provienen de una población infinita.

4ª Los tiempos de servicio varían de un cliente al siguiente y son independientes unos de otros.

5ª Los tiempos de servicio ocurren de acuerdo con la distribución de probabilidad exponencial." Render; 356.

Bajo estas condiciones podemos desarrollar las siguientes fórmulas de aplicación directa para modelos M/M/1 (Schroeder, 1992).

- Factor de utilización del servidor:

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

- Probabilidad de que el sistema esté vacío:

$$P_0 = 1 - \frac{\lambda}{\mu}$$

- Número promedio de unidades en cola:

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

- Número promedio de unidades en el sistema:

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

- Tiempo promedio de espera en la cola:

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

- Tiempo promedio en el sistema:

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

No obstante a continuación y siguiendo a Cao 2002 avanzaremos sobre la obtención de las formulas anteriores.

Teniendo en cuenta las características de este modelo, comenzamos por deducir la tasa de llegadas:

$$\lambda_n = \lambda, para\ todo\ n = 0, 1, \ldots$$

Y las tasas de servicio como:

$$\mu_n = \mu, para\ todo\ n = 1, 2, \ldots$$

Siendo λ_n número medio de unidades que llegan al sistema y μ_n número medio de unidades que son atendidos en el sistema

Utilizando las fórmulas para resolver un proceso de nacimiento y muerte (ver Cao, 2001), se tiene que:

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} c_n} = \frac{1}{1 + \frac{\rho}{1-\rho}} = \frac{1}{\frac{1-\rho+\rho}{1-\rho}} = \frac{1}{1-\rho}$$

Dónde:

$$c_n = \frac{\lambda_{n-1}\,\lambda_{n-2}\ \ldots\ \lambda_0}{\mu_n \mu_{n-1} \ldots\ \mu_1} = \frac{\lambda^n}{\mu^n} = \left(\frac{\lambda}{\mu}\right)^n = \rho^n, para\ n = 1, 2, \ldots$$

Siendo $\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} \rho^n$ una serie geométrica, y será convergente si $|\rho| < 1$, además de $\rho > 0$ .Quiere decir que no será posible que la tasa de utilización $\rho \geq 1$ (100%). Otra interpretación, y remitiéndonos a la condición uniforme explicada en apartados anteriores es que siempre que $\lambda < \mu$, se dice que el sistema es estacionario.

En conclusión sobre lo explicado, el número medio de clientes que entran en el sistema por unidad de tiempo será menor que el número medio clientes que pueden ser atendidos en la misma unidad de tiempo.

Para continuar y suponiendo que el sistema es estacionario, se debe calcular la suma de la serie de las c_n:

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} \rho = \frac{\rho}{1-\rho}$$

Además, para cualquier $n \geq 1$, se obtiene que $p_n = c_n . p_0 = (1-\rho)\rho^n$, la masa de probabilidad de la variable "número de clientes en el sistema":

$$P(N = n) = p_n = (1-\rho)\rho^n, n = 0, 1, \ldots$$

Así N tiene una distribución geométrica de parámetro $1 - \rho$, probabilidad del número de fracasos antes del primer éxito. En base a la fórmula de la media de esta distribución, se puede calcular directamente $L = E(N)$ (número medio de individuos en el sistema) como:

$$X \sim Geometrica\ (p) = n$$

$$E[x] = \frac{1-p}{p}$$

$$L = E[n] = \frac{1-(1-\rho)}{1-\rho}$$

Con razonamiento análogo podemos obtener L_q (número medio de individuos en cola)

$$L_q = L - (1 - p_0) = \frac{\rho}{1-\rho} - \big(1 - (1-\rho)\big) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

A partir de las fórmulas de Little siguiendo lo que se ha obtenido se pueden expresar las fórmulas que hacen referencia a los tiempos de espera tanto en el sistema W, o en cola W_q:

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

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

7.2. Modelo de múltiples canales y una sola fase M/M/S

"*Se usa cuando la demanda es grande, por lo que es necesario que varios servidores presten un mismo servicio. Los clientes forman una sola cola y el que se encuentre en primer lugar, acude al primer servidor que se desocupe para ser atendido*. "Krajewski; 330.

Este modelo es la generalización del modelo anterior, por lo que se mantiene la suposición de que la población potencial y la capacidad de cola son infinitas, de igual manera que se sigue trabajando con los mismos parámetros y medidas de eficiencia.

Será el modelo que nos facilite el análisis de multitud de servicios, cualquiera en los cuales exista más de un servidor: (banco, supermercado, gasolinera…etc.)

A continuación se expondrán las fórmulas de aplicación para el modelo M/M/S, así mismo y por la complejidad del desarrollo de las mismas no se detallará el proceso para el cual se facilitará bibliografía especializada.

- Factor de utilización del servidor:

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

Al igual que el modelo M/M/1 el factor de utilización dependerá de la tasa de llegada y atención de los servidores, teniendo en cuenta ahora s (número de servidores).

- Probabilidad de que en sistema haya un número n de unidades o clientes o en su caso de que el sistema esté vacío:

$$p_0 = \frac{1}{1 + \sum_{n=1}^{s-1} \frac{\lambda^n}{n!\,\mu^n} + \frac{\lambda^s}{(s-1)!\,\mu^{s-1}(s\mu - \lambda)}}$$

De acuerdo con Cao (2002) está formula puede generar errores si s es muy elevado, ya que generaría denominadores muy elevados, por lo tanto se presenta este cálculo como recursivo de c_n (relación tasas de llegada con tasas de servicio) de la siguiente forma:

Partiendo de:

$$\lambda_n = \lambda \; \forall n$$

$$\mu_n = \begin{cases} n\mu & \forall\, n \leq s \\ s\mu & \forall\, n > s \end{cases}$$

A partir de:

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

Si se define $c_0 = 1$

$$c_n = c_{n-1}\frac{\lambda_{n-1}}{\mu_n} = c_{n-1}\frac{\lambda}{n\mu}, para\ n = 1, 2, \dots, s-1$$

Para mostrarlo de forma más clara, podemos observar porque se definen expresiones recurrentes, por ejemplo:

$$c_2 = c_1\frac{\lambda_0}{\mu_1}$$

$$c_3 = c_2\frac{\lambda_1}{\mu_2}$$

$$c_4 = c_3\frac{\lambda_2}{\mu_3}$$

Para terminar, se obtiene la expresión final, la cual será la aplicada en la parte práctica.

