

**UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL
CUSCO**

**ESCUELA DE POSGRADO
MAESTRIA EN ESTADÍSTICA**



**“SIMULACIÓN DE SISTEMAS DINÁMICOS PARA LÍNEAS DE ESPERA,
APLICADA AL TRÁFICO EN LAS VENTANILLAS DE UN BANCO”**

TESIS PRESENTADO POR:

Br. JOEL GRIMALDO OLARTE ESTRADA

**PARA OPTAR EL GRADO ACADÉMICO DE
MAESTRO EN ESTADÍSTICA**

ASESOR:

Mgt. GUILLERMO PAUCAR CARLOS

CUSCO – PERÚ

2019

DEDICATORIA

A la Memoria de mi Madre Margarita

Quien fue pilar fundamental en mi formación como profesional

A mi Esposa Sofía

Por brindarme siempre su apoyo y comprensión en los momentos que sentía decaer, por sus palabras de aliento y motivación que me enseñaron valores para alcanzar grandes metas, por su cariño, amor y sobre todo darme la dicha de ser doblemente Padre

A mi Princesa Anghely

Porque desde que naciste me llenas de alegría cada día, tu amor, tus besos y abrazos son grandes manifestaciones de afecto que Dios puso en mi camino, con tu sonrisa me das la fortaleza que necesito para conseguir mis objetivos.

A mi Bebé Joel Augusto

Porque completaste mi felicidad y formas parte de nuestro hogar, te llevaré por siempre en mi mente y en el corazón, de seguro que seguirás mis pasos en este maravilloso mundo de la Ciencia Matemática

AGRADECIMIENTO

Agradezco a la Universidad Nacional de San Antonio Abad del Cusco, Institución donde me he formado profesionalmente y realizado los estudios de maestría, que me dio la posibilidad de laborar y transmitir lo aprendido a los estudiantes.

Agradecimiento especial a mi Asesor Mgt. Guillermo Paucar Carlos por el constante apoyo para el desarrollo de la Tesis.

Agradecimiento también especial al Dr. Alfredo Valencia Toledo por las enseñanzas de la programación en R, el tiempo empleado en largas horas de labor, haber compartido sus conocimientos y sobre todo por la amistad brindada.

A los profesionales Ing. Manuel Arizabal Vera, Blgo. Pavel Atauchí Rojas por el apoyo brindado en la culminación de la Tesis.

A mi Padre Grimaldo y mis hermanos Percy y Magaly por su constante apoyo moral.

A todos los Profesores, colegas y amigos que compartieron sus conocimientos y fortalecieron mis saberes.

ÍNDICE GENERAL

Presentación	i
Resumen	ii
Abstract	iii
INTRODUCCIÓN	1
<i>I. PLANTEAMIENTO DEL PROBLEMA</i>	2
1.1. <i>Situación problemática</i>	2
1.2. <i>Formulación del problema</i>	3
1.3. <i>Justificación de la investigación</i>	4
1.4. <i>Objetivos de la investigación</i>	5
<i>II. MARCO TEÓRICO</i>	6
2.1. <i>Bases teóricas</i>	6
2.1.1. <i>Introducción a la teoría de colas</i>	6
2.1.1.1. Definiciones iniciales	6
2.1.1.2. Origen	8
2.1.1.3. Modelo de formación de colas	8
2.1.1.4. Objetivos de la Teoría de Colas	13
2.1.1.5. Elementos de la Teoría de Colas	13
2.1.1.6. Descripción de un Sistema de Colas	16
2.1.1.7. Notación de Kendall	22
2.1.1.8. Cola M/M/1	25
2.1.1.9. Un poco más de notación	25
2.1.1.10. Fórmulas de Little	29
2.1.1.11. Un sistema D/D/1	34
2.1.1.12. Herramientas Probabilísticas: Procesos Estocásticos	37
2.1.1.13. Procesos de Poisson	37
2.1.1.14. Relación entre los procesos de Poisson y la distribución exponencial	38
2.1.1.15. Modelos de colas con tasas de llegada y servicio de Poisson	40
2.1.1.16. El modelo M/M/1	41
2.1.1.17. El modelo M/M/s	45
2.1.1.18. Series de tiempo	48
2.1.1.19. Procesos Estacionarios	49
2.1.1.20. Modelos estocásticos de Series temporales	50
2.1.1.21. Sistemas, modelos y simulación.	51