$$p_n = c_n p_0 = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n}p_0 & si\ n = 1, 2, \dots, s \\ \dfrac{\lambda^s}{s!\,\mu^s}\rho^{n-s}p_0 & si\ n = s+1, s+2, \dots \end{cases}$$

- Número promedio de unidades en cola:

$$L_q = \frac{\lambda^{s+1}p_0}{(s-1)!\,\mu^{s-1}(s\mu - \lambda)^2}$$

- Tiempo promedio de espera en la cola:

$$W_q = \frac{L_q}{\lambda}$$

- Tiempo promedio en el sistema:

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

- Número promedio de unidades en el sistema:

$$L_s = \lambda W$$

8. Modelos avanzados de colas

A continuación se harán indicaciones básicas sobre tres modelos en extensión a los anteriores, los cuales tienen en cuentan factores de aplicación real. El desarrollo matemático de estos, es demasiado complejo y se escapa del objetivo del trabajo, se facilitará bibliografía al respecto.

8.1 Modelo M/M/1K

Siguiendo a Cao (2002), este variación del modelo M/M/1 aplica una limitación en el tamaño de la cola. Se entiende K como el número límite de individuos o elementos que pueden hacer cola. Sin más variaciones al respecto, el tiempo de llegadas y de atención se explica bajo una distribución exponencial, la población potencial se supone infinita y la disciplina de cola FIFO.

El problema que se trata de modelizar en esta variante es que si la cola está llena y el servidor no avanza en su proceso de atención los nuevos clientes no pueden entrar en el sistema. Se trata de una limitación lógica y esperada ya que las condiciones reales de la mayoría de los servicios aplican restricciones sobre el tamaño de la cola, sea físicamente, por razones lógicas, o informática en el caso de las telecomunicaciones.

Como ejemplo, en una peluquería en la que solo trabaje un peluquero. Considerando la clientela potencial infinita, cualquiera puede entrar. El tamaño de la cola es limitado, podrán esperar el número máximos de persona que quepan en teoría en el local, y permita desempeñar la labor de atención.

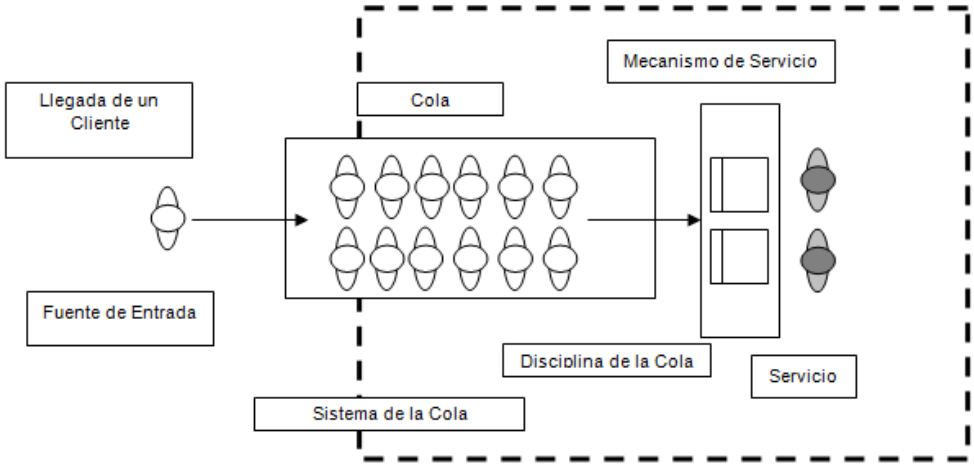
Con esto, se considera un sistema colapsado cuando un individuo no accede a hacer cola para esperar a ser atendido, no solo si físicamente no puede, sino que también si no lo hace porque considera que el tiempo de espera es demasiado elevado. Se conoce como probabilidad de bloqueo. Este factor dependerá por supuesto de cada persona, no

se debe olvidar que el estudio del comportamiento humano no es posible modelizarlo por completo.

8.2 Modelo M/M/S/K

"Se trata de la extensión del modelo anterior al caso en que existe un número entero cualquiera, s, de servidores. Las tasas de llegada son idénticas a las del modelo M/M/1, mientras que las de servicio son exactamente iguales a las M/M/S" Cao; 158.

Ilustración 3. Representación de un modelo M/M/S/K



Fuente: Imagen libre de autor

Esta variante no es más que una limitación natural al modelo anterior, además de limitar el tamaño de la cola, definimos de forma concreta el número de servidores, yendo un paso más cerca de un modelo de aplicación más práctico.

Como ejemplo, la oficina de una entidad bancaria en la cual el mecanismo de servicio esté compuesto por más de un servidor. Los clientes según entran en la oficina esperan en una sola cola a ser atendidos por el servidos que primero quede libre.

La ilustración 3 hace referencia al modelo antes indicado, como podemos observar dos servidores prestan el servicio simultáneamente, el tamaño de la cola es limitado y los clientes son atendidos por orden de llegada, disciplina FIFO.

8.3 Modelo M/M/1/∞/H

Como indica su nomenclatura, en este modelo limitamos la población potencial. Ahora debemos tener en cuenta que las llegadas dependerán de cuantos clientes se encuentre dentro del sistema. En este modelo suponemos un tamaño de cola infinito para ser servido en un solo servidor.

Para observar de forma práctica este modelo nos situaremos en única tienda de un pueblo cualquiera, si la atención se da solo por una persona. Nunca consideraríamos el sistema colapsado ya que el que tenga la necesidad de ir a comprar esperará, en teoría, hasta que le llegue su turno porque no tiene otra opción.

Ahora bien las llegadas de clientes a la tiende dependerá de los clientes que ya se encuentren allí. Es decir si en el pueblo hay cincuenta habitantes y en la cola de la tienda esperan diez, existe más probabilidad de que lleguen más clientes que si en la cola hubiese más de diez.

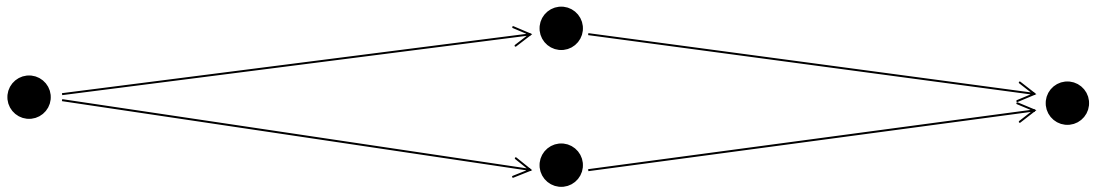
9. Redes de colas

"Una red de colas no es más que una red en la que cada nodo está constituido por el sistema de una cola. Se trata, por tanto de un grafo orientado en el que se pueden producir transacciones de clientes que salen servidos de un nodo hacia otro nodo. La forma más habitual para modelizar el modo en el que los clientes servidos en un nodo se dirijan a otro es considerado que lo hacen de acuerdo con una distribución discreta" Cao; 181.