2.1.1.22.	Funciones del modelo	53
2.1.1.23.	Ventajas y consideraciones de la simulación	55
2.1.1.24.	Elementos de un Modelo de Simulación	56
2.1.1.25.	Modelización de las Entradas	57
2.1.1.26.	Análisis de Resultados.....	58
2.1.1.27.	Validación del Modelo	59
2.1.1.28.	Programación de la simulación.....	60
2.1.1.29.	Simulando con PROMODEL	62
2.2.	<i>Marco Conceptual.....</i>	65
2.3.	<i>Antecedentes.....</i>	66
III.	<i>HIPOTESIS.....</i>	70
3.1.	<i>Hipótesis.....</i>	70
3.2.	<i>Identificación de variables</i>	70
3.3.	<i>Operacionalización de variables.....</i>	71
	<i>CAPÍTULO IV</i>	72
IV.	<i>METODOLOGÍA.....</i>	72
4.1.	<i>Ámbito de estudio.....</i>	72
4.2.	<i>Tipo y Nivel de Investigación</i>	72
4.3.	<i>Unidad de Análisis</i>	72
4.4.	<i>Población de estudio.....</i>	72
4.5.	<i>Tamaño de muestra</i>	73
4.6.	<i>Técnicas de recolección de muestra</i>	73
4.7.	<i>Técnicas de recolección de información</i>	73
4.8.	<i>Técnicas de análisis e interpretación de la información.....</i>	75
4.9.	<i>Técnicas para demostrar la verdad o falsedad de hipótesis planteadas.....</i>	75
V.	<i>RESULTADOS</i>	76
5.1.	<i>Procesamiento, análisis, interpretación y discusión de resultados.....</i>	76
5.2	<i>Presentación de resultados</i>	95

Índice de tablas

TABLA 1 POSIBLES VALORES SEGÚN LA NOTACIÓN A / B / X / Y / H / Z.....	24
TABLA 2 RESUMEN DE RESULTADOS PARA SISTEMAS G/G/s.	33
TABLA 3 ANÁLISIS DESCRIPTIVOS DE DATOS.....	76
TABLA 4. RESULTADOS DE LAS COMPARACIONES PARA LOS TIEMPOS DE LLEGADA.....	79
TABLA 5. RESUMEN DE LAS COMPARACIONES PARA LOS TIEMPOS DE SERVICIO	81
TABLA 6 NÚMERO PROMEDIO DE CLIENTES QUE LLEGAN AL BANCO EN INTERVALOS DE 5 MINUTOS.....	83
TABLA 7 RESUMEN DESCRIPTIVO DE LAS SERIES EN MINUTOS	96
TABLA 8 RESUMEN DE RESULTADOS G1 (UN SERVIDOR)	98
TABLA 9 RESUMEN DE RESULTADOS G1 (DOS SERVIDORES)	99
TABLA 10 RESUMEN DE RESULTADOS G2 (UN SERVIDOR)	100
TABLA 11 RESUMEN DE RESULTADOS G2 (DOS SERVIDORES)	100

Índice de figuras

FIGURA 1 ELEMENTOS PRINCIPALES DE UN SISTEMA DE COLAS	12
FIGURA 2 CLASIFICACIÓN DEL MODELO DE COLAS	12
FIGURA 3 ILUSTRACIÓN DE LAS FÓRMULAS DE LITTLE.....	30
FIGURA 4 TIEMPOS DE ESPERA ENTRE CLIENTES CONSECUTIVOS EN UN SISTEMA G/G/1 ...	34
FIGURA 5 EVOLUCIÓN DE UN SISTEMA D/D/1 CON $\lambda = 1/3$ Y $\mu = 1/5$	36
FIGURA 6 DIAGRAMA DE CAJAS EN INTERVALOS DE 5 MINUTOS DE LOS	84
FIGURA 7 DISPERSIÓN DE LA SERIE DE DATOS PARA EL DÍA LUNES	85
FIGURA 8 DISPERSIÓN DE LA SERIE DE DATOS PARA EL DÍA MARTES	86
FIGURA 9 DISPERSIÓN DE LA SERIE DE DATOS PARA EL DÍA MIÉRCOLES	86
FIGURA 10 DISPERSIÓN DE LA SERIE DE DATOS PARA EL DÍA JUEVES.....	87
FIGURA 11 DISPERSIÓN DE LA SERIE DE DATOS PARA EL DÍA VIERNES	88
FIGURA 12 FUNCIÓN DE DENSIDAD PARA LA SERIE TLL1.....	89
FIGURA 13 COMPARACIÓN DE FUNCIÓN DE DISTRIBUCIÓN ACUMULADA.....	89
FIGURA 14 PRUEBA DE K-S PARA EL DÍA MARTES.....	90
FIGURA 15 PRUEBA DE K-S PARA EL DÍA MIÉRCOLES.....	91
FIGURA 16 PRUEBA DE K-S PARA EL DÍA JUEVES	91
FIGURA 17 PRUEBA DE KOMOGOROV PARA EL DÍA VIERNES.....	91
FIGURA 18 HISTOGRAMA DE FRECUENCIAS PARA LAS SERIES DEL TIEMPO DE SERVICIO.....	92
FIGURA 19 COMPARACIÓN DE FUNCIÓN DE DISTRIBUCIÓN ACUMULADA.....	93
FIGURA 20 FUNCIÓN DE DENSIDAD PARA EL TIEMPO DE SERVICIO	93
FIGURA 21 COMPARACIÓN DE FDA PARA TA1	94
FIGURA 22 PRUEBA DE K-S PARA TA1	94
FIGURA 23 PRUEBA DE K-S PARA TA2	95
FIGURA 24 CAPTURA DE PANTALLA DE PROMODEL, SIMULACIÓN	97
FIGURA 25 CAPTURA DE PANTALLA DE PROMODEL, SIMULACIÓN	99

PRESENTACIÓN

El presente trabajo de Tesis intitulado “**SIMULACIÓN DE SISTEMAS DINÁMICOS EN LÍNEAS DE ESPERA, APLICADA AL TRÁFICO EN LAS VENTANILLAS DE UN BANCO**” se desarrolla aplicando la teoría de colas y la simulación en ProModel con el propósito de encontrar un sistema que optimice la atención de los usuarios que acuden a las ventanillas de un banco a realizar sus transacciones financieras, que como se sabe muchas veces se invierte tiempo en las colas de espera para el momento de su atención, generando una pérdida de dinero tanto en los usuarios como en las empresas bancarias.

Este trabajo se realizó con la idea de motivar a los estudiantes a investigar en el uso de programas informáticos referentes a la simulación que ayuda a minimizar los costos de experimentación, así como también ampliar el horizonte de conocimientos hacia la estadística experimental, que estoy seguro será de mucho interés también de Profesionales dedicados a la investigación en nuestra universidad tanto a nivel local y nacional.

Br. Joel Grimaldo Olarte Estrada

Enero, 2019

RESUMEN

El propósito de este estudio es el de analizar las líneas de espera del tráfico en las ventanillas en una agencia de un banco de la ciudad del Cusco, para esto se ha recolectado datos durante una semana en el mes de octubre del año 2017, que fueron analizados mediante la teoría de colas y simulados con el programa ProModel, la metodología considera un nivel descriptivo de tipo aplicado cuyo propósito es solucionar el problema real, de tal manera que se pretende especificar propiedades y características importantes del fenómeno en estudio; en primer lugar la recolección de datos se realizó a los clientes que fueron servidos en la agencia del banco, durante cinco días de la semana, aplicando la técnica censal, recolectando información cuantitativa; en segundo lugar se realizó un análisis de las series de las variables involucradas (tiempo entre llegadas (TLL) y tiempo de servicio (TA)) utilizando el Software libre R (V 3.5.1); por último se aplicó una simulación con el Software libre ProModel (V 2016) para determinar el comportamiento del sistema. Los resultados obtenidos nos indican que los clientes llegan al banco con una distribución de probabilidad exponencial con una media de 1.29 minutos, el tiempo de atención al cliente en el servidor con una distribución de probabilidad exponencial de media 1.6 minutos, determinando que $\rho > 1$ por lo que la cola no se estabiliza, generando un modelo M/M/1; por lo que hubo la necesidad de simular el sistema con dos servidores utilizando el ProModel y así conseguir la estacionariedad del sistema. Por tanto se pudo concluir que el sistema funciona con mayor eficiencia aplicando un modelo M/M/2, analizando el impacto de las mejoras de nuestro estudio y sus respectivos costes que resultan óptimos para la agencia bancaria y sobre todo lo que más importa la satisfacción de los clientes.