En teoría de colas se entiende por nodo un subsistema que conforma una red de colas, este subsistema a su vez es también un sistema de colas, pudiendo ser compuesto por cualquiera de los modelos vistos anteriormente. (Taha, 2012)

Se define grafo orientado como conjunto de nodos enlazados de forma direccional.

Ilustración 4. Representación de una red de colas



Fuente: Elaboración propia

Se puede observar como ejemplo de una red de colas la Ilustración 4. Se trata de una red de colas formada por cuatro nodos, cada uno de ellos con su propio sistema de colas.

Para aclarar este concepto pondremos de ejemplo una asesoría empresarial. En primer lugar los clientes llegan y esperan a ser atendidos por el servicio de recepción, una vez allí serán derivados al servicio solicitado (contable o fiscal por ejemplo), allí esperan la cola correspondiente y una vez atendidos, tendrán que hacer cola en un servicio de gestión de cobros.

La teoría de redes de colas contempla dos modelos. Así se entiende que las redes pueden ser cerradas o abiertas en función de que los clientes entren y salgan o solamente se desplacen de nodo a nodo en su interior.

9.1 Redes de Jackson abiertas

Se llaman redes de Jackson abiertas las redes que contemplan la posibilidad de entrada de clientes desde el exterior. De acuerdo con Cao (2002) las redes abiertas verifican tres propiedades:

1. Cada nodo *i* mantiene un número *s* de servidores de igual tiempo de servicio bajo una distribución exponencial.
2. El tiempo que transcurre entre dos llegadas que provienen del exterior del sistema cumplen una distribución exponencial.
3. Los clientes que son atendidos en un nodo se dirigen a cualquier otro nodo de servicio de forma aleatoria, o bien sale del sistema. Es decir no existen nodos preferidos.

Se puede observar una red de colas abierta de forma práctica en un hospital, en cada servicio que se deriva a un paciente debe realizar cola. Desde que llega un paciente al servicio de urgencias, primero realiza cola para que le tomen los datos, hace cola esperando que le atiendan y posteriormente si lo necesita será atendido, previa cola, en cualquiera de los servicios hospitalarios que necesite.

9.2 Redes de Jackson cerradas

La redes cerradas son sistemas compuestos por K subsistemas o nodos en el cual no entra ni sale ningún cliente. Estando todos los tiempos de atención modelizados bajo una distribución exponencial, el elemento a tener en cuenta es el número de clientes N dentro de la red, el cual permanecerá constante.

En este tipo de redes el estudio relevante son las probabilidades de que haya más o menos clientes en cada subsistema de atención o nodo. (Cao, 2002)

Al igual que el apartado anterior, por la complejidad de las operaciones y no siendo el objetivo del trabajo, se facilitará bibliografía en la cual se detalla la formulación de todos los resultados.

El servicio de las antenas de telefonía es un buen ejemplo, del funcionamiento de una red de colas cerrada en el ámbito de las telecomunicaciones, a medida que nos desplazamos, la señal cambia de servidor de forma automática al nodo que ofrece mayor frecuencia de cobertura.

Estudio empírico

A continuación, se realizará una aplicación práctica del modelo M/M/S, sobre un escenario real, hipermercado Carrefour, comparando sus medidas de eficiencia con un modelo simple M/M/1.

Además, se mostrará el efecto que pequeñas variaciones de los factores generarían en la eficiencia del sistema, tales como el número de cajas abiertas disponibles para cobrar a los clientes.

Para finalizar, y completando el estudio; se cuantificará las medidas de eficiencia en euros, para así, observar desde un punto de vista económico la influencia de los factores. Se enunciará como análisis económico.

10.1 Ámbito de aplicación

Los hipermercados son establecimientos que atienden un gran volumen de clientes en grandes superficies; la capacidad que tengan para gestionarlos será fundamental para su éxito económico y comercial.

Carrefour Ourense (Barbadás), con 9500 m^2 de superficie comercial atiende a 300 clientes, promedio por hora, de 9 am hasta las 22 pm, de lunes a sábado[8]. Volumen suficientemente importante como para que las decisiones al respecto no sean tomadas sin consideración.

Realizar la aplicación práctica sobre esta superficie no ha sido una decisión solamente relacionada con su volumen e importancia para el comercio de la ciudad de Ourense, sino que se decidió, con la venia del tutor, por haber implantado recientemente un sistema de cola única informatizado[9], sustituyendo el sistema tradicional de múltiples colas.

[8] Datos proporcionados por Responsable de área de Ourense.
[9] El sistema indica cual es la caja a la que debes ir para ser atendido.

10.2 Metodología y parámetros básicos

En primer lugar, para estudiar el modelo de colas con el que Carrefour atiende a sus clientes, se sabe que el hipermercado dispone de 29 cajas. Normalmente, según la encargada de cajas, operan con 10 o 12, según la necesidad. Solamente en Rebajas, Navidad y ciertos momentos puntuales se operan con más cajas.

Partiendo de esta información, se observa que de lunes a viernes hasta las 19 pm el hipermercado trabaja con 10 cajas abiertas y solamente pasa a operar con 12 algún día puntual. Por esto asumiremos, para simplificar el modelo, que el sistema utiliza de forma constante 10 cajas.