Palabras clave: Simulación, sistemas dinámicos, líneas de espera, teoría de colas

ABSTRACT

The purpose of this study is to analyze the waiting lines of the traffic in the servers in an agency of a bank in the city of Cusco, for this data has been collected during a week in the month of October of the year 2017, which were analyzed using the theory of queues and simulated with the ProModel program, the methodology considers a descriptive level of applied type whose purpose is to solve the real problem, in such a way that it is intended to specify properties and important characteristics of the phenomenon under study; First of all, the data collection was done to the clients that were served in the bank's agency, during five days of the week, applying the census technique, collecting quantitative information; Secondly, an analysis of the series of the variables involved (time between arrivals (TLL) and time of service (TA)) was carried out using Free Software R (V 3.5.1); Finally, a simulation was applied with the Free Software ProModel (V 2016) to determine the behavior of the system. The results obtained indicate that customers arrive at the bank with an exponential probability distribution with an average of 1.29 minutes, customer service time on the server with an exponential probability distribution of 1.6 minutes average, determining that $\rho > 1$ per what the tail does not stabilize, generating a model $M / M / 1$; so there was the need to simulate the system with two servers using the ProModel and thus get the stationarity of the system. Therefore it was concluded that the system works with greater efficiency by applying an $M / M / 2$ model, analyzing the impact of the improvements of our study and their respective costs that are optimal for the banking agency and above all what matters most the satisfaction of the clients.

Keywords: Simulation, dynamic systems, waiting lines, queuing theory

INTRODUCCION

El propósito de este estudio es el de analizar las líneas de espera en una agencia de un banco de la ciudad del Cusco. Para ello se han recolectado datos durante una semana del mes de octubre del año 2017, que tendrán que ser modelados mediante un análisis de teoría de colas y simulación.

En el banco en estudio, encontramos un sistema de atención al público. Estos generalmente son gente local, nacional y extranjeros, los cuales llegan individualmente. El método de atención parece bueno en determinadas horas y épocas del año, pero en las últimas horas de la tarde cuando hay afluencia alta de clientes al banco esta atención es ineficiente, por lo cual es necesario hacer un estudio de Teoría de Colas.

En este estudio primeramente haremos una revisión bibliográfica sobre la teoría de colas, series de tiempo y simulación. Luego utilizando nuestros datos que se tomaron durante una semana, haremos el correspondiente análisis, haciendo uso de la teoría de colas.

La Tesis está estructurada en 5 capítulos, En el primer capítulo realizamos el planteamiento del problema, en el segundo capítulo damos a conocer el Marco Teórico necesario para el desarrollo de la investigación, en el capítulo III determinamos la hipótesis identificando las variables en estudio, en el capítulo IV determinamos la metodología de la investigación y en el Capítulo V determinamos los resultados obtenidos aplicando la simulación mediante los Software informáticos de R Project (R) y ProModel, para finalmente analizar el impacto de las mejoras de nuestro estudio y sus respectivos costes.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. Situación problemática

En los últimos años la Ciudad del Cusco presentó un crecimiento poblacional exponencial (según censo nacional 2017 INEI) generando mayor demanda en los servicios bancarios, que en algunos casos se presentó disconformidad de los usuarios, generando muchos reclamos a las gerencias de los bancos por parte de los clientes debido sobre todo a la falta de servidores.

El uso de la Estadística auxilia tareas profesionales importantes procurando sintetizar y representar en forma comprensiva la información cotidiana de un conjunto de datos. Una cola implica clientes que llegan y esperan a ser servidos en las ventanillas de un banco que proporciona el servicio, si la cola que se forma es larga, los clientes pueden impacientarse y marcharse, de manera que el banco dejaría de percibir ingresos por parte de los clientes, por lo que la gerencia del banco debe decidir si vale la pena implementar más ventanillas de atención y si el costo por éste es cubierto por las comisiones de las operaciones bancarias atendidas.

Por lo tanto es necesario realizar un estudio del comportamiento de los tiempos de llegada y de atención para determinar el tiempo que un cliente espera para ser atendido en ventanilla, en base a estos datos realizar una simulación que permita hacer una propuesta para mejorar la gestión de toda entidad bancaria.

1.2. Formulación del problema

a. Problema general

¿Cuál es la eficiencia de las líneas de espera aplicadas al tráfico en las ventanillas de una agencia bancaria de la ciudad de Cusco?

b. Problemas específicos

1. ¿Cómo es el comportamiento del tráfico en las ventanillas del banco?
2. ¿Cuál es el modelo de teoría de colas en el tráfico en las ventanillas del banco que garantice mayor eficiencia mediante la simulación?

1.3. Justificación de la investigación

El trabajo de investigación titulado “**Simulación de sistemas dinámicos para líneas de espera, aplicada al tráfico en las ventanillas de un banco**” se justifica por la necesidad que se tiene de encontrar un servicio eficiente, en el cual no se tenga que esperar mucho o no ser servido adecuadamente por las ventanillas del banco. La agencia del banco es una entidad financiera cuyo fin es brindar sus servicios a clientes internos y externos, en general toda la comunidad. La agencia del banco, tiene horario de atención de 9:00am a 1:00pm y de 3:00pm a 6:00pm de lunes a viernes y los días sábados de 9:00am a 12:30pm. Para el banco los clientes son lo primero y cada día todos sus clientes confían sus logros, sus proyectos empresariales, proyectos familiares y personales, es por eso que este banco para cada acción e idea que realiza debe estar fundamentado por la existencia de los clientes, ya que debido a ellos existe el banco. Es por ésta razón que se justifica hacer un estudio de las líneas de espera de los clientes en el banco.

1.4. Objetivos de la investigación

a. Objetivo general

Analizar y optimizar el sistema de las líneas de espera aplicada al tráfico en las ventanillas de una agencia bancaria de la ciudad de Cusco.

b. Objetivos específicos

1. Describir el comportamiento del tráfico en las ventanillas de un banco.
2. Determinar el modelo de teoría de colas en el tráfico en las ventanillas del banco mediante la simulación

CAPÍTULO II

MARCO TEÓRICO

2.1. Bases teóricas

2.1.1. Introducción a la teoría de colas

Las "colas" son un aspecto de la vida moderna que encontramos continuamente en nuestras actividades diarias. En la caja de un supermercado, en los cajeros automáticos, en el comedor universitario, en los Bancos, etc.; el fenómeno de las colas surge cuando unos recursos compartidos necesitan ser accedidos para dar servicio a un elevado número de trabajos o clientes. El estudio de las colas es importante porque proporciona tanto una base teórica del tipo de servicio que podemos esperar de un determinado recurso, como la forma en la cual dicho recurso puede ser diseñado para proporcionar un determinado grado de servicio a sus clientes (Hiller & Lieberman, 2002)

2.1.1.1. Definiciones iniciales

Existen varias definiciones sobre la teoría de colas, una de ellas y de suma importancia es la que menciona Jaime Enrique Varela en el libro *Introducción a la Investigación de Operaciones*, en el que indica que la Teoría de Colas se ocupa del análisis matemático de los fenómenos de las líneas de espera o colas. Además, menciona que las colas se presentan con frecuencia cuando se solicita un servicio por parte de una serie de clientes y tanto el servicio como los clientes son de tipo probabilístico (Cao, 2002a)

La teoría de colas estudia matemáticamente el comportamiento de las líneas de espera. Una línea de espera se presenta cuando un conjunto de personas "clientes" llegan a un determinado "lugar" solicitando un servicio a un "servidor", que tiene una cierta capacidad de atención. Observando que si el servidor no está disponible inmediatamente y que el cliente decide esperar, entonces se genera una línea de espera (Hiller & Lieberman, 2002)

Un sistema de colas es un modelo que proporciona servicio, es decir, puede representar cualquier sistema en donde los clientes u objetos llegan buscando un servicio de un tipo específico y se retiran después de que dicho servicio haya sido atendido. Es posible modelar los sistemas de este tipo así como las colas sencillas o un sistema de colas que están interconectadas formando una red de colas. (Hiller & Lieberman, 2002; Martínez, Morales, Barber, & Aparicio, 2007)

Uno de los problemas de teoría de colas es determinar la capacidad o tasa de servicio que proporciona el balance correcto. Esto se hace difícil, puesto que un cliente no siempre llega en un horario fijo, es decir, no se sabe con exactitud el momento que llegarán los clientes y en muchas ocasiones el tiempo de servicio no es constante. (Dominguez & Bocanegra, 2016)

Muchos problemas de "colas" se presentan continuamente en nuestra vida diaria: un estudio realizado en EEUU concluyó que en término medio, una persona pasa cinco años de su vida esperando en diferentes colas, y de éstas casi seis meses pasa parado en los semáforos (Hiller & Lieberman, 2002)