A continuación y siguiendo el marco teórico, debemos conocer la tasa media de llegadas λ, y la tasa media de servicio μ. Para ello, debemos obtener el tamaño de la muestra. Teniendo en cuenta que los datos que necesitamos estimar son medias, se debe hallar las medias y varianzas muestrales; para ello se selecciona una premuestra que nos indiquen dichos parámetros. (Rodríguez, 1991)

Para obtener los datos de las premuestras, para la tasa de llegada, se cuentan los clientes que entran en franjas de 15 minutos, elegidas estas últimas al azar:

Tabla 1. Premuestra Tasa de llegadas

Premuestra Tasa llegadas (15 minutos)	
66	112
87	63
73	108
69	44
88	60
Media ($\bar{Y}$)	**77**
Varianza (S^2)	**464.6667**

Fuente: Elaboración propia

Una vez recogida la premuestra, se procede al cálculo del tamaño muestral óptimo para operar con un error máximo relativo a la media del 5%, de la siguiente forma[10]:

$$n = \frac{z^2 \ \hat{S}^2}{e^2}$$

[10] Rodríguez, (1991)

Siendo:

n = tamaño muestral

z = Es la desviación del valor medio que aceptamos para lograr el nivel de confianza deseado. En función del nivel de confianza que busquemos, usaremos un valor determinado que viene dado por la forma que tiene la distribución de Gauss.

Para un nivel de confianza 95% se obtiene un z =1,96

e = error, en nuestro caso, siendo para la media, será 5% más menos sobre la media.

$$e = (0.05\ \bar{Y})$$

$\hat{S}^2$= Quasivarianza muestral

Entonces el tamaño muestral para la tasa de llegadas:

$$n = \frac{z^2\ \hat{S}^2}{e^2} = \frac{1.96^2 * \ 464.667}{(0.05 * 77)^2} = 120.429$$

Se debería de recoger 121 datos para trabajar con un error muestral del 5%.

A continuación, en base a un razonamiento análogo se recoge la premuestra para obtener el tamaño muestral de la tasa de servicio, se miden el tiempo que el cajero/a tarda en atender a diversos clientes y en cajas al azar, obteniendo:

Tabla 2. Premuestra Tasa de Servicio

Premuestra Tasa de servicio	
02:57	01:06
02:04	02:51
01:07	01:15
00:59	01:24
01:40	02:02
Media ($\bar{Y}$)	**01:44**
Varianza (S^2)	**0,5135**

Fuente: Elaboración propia

Se calcula el tamaño muestral:

$$n = \frac{z^2\, S^2}{e^2} = \frac{1.96^2 * \ 0.5135}{(0.05 * 1.733)^2} = 262.73$$

Se debería recoger 263 datos para poder trabajar con un error muestral del 5%.

Debido a los resultados obtenidos sobre los tamaños muestrales necesarios, por falta de tiempo, y conocido el procedimiento que hubiera sido necesario para realizar el análisis con el error indicado, se decide asumir un error más elevado y operar con muestras más reducidas.

Se seleccionan franjas horarias aleatoriamente y de manera independiente durante todos los días de la semana, para así contar, con ayuda del arco de seguridad, las unidades potenciales de compra[11] que entran, se debe indicar que varias personas pueden realizar una única compra, se observa su comportamiento a la entrada, para conocer si las compras serán independientes o conjuntas. El objetivo es conocer cuántos clientes entran en promedio al hipermercado.

El procedimiento que se ha utilizado para obtener los datos ha sido el Proceso de contar de Poisson, como se explica en el marco teórico se cuenta cuantos sucesos ocurren en una determinada franja de tiempo, en nuestro caso cuantos clientes entran en 15 minutos.

Los datos obtenidos son los siguientes:

Tabla 3. Muestra tasa de llegadas

Nº llegadas en franjas de 15min	
29	67
44	142
79	42
88	126
60	61
86	74
87	54
33	139
112	78

Fuente: Elaboración propia

[11] Persona o grupo de personas que acceden para realizar una compra única.

Así se puede obtener, el promedio de clientes que entran en el hipermercado, partiendo del promedio de clientes que llegan en 15 minutos, obtenemos la tasa de llegadas por minuto:

$$\lambda_{15\min} = 77.8333\ clientes$$

$$\lambda_{min} = 5.1889\ clientes$$

Es decir, en esta superficie comercial entran, en promedio, más de 5 personas por minuto.

Por la relación existente entre la distribución de Poisson y la distribución Exponencial, obtenida λ, como tasa de llegada podemos obtener el tiempo promedio que pasa entre cliente y cliente como:

$$\frac{1}{\lambda} = \frac{1}{5.1889} = 0.1927\ minutos \sim 11.6\ segundos$$

Por lo tanto, es lo mismo decir que entran 5.1889 clientes por minuto que entra un cliente cada 11.6 segundos.

El siguiente paso, una vez obtenida la tasa de llegadas es conocer la tasa de atención. Para obtener este dato, se mide el tiempo que tarda el operario (cajero/a) en cobrar a cada cliente.

Con cronómetro en mano, con la autorización de la encargada del sistema de cajas, se realizan múltiples mediciones para obtener el promedio de tiempo que tarda un cliente desde que el sistema le asigna una caja hasta que termina de ser atendido.

Los datos obtenidos son los siguientes:

Tabla 4. Muestra tasa de servicio

Tiempo de Atención 1/μ			
01:39	03:16	02:52	00:53
01:27	02:50	04:16	00:51
01:09	02:01	02:30	01:36
01:41	02:58	03:05	01:50
01:13	01:06	01:31	02:02
02:12	00:47	01:12	02:21
00:41	01:02	01:56	01:46
00:24	00:59	03:02	01:20
02:06	01:13	00:37	01:01
00:54	02:05	00:19	02:15
00:50	01:27	01:16	01:41

01:29	01:44	01:08	01:25
01:12	02:24	01:09	02:45

Fuente: Elaboración propia

Se puede obtienen así, un promedio de tiempo de atención, por el cual podemos calcular la tasa de servicio en base a la relación explicada con anterioridad.

$$\frac{1}{\mu} = 1{,}6821\ minutos$$

$$\mu = \frac{1}{1{,}6821} = 0{,}5945\ clientes$$

Es decir, cada caja tarda un promedio de 1.6821 minutos, 1 minuto y 18 segundos en servir a cada cliente, esta información sirve para obtener la tasa de servicio $\mu = 0.5945$, cada minuto se atienden a 0.5945 clientes.

Una vez obtenido el tamaño muestral, se realizará el cálculo del error muestral en base al valor de la distribución t de Student con n-1 grados de libertad (Rodríguez, 1991)

$$e^2 = \frac{t_{n-1}^2\ \hat{S}^2}{n}$$

Para un nivel de confianza del 95% es t_{18-1}= t_{17}= 2.1098

Para la tasa de llegadas con n = 18 y $\hat{S}^2$= 1160.3823 se obtiene:

$$18 = \frac{2.1098^2 * 1160.3823}{e^2} =$$

$$e^2 = \frac{5165.1587}{18} = 286.953$$

$$e = \sqrt{286.953} = 16.939\ peronas\ en\ 15\ minutos$$

$$e = 1.1293 personas/minuto$$

Obtenemos el error real de 1.1293 personas/minuto, sobre una media de llegadas de 5.1889 clientes, lo que significa en tanto por ciento un errores próximos al 21.76%.