2.1.1.2. Origen

El origen de la Teoría de Colas se remonta en el esfuerzo de Agner Kraup Erlang (Dinamarca, 1878 – 1929) quien en 1909 realizó el estudio de la congestión de tráfico telefónico cuyo objetivo era satisfacer la demanda que se observaba muy incierta sobre los servicios en el sistema telefónico de la ciudad de Copenhague. Las investigaciones realizadas dieron a conocer el fundamente de la Teoría de colas. Esta teoría hoy en día es una herramienta valiosa en muchas disciplinas que puedan ser caracterizados como problemas de congestión con un sistema de llegada–salida (Kaufmann, 1978).

2.1.1.3. Modelo de formación de colas

En los problemas de formación de cola, se menciona muy a menudo a los clientes, como personas que esperan la desocupación de un determinado servicio y que puedan ser atendidos, la espera de máquinas para ser reparadas, los aviones que esperan aterrizar y estaciones de servicios, tales como mesas en un restaurante, operarios en un taller de reparación, pistas en un aeropuerto, etc. En

los problemas de formación de colas se presenta una velocidad variable de llegada de clientes que solicitan cierto tipo de servicio, y una velocidad también variable del tiempo utilizado en la estación de servicio (Hiller & Lieberman, 2002)

Cuando hablamos de las líneas de espera, nos referimos a las que se generan por los clientes o también por las estaciones que prestan servicio. Un cliente tiene que esperar en la cola por que los medios que utilizan los servidores son inadecuados para satisfacer la demanda del servicio; en este caso, se dice que la cola tiende a ser explosiva, es decir, cada vez se hace mas larga a medida que el tiempo avanza. Las estaciones de servicio pueden estar esperando por que los medios existentes son excesivos en relación con la demanda de los clientes; en este caso, las estaciones de servicio podrían permanecer esperando la mayor parte del tiempo. Los clientes pueden esperar un cierto intervalo de tiempo, aunque las instalaciones de servicio sean adecuadas, por que los clientes llegados anteriormente están siendo atendidos. Las estaciones de servicio pueden esperar temporalmente cuando, las instalaciones sean adecuadas a largo plazo o haiga una escasez de demanda de los usuarios debido a un hecho ocasional. Estos dos últimos casos muestran una situación equilibrada que tiende constantemente hacia el equilibrio, o una situación estable

Un sistema de colas se especifica por seis características principales:

- a. El tipo de distribución de entradas o llegadas (tiempo entre llegadas)
- b. El tipo de distribución de salidas o retiros (tiempo de servicio)
- c. Los canales de servicio
- d. La disciplina del servicio
- e. El número máximo de clientes permitidos en el sistema
- f. La fuente o población.

El patrón de llegadas de los clientes generalmente está especificado por el tiempo entre llegadas, que es el tiempo que transcurre entre la llegada de un cliente a otro al sistema que ofrece el servicio. Es importante indicar que a veces los clientes prefieren no esperar en la cola para recibir el servicio y es cuando observamos dos casos, el rechazo y el abandono, el primero ocurre cuando el cliente observa una cola demasiado grande o larga y prefiere no ingresar a ella, el segundo caso se presenta cuando un usuario se encuentra en la cola, pero prefiere dejarla (Cao, 2002a)

El patrón de servicio está especificado por el tiempo de servicio, que es el tiempo que le toma a un servidor atender a un cliente. En esta parte es importante determinar si un servidor atiende por completo a un cliente o si el cliente requiere una serie de servidores. Para este

estudio se necesita que un servidor atienda por completo a un usuario (Carro & Gonzales-Gonzales, 2012)

Los canales de servicio son el proceso o sistema que está efectuando el servicio para el cliente de manera complementaria, y podrían ser canales de servicio en serie, paralelo o mixto, es decir una combinación de ambas. La diferencia entre un canal en serie y el paralelo es el número de clientes que pueden ser atendidos de manera simultánea, esto es que se pueden atender varios clientes al mismo tiempo en un canal paralelo, sin embargo en un canal en serie los clientes tendrán que pasar por todos los canales hasta obtener el servicio (Kaufmann, 1978).

La disciplina del servicio determina la forma en que los usuarios son atendidos por el sistema, se considera la disciplina FIFO, LIFO y RSS.