Tabla 5. Ficha Técnica. Análisis de la tasa de llegadas

Ficha técnica	
Objetivo	Conocer tasa promedio de llegadas
Universo de estudio	Franjas de 15 minutos durante todo el periodo de apertura de Carrefour (Barbadás)
Metodología	Muestreo aleatorio simple
Tamaño muestral	18 mediciones de 15 minutos
Error	21.76 %
Período	Mayo de 2016

Fuente: Elaboración propia

Con un razonamiento análogo se calcula el error real para la tasa de servicio

Para un nivel de confianza del 95% es $t_{52-1}= t_{51}= 2.0076$

Y n = 52 y $\hat{S}^2$ = 0.6763 se obtiene:

$$52 = \frac{2.0076^2 * 0.6763}{e^2} =$$

$$e^2 = \frac{2.72579}{52} = 0.05241$$

$$e = \sqrt{0.05241} = 0.2289\ minutos$$

Obtenemos un error real sobre la media de 0.2289 minutos, sobre una media de servicio de 1.6821 minutos, lo que significa en tanto por ciento un error próximo al 13.61%.

Tabla 6. Ficha Técnica. Análisis de la tasa de servicio

Ficha técnica	
Objetivo	Conocer tasa promedio de servicio
Universo de estudio	Unidades potenciales de compra (Barbadás)
Metodología	Muestreo aleatorio simple
Tamaño muestral	52 mediciones de tiempo de servicio
Error	13,61 %
Período	Junio de 2016

Fuente: Elaboración propia

10.3 Medidas de eficiencia para el sistema M/M/10

Una vez obtenidos los parámetros principales λ y μ, se marcan las hipótesis de partida para obtener las medidas de eficiencia del modelo:

- Los tiempos entre llegadas de los clientes y de atención de las cajas obedecen a una distribución exponencial
- Se opera con 10 cajas
- La capacidad de cola se considera infinita
- La población potencial se considera infinita
- La disciplina de cola es FIFO

Con respecto a la primera de las hipótesis, se debe indicar que se necesitaría un contrataste no paramétrico para justificar que las tasas obedecen a una distribución exponencial, sin embargo no se realiza. Por un lado, porque no es objetivo del trabajo la realización de este tipo de contrastes y por otro, al disponer de pocos datos el resultado no sería concluyente. Se asume entonces que las hipótesis de partida son ciertas para poder realizar el análisis.

Por lo tanto el modelo se define: **M/M/10/∞/∞/FIFO**

Parámetros principales:

Tabla 7. Parámetros principales

λ	μ
5,1889	0,5945

Fuente: Elaboración propia

Obtenemos:

1. Factor de utilización de las cajas

$$\rho = \frac{\lambda}{s\mu} = \frac{5.1889}{10 * 0.5945} = 87.28\ \%$$

Cada caja está ocupada, en promedio, entorno al 87% del tiempo, o bien está sin atender a ningún cliente aproximadamente un 13% del tiempo.

2. Probabilidad de que no haya nadie en el hipermercado

Siguiendo el cálculo por recurrencia

$$c_0 = 1$$

$$c_1 = \frac{1 * 5.1889}{1 * 0.5945} = 8.728$$

$$c_2 = \frac{c_1 * 5.1889}{2 * 0.5945} = 38.089$$

$$\{\dots\}$$

$$c_9 = \frac{c_8 * 5.1889}{9 * 0.5945} = 810.1174$$

$$\sum_{c_0}^{c_9} c_i = 3847.7501$$

Entonces:

$$p_0 = \frac{1}{3847.7501 + \frac{5.1889^{10}}{9! * \ 0.5945^{(10-1)} * (10 * 0.5945 - 5.1889)}} = 0.01063\%$$

La probabilidad de que no se encuentre nadie dentro del sistema, esperando en cola o bien pagando, es ínfima.

3. Número promedio de clientes en cola

$$L_q = \frac{\lambda^{s+1} p_0}{(s-1)!\ \mu^s (s\mu - \lambda)} =$$

$$= \frac{5.1889^{10} * 0.001063}{(10-1)! * 0.5945^{10}(10 * 0.5945 - 5.1889)} = 4.05565\, personas$$

Una vez realizada la compra, y dispuestos a hacer cola, el cliente se encontrará un promedio de ocho clientes esperando antes que él.

4. Tiempo promedio de espera en la cola

$$W_q = \frac{L_q}{\lambda} = \frac{4.05565}{5.1889} = 0.7816\, minutos$$

Una vez en la cola, el cliente debe esperar de media 0.78 minutos, aproximadamente 47 segundos, para ser atendido por el cajero/a.

5. Tiempo promedio total, (tiempo de cola más tiempo de atención)

$$W_s = W_q + \frac{1}{\mu} = 0.7816 + \frac{1}{0.5945} = 2.4636\, minutos$$

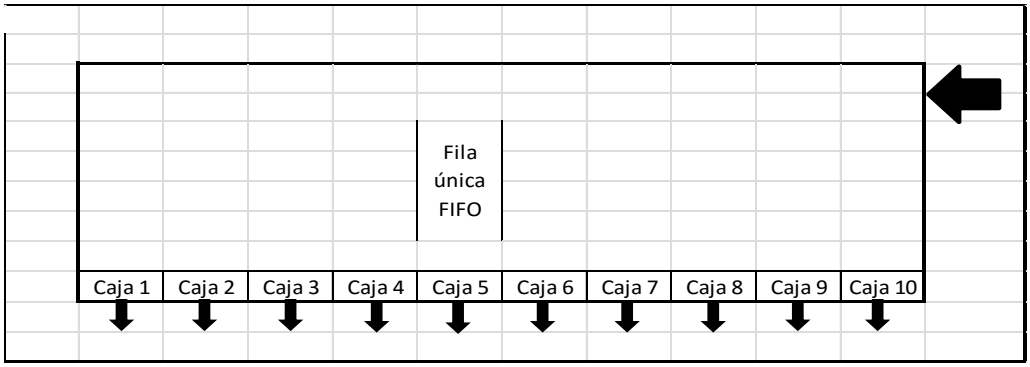
El cliente, teniendo en cuenta el tiempo que espera para ser atendido en promedio y el tiempo que tarda el cajero/a en cobrarle, se estiman 2.46 minutos, aproximadamente 2 minutos y 27 segundos.

6. Número promedio de unidades en el sistema.

$$L_s = \lambda\, W_s = 5.1889 * 2.4636 = 12.7838\, personas$$

Con esta última medida, se obtiene el promedio de clientes que se encuentran en el sistema, es decir, cuantas personas están en la cola o bien siendo atendidos en las cajas

Ilustración 5. Representación de un sistema M/M/10



Fuente: Elaboración propia

Tabla 8. Resumen de los resultados del modelo

		Promedio
λ	Tasa de llegada (personas)	5.1889
μ	Tasa de servicio (personas)	0.5945
ρ	Factor de utilización	87.28%
p_0	Probabilidad de sistema vacío	0,01063%
L_q	Personas en cola	4.0556
W_q	Tiempo en cola (minutos)	0.7816
W_s	Tiempo en el sistema (minutos)	2.4636
L_s	Personas en el sistema	12.7838

Fuente: Elaboración propia

10.4 Variaciones de eficiencia sobre el modelo M/M/S

A continuación, se plantea el escenario en el cuál Carrefour decide operar con nueve cajas u once en lugar de con diez, con el objetivo de comprobar si la decisión de trabajar con diez cajas es acertada. Para realizar la comparativa asumimos constantes las tasas promedio de llegada y atención: 5.1889 y 0.5945 respectivamente.

En primer lugar, y para analizar las variaciones posibles en la eficiencia del sistema, atenderemos al factor de utilización de las cajas "ρ". Como se puede observar en la Tabla 7, operando con diez cajas, se mantienen ocupadas un 87.28% del tiempo, como es lógico, a tasas constantes, el factor de utilización es superior, 96.9%, si operase con una caja menos y viceversa si atendiesen con una caja más. Este es una medida expresamente de interés para el gestor del hipermercado, se concluirá en el análisis económico sobre su uso y aplicación.

Tabla 9. Variaciones de eficiencia

	9 cajas	10 cajas	11 cajas
ρ	96.9%	87.28%	79.3%
p_0	0.003%	0,01063%	0.014%
L_q	27.79	4.0556	1.42
W_q	5.35	0.7816	0.27
W_s	7.03	2.4636	1.95
L_s	36.51	12.7838	10.14

Fuente: Elaboración propia

En segundo lugar, como se muestra en la Tabla 9, el número promedio de clientes en cola "L_q" aumenta ampliamente con solo reducir una caja, sin embargo al operar con una caja más la reducción no es tan significante. Los clientes que esperan en cola a ser atendidos, en promedio, y por lo tanto el tiempo que se debe esperar, varía sustancialmente si se operase con una caja más o menos. En cuanto al tiempo, está directamente relacionado, se pasa de menos de un minuto de espera a más de cinco.

En tercer lugar, atendiendo al número total de personas en el sistema (clientes en cola y siendo atendidos) "L_s" y el tiempo relacionado "W_s" también se comprueba la misma relación que en las medidas de eficiencia anteriores. De permanecer en el sistema dos minutos y medio se pasaría a más de siete si se redujese una caja.

En cuanto a la probabilidad de que el sistema está vació, que no se encuentre siendo atendido ni por supuesto en cola, se puede observar que sus tasas son insignificantes.

Como conclusión a este apartado, destaca: la decisión de Carrefour de operar con diez cajas no ha sido tomada sin estudio. Si se operara con once cajas los clientes esperarían menos, sería una atención casi instantánea, eso sí, Carrefour tendría que soportar un factor de utilización inferior al 80%, o lo que es lo mismo; sus cajas estarían ociosas un 20% del tiempo, lo que sería antieconómico.

10.5 Comparativa de eficiencia: M/M/10 y diez sistemas M/M/1

Llegados a este punto, nos planteamos sí es más eficiente operar con un sistema de cola única o un sistema tradicional simple de múltiples colas. Si es así, ¿cuál es la diferencia de eficiencia real entre los dos sistemas de colas?

Obtenidos anteriormente los resultados del sistema de cola única de Carrefour, siguiendo el marco teórico, se hallarán las medidas de eficiencia para 10 sistemas M/M/1 con los mismos parámetros que se han utilizado en los apartados anteriores.

Para calcular los parámetros que representen un modelo basado en 10 cajas individuales con sus respectivas colas, se suponen que los clientes pueden acudir a cualquiera de las cajas sin preferencias. Por lo tanto se debe dividir uniformemente la tasa de llegadas entre las diez cajas.

A continuación, siguiendo las tasas de llegada y servicio obtenidas en el trabajo de campo en Carrefour, se calcula las medidas de eficiencia para un sistema M/M/1 en función de lo detallado en el marco teórico.

Se compara con los resultados anteriormente obtenidos para el sistema de una sola cola.

- Tasa de llegada: λ= 0.51889
- Tasa de servicio: μ= 0.5945

Tabla 10. Comparativa de eficiencia

	10 - M/M/1	M/M/10
ρ	87.28%	87.28%
p_0	12.72%	0,01063%
L_q	5.9898	4.0556
W_q	11.54	0.7816
W_s	13.225	2.4636
L_s	6.8627	12.7838

Fuente: Elaboración propia

Como se muestra en la Tabla 10, la comparativa de eficiencia no deja en buen lugar al sistema tradicional de múltiples colas frente al sistema de cola única. A pesar de que el factor de utilización de las cajas es el mismo, los cajeros/as trabajan con la misma eficiencia, las diferencias se manifiestan de cara a la espera del consumidor

En primer lugar, atendiendo al tiempo de espera para ser atendido, W_q, el sistema tradicional genera, en promedio, tiempos de espera aproximadamente entorno a doce minutos por caja; tiempos muy alejados de los promedios alcanzados por el sistema de una sola cola, 47 segundos para ser atendido. He aquí la principal diferencia entre estos dos sistemas, ser servido en la primera caja que esté disponible evita esperas innecesarias. El bloqueo de una de las cajas, en el sistema tradicional, aumenta sustancialmente el tiempo de espera, sin embargo en el modelo M/M/10, los clientes se reparten de forma eficiente según las cajas queden disponibles.

En relación al número de personas que esperan en la cola, L_q, también es sustancialmente superior en el modelo tradicional. Como se muestra en la Tabla 8, el número de clientes que esperan en la cola tradicional, en promedio son próximos a 6 por caja, sin embargo, en el sistema M/M/10 la media se sitúa cerca de 4, pero recordamos que la cola es única. Es decir en el sistema tradicional esperaran en todas las colas

individuales 6 personas por cola, un promedio de 60 personas, tasas muy superiores que en modelo de una sola cola.

Se debe mencionar: en la tabla 8 y con respecto a la cantidad de clientes en el sistema, L_s, aparecen un promedio de 6.86 para el sistema tradicional y de 12.78 para el M/M/10, pero estos datos no son comparables en efecto inmediato, ya que para el sistema tradicional sería 6.86 por caja, es decir 68 frente a 12.78.

A la vista de los resultados obtenidos, el sistema tradicional, a pesar de ofrecer un factor de utilización igual que el sistema de una sola cola, permite económicamente dar el mismo aprovechamiento a los factores productivos: caja, cajero/a; es ineficiente de cara al consumidor.

A parte de la velocidad del sistema, que ha sido demostrada, la percepción que se genera en un sistema, por parte del consumidor, cuando necesitas ser atendido y hay 60 personas, repartidas en las diferentes colas; es totalmente diferente si hay 4 o 5 en una cola única.

10.6 Análisis económico

Los factores que inciden en las decisiones operativas de un hipermercado son múltiples, desde la desgranación de los costes, pasando por la propia eficiencia operativa del sistema de colas o incluso los diferentes perfiles de consumidores a los que venden.

En este apartado se podría desgranar todos estos aspectos, pero se sale del objeto del trabajo, no obstante será interesante mostrar como la disciplina de teoría de colas se muestra útil para decidir sobre múltiples aspectos de contratación.

Partiendo de un indicador básico de referencia, como es el salario base bruto de los cajeros/as, se realizará una comparativa de distintos escenarios en los cuales se podrán comparar las posibilidades de contratación en base la eficiencia del modelo M/M/10.

Para realizar este análisis, en primer lugar se consulta lo resuelto en el BOE el 8 de Abril de 2013, por la Dirección General de Empleo, por la que se registra y publica el Convenio Colectivo estatal de grandes almacenes, válido hasta el 31/12/2016, por el que se rige Carrefour.

Se parte de la hipótesis que la política de la empresa es la de mantener las 10 cajas operativas, lo que se plantea estudiar es las posibilidad de operar con cajeros de más o menos experiencia, en base a la suposición de que los cajeros/as con experiencia son más rápidos y sin ella más lentos a la hora de atender.

Previamente se indica que las medidas de eficiencia con las que hasta ahora se ha trabajado, fueron obtenidas de mediciones a cajeros/as con experiencia inferior a cuatro años, factor relevante a partir de este punto.

Siguiendo el convenio colectivo mencionado, se indica lo siguiente:

- Salario Base Bruto (anual) Cajero sin experiencia: 13.951´11€
- Salario Base Bruto (anual) Cajero con experiencia: 14.369´65€
- Complemento Cuatrienal (anual): 255´79€

A continuación, se recuerda que el tiempo promedio obtenido de las mediciones había sido de 1.6821 minutos, con lo que se obtenía una tasa de atención promedio de 0.5945, como se explica anteriormente, un cajero/a atiende, en promedio, a 0.5945 personas por minuto.

Se plantea los siguientes escenarios:

Escenario A (real)

Cajeros/as con experiencia inferior a cuatro años, medidas de eficiencia obtenidas en apartado 10.3. De acuerdo con el Convenio Colectivo estos operarios perciben un salario bruto anual de 14.369´65€, 42´76€ día o bien 5´34€/hora.

Se puede decir, entonces, que dicho salario sirve para financiar las medidas de eficiencia que se han obtenido. Desde el punto de vista del gestor de Carrefour, el factor de utilización de las cajas, 87.28%, y como se analiza pormenorizadamente en apartados anteriores los tiempos de espera y el número de personas en cola son asumibles para el consumidor (tiempo promedio total en el sistema 2.4636 minutos).

Desde el punto de vista económico, teniendo un índice de cajas ociosas del 12.72%, el coste objetivo de este desaprovechamiento sería en base al salario convenido de 1827.82€ mensuales.

Sobre este escenario: manifiesta la realidad del hipermercado, llegados a su cola, encontraremos 4 o 5 personas esperando, y en menos de un minuto seremos llamados a una caja para ser atendidos. El cajero/a que nos atiende tendrá menos de cuatro años de experiencia y el tiempo total que se estará en cola y siendo atendidos será de dos minutos y medio aproximadamente.

Las decisiones que se llegan a resolver para que este escenario se dé, se justifican por si solas, en base a la suposición de que la tasa de llegadas es constante, la velocidad de atención de este tipo de cajero/a es el idóneo, como manifiestan los indicadores. ¿Podría ser más eficiente?, a continuación se plantea un escenario que lo analice.

Escenario B (rápido)

Cajeros/as con experiencia superior a cuatro años, suponemos más rápido, conforme el Convenio Colectivo estos operarios perciben un salario bruto anual de 14.369´65€ más un complemento anual por cuatrienio de 255´79€, haciendo un total de 14.625´44€, 43´52€/día o bien 5´44€/hora.

Este escenario se plantea que este cajero/a pasa de la tasa de atención de 0.5945 personas por minuto a un 0.6539, 10% más rápido. Así asumiendo la tasa de llegada constante 5.1889 personas/minuto, se calcula las medidas de eficiencia:

Tabla 11. Medidas eficiencia para Escenario B

	promedio
p	79,35%
Po	0,030%
Lq	1,518
Wq	0,292
Ws	1,822
Ls	9,453

Fuente: Elaboración propia

En primer lugar, en mención a la Tabla 11, el tiempo de espera disminuye sustancialmente a la par que las personas que esperan en la cola. Ahora bien, el incremento de la tasa de servicio, a una misma tasa de llegada, provoca una bajada relevante del factor de utilización, desde el 87.28% al 79.35%.

Se estaría financiando con salarios más altos un 20% de cajas ociosas, desde el punto de vista económico el 20% de los salarios no serían productivos, teniendo en cuenta el convenio, 2925.08€ por trabajador y mes. Un nada despreciable 60% más que en el caso anterior.

Visto el resultado, parece justificar la política de formar cajeros de manera que se ajusten su tasa de atención promedio a los 0.60 clientes/minuto, que ofrece unas medidas de eficiencia estables y un factor de utilización aceptable para la tasa de llegada obtenida.

Siendo el escenario real el más rentable, hasta ahora, se plantea un tercero, ¿sería aceptable cajeros sin experiencia por un salario menor?

Escenario C (lento)

Partiendo de la información obtenida del Convenio Colectivo, al igual que en los demás escenarios, indicamos: Cajero sin experiencia 13.951´11€ anuales, 41´52 €/día o bien 5.19 €/hora.

Este escenario se plantea que este cajero/a pasa de la tasa de servicio de 0.5945 personas por minuto a 0.535 personas/minuto, 10% menos, se calcula las medidas de eficiencia:

A primera vista, la variación de la tasa promedio de servicio en unos segundos no parece que fuese muy relevante; en este caso cajeros/as sin experiencia se ponen al frente de las cajas de nuestro modelo de colas.

Ahorramos un 3% sobre los costes del salario de los cajeros/as, pero atendiendo a los resultados expuestos en la Tabla 12: el factor de utilización llega a índices próximos al 100%, lo que a priori sería ideal, no lo es. Siendo próximo a 100% el factor de utilización de las cajas expone al sistema a un alto riesgo de colapso.

Tabla 12. Medidas de eficiencia para Escenario C

	promedio
p	97%
Po	0,0013%
Lq	28.759
Wq	5.542
Ws	7.412
Ls	38.458

Fuente: Elaboración propia

En segundo lugar y teniendo en cuenta los indicadores que muestran el promedio de clientes en cola, se puede concluir que el ahorro no es económicamente productivo, siendo el tiempo de espera en cola 5.54 minutos, y el número de personas en la cola próximo a 30, supondría un coste en cuestión de calidad percibida importante.

Partiendo de los tres escenarios detallados, se observa que Carrefour opera justificadamente con cajeros/as jóvenes con una experiencia mínima que permita obtener la tasa de atención promedia muy próxima a nuestros 0.5945 clientes/minuto, sirviendo a cientos de clientes al día de forma técnica y económicamente eficiente.

Para concluir, si fuésemos capaces de cuantificar el coste que tiene para la empresa el tiempo de espera de un cliente, podríamos hacer un estudio de contratación basado en el tiempo que está esperando un cliente antes de ser servido. Comparando así, por ejemplo, si el aumento que se produce en el salario, al contratar empleados con más experiencia, es inferior al beneficio obtenido/percibido por la disminución del tiempo de espera para el cliente

10. Conclusiones

Llegados a este punto, se debe echar la vista atrás y analizar si se han cumplido los objetivos que se habían planteado.

En primer lugar, se ha justificado la realización de este trabajo, en base a la importancia y necesidad de las decisiones operativas que permite tomar la Teoría de Colas. Para ello se ha analizado en profundidad su base teórica, apoyándonos en numerosos ejemplos que ayudan a visualizar sus pormenores.

En segundo lugar, con respecto al estudio empírico, se han aplicado los modelos M/M/1 y M/M/S en un caso real, obteniendo resultados prácticos y útiles. Además de poder compararlos entre sí, se ha comprobado la eficiencia del modelo aplicado por Carrefour para la gestión de sus clientes.

Con respecto al estudio empírico, en primer lugar se han aplicado los modelos que se habían marcado como objetivo en el escenario real, sin embargo, por falta de tiempo no se ha podido trabajar con muestras suficientemente grandes para garantizar un erro máximo más pequeño. No obstante se considera cumplido el objetivo ya que los resultados obtenidos, a pesar de no ser estadísticamente significativos, representan una realidad posible en la práctica.

Siguiendo con la parte práctica, se debe concluir que Carrefour de un tiempo a esta parte ha cambiado su modelo de atención; de un modelo tradicional a un modelo de cola única, teniendo en cuenta numeroso aspectos, enfocados principalmente en el consumidor. La eficacia que logran en la gestión de cientos de clientes al día pasa por un eficiente sistema de una sola cola. En base a los resultados se justifica que el sistema de una sola cola es más eficiente que el de múltiples colas independientes.

Atendiendo al análisis económico, se ha demostrado que no siempre un grado más de especialización es mejor. Pequeñas variaciones en la tasa de servicio provocan grandes distorsiones en el sistema, por lo que su ajuste es el factor más determinante en la eficiencia de un sistema de colas.

Finalmente, con respecto al aprendizaje por mi parte, puedo concluir que durante el desarrollo de este trabajo, he ampliado mis conocimientos sobre Teoría de Colas de forma práctica y teórica, aportándome nuevas competencias que podré aplicar en un futuro próximo.

Bibliografía

Cao, R. (2002). *Introducción a la Simulación y a la Teoría de Colas.* A Coruña: Netbiblo.

Carro, R. (2009). *Investigación de Operaciones en Administración.* Mar de Plata: Pincu

De la Fuente, G.D., Pino, D.R. (2001). *Teoría de líneas de espera: modelos de colas.* Gijón: Universidad de Oviedo, Servicio de Publicaciones.

Krajewski, L.J., Riztman, L.P. (2000). *Administración de operaciones: estrategia y análisis.* México: Pearson

Montero, JM (2007). *Estadística descriptiva.* Madrid: Paraninfo

Render, B., Heizer, J. (2007). *Dirección de la producción y de operaciones: decisiones tácticas 8ª edición.* Madrid: Pearson

Rodríguez, O. J. (1991). *Métodos de muestreo.* Madrid: Centro de investigaciones sociológicas

Schroeder, R. (1999). *Administración de Operaciones*. 3ª Edición. España: Mc Graw-Hill

Taha, H.A. (2012). *Investigación de Operaciones*. México: Pearson

Winston, W.L. (2004). *Investigación de Operaciones Cuarta edición.* México: Thomson

Bibliografía adicional

Blanca Azucena C. (2005). *La teoría de colas como herramienta para optimizar el servicio de una entidad municipal.* (Tesis inédita de maestría). Universidad San Carlos de Guatemala.

Vega, J.L. (2004). *Diseño e implementación de una herramienta para la enseñanza y el aprendizaje de la teoría de colas.* Proyecto de fin de carrera. Universidad da A Coruña.

Willing, A. (1999). *A Short Introduction to Queueing Theory.* Berlín: Techical Univesity Berlin.

Zukerman, M. (2015). *Introduction to Queueing Theory and Stochastic Teletraffic Models.* Hong Kong: City University of Hong Kong