El número máximo de clientes permitidos, es el cupo de clientes que pueden estar en la cola, dependiendo de las características que presenta el sistema, es decir, se podrá tener una cola infinita o finita. Si una cola es infinita no hay problema en que lleguen mil clientes ya que los mil clientes podrán ser atendidos; mientras que en una cola finita hay un cupo máximo o límite y cuando la cola se encuentra llena, los demás clientes serán rechazados, en este caso se le conoce como frustración (Kaufmann, 1978).

Finalmente, la fuente o población representa un factor importante en el análisis de teoría de colas ya que el modelo de llegadas depende de la fuente de donde provienen los clientes. Así pues la cola puede verse de la siguiente manera:

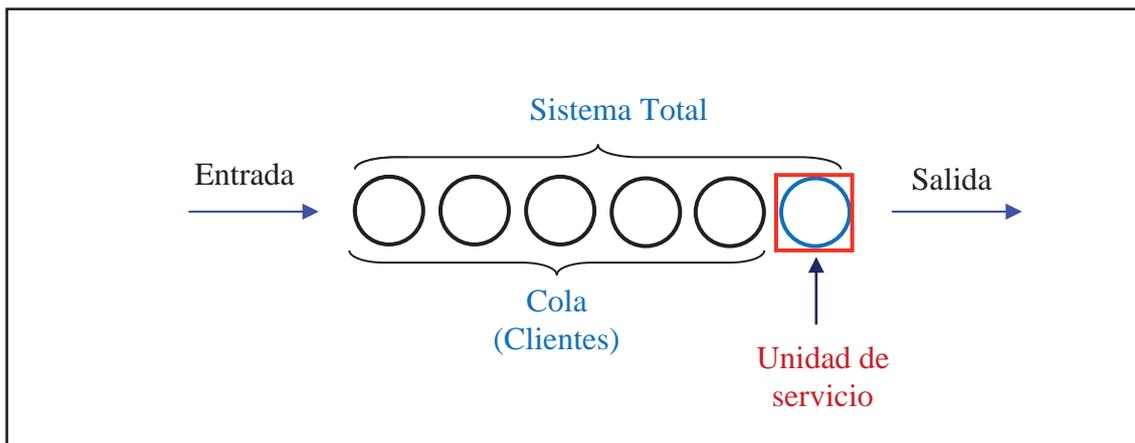


Figura 1. Elementos principales de un sistema de colas

Nota: El esquema presentado fue diseñado por el autor

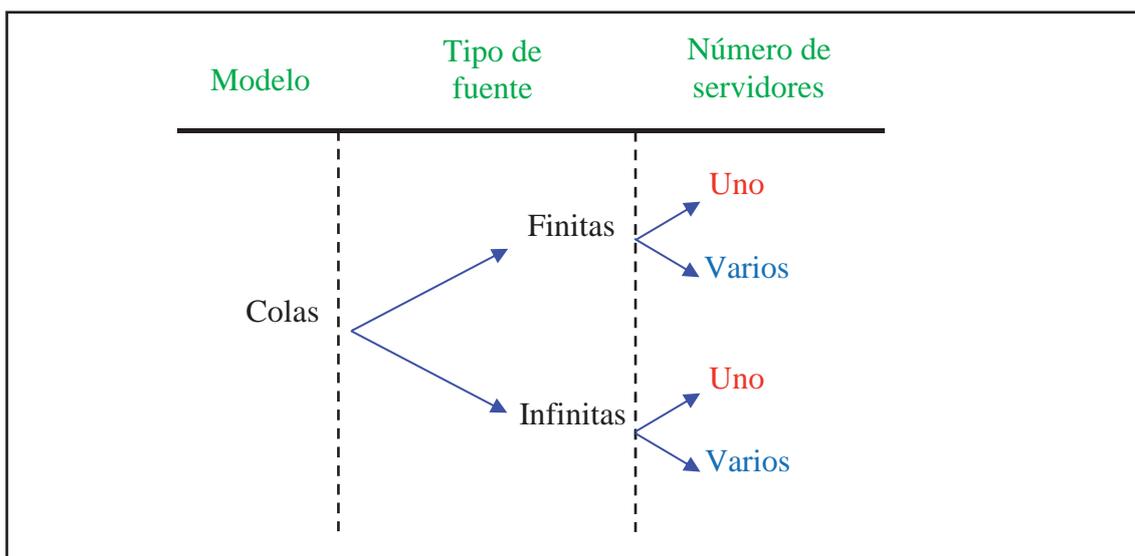


Figura 2. Clasificación del modelo de colas

Nota: El esquema presentado fue diseñado por el autor

2.1.1.4. Objetivos de la Teoría de Colas

Los objetivos principales de la teoría de colas son:

- a. Identificar el nivel óptimo de la capacidad del sistema en el que se obtenga el mínimo costo global del mismo.
- b. Evaluar el impacto que las posibles alternativas de modificación de la capacidad del sistema tendrían en el coste total del mismo.
- c. Encontrar un balance equilibrado ("balance óptimo") entre las consideraciones cuantitativas de los costes y las cualitativas aplicadas al servicio.
- d. Se debe prestar atención al tiempo que permanece un cliente en el sistema o en la cola: la "paciencia" de los clientes dependerá también del tipo de servicio que se brinda y eso puede llevar al abandono del cliente en el sistema.

2.1.1.5. Elementos de la Teoría de Colas

Fuente de entrada o población potencial: Es el conjunto de individuos u objetos (no necesariamente seres vivos) que pueden llegar a solicitar un servicio. Se puede considerar finita o infinita. Aunque el caso de infinita no es realista sino relativa, puesto que permite (por extraño que parezca) resolver de forma más sencilla muchas situaciones en las que, la población es finita pero muy grande. El supuesto de población infinita no resulta restrictiva cuando, aun siendo finita la población potencial, su número de elementos es tan grande que el número de individuos que ya están solicitando el citado servicio prácticamente no afecta a la frecuencia con la que la

población potencial genera nuevas peticiones de servicio (Cao, 2002, p. 129).

Cliente: Es todo elemento de la población potencial que solicita un servicio. Suponiendo que los tiempos de llegada de clientes consecutivos son $0 < t_1 < t_2 < \dots$, será importante conocer la distribución de probabilidad según la cual, la fuente de entrada genera clientes. Lo más recomendable es tomar como referencia los tiempos entre las llegadas de dos clientes consecutivos: $T\{k\} = t_k - t_{k-1}$, fijando su distribución de probabilidad. Normalmente, cuando la población potencial es infinita se supone que la distribución de probabilidad de los T_k (que será la llamada distribución de los tiempos entre llegadas) no depende del número de clientes que estén en espera de completar su servicio, mientras que en el caso de que la fuente de entrada sea finita, la distribución de los T_k variará según el número de clientes en proceso de ser atendidos (Ferreira, 2004).

Capacidad de la cola: Es la cantidad máxima de clientes que pueden estar haciendo cola esperando a ser servidos. Lo más sencillo, a efectos de simplicidad en los cálculos, es asumir que sea infinita. Aunque en la mayor parte de los casos aplicados a la realidad la capacidad de la cola es finita, no es una gran restricción el suponerla que sea infinita si es extremadamente improbable que no puedan entrar clientes a la cola por haberse llegado a ese número límite en la misma (Ferreira, 2004).

Disciplina de la cola: Es la forma en el que los clientes ingresan para ser servidos. Las más habituales son:

La disciplina **FIFO** (first in first out), (primero en llegar, primero en salir) denominada también FCFS (first come first served): según la cual se atiende primero al cliente que llegó primero.

La disciplina **LIFO** (last in first out), (primero en llegar, último en salir) también conocida como LCFS (last come first served) o pila: que consiste en atender primero al cliente que ha llegado al último.

La **RSS** (random selection of service), (selección en forma aleatoria de servicio) o **SIRO** (service in random order), que selecciona a los clientes de forma aleatoria.

Mecanismo de servicio: Es el procedimiento mediante el cual se da servicio a los clientes que lo solicitan. Para determinar el mecanismo de servicio se debe conocer el número de servidores que funcionan en dicho mecanismo (si dicho número fuese aleatorio, la distribución de probabilidad del mismo) y el tipo de distribución de probabilidad del tiempo de atención que le lleva a cada servidor dar un servicio. En caso de que los servidores tengan distinta destreza para dar el servicio, se debe especificar la distribución del tiempo de servicio para cada uno (Cao, 2002).

La cola, propiamente dicha, es el conjunto de clientes que hacen espera, es decir los clientes que ya han solicitado el servicio pero que aún no han pasado al mecanismo de servicio (Cao, 2002).

Un modelo de sistema de colas debe especificar la distribución de probabilidad de los tiempos de servicio para cada servidor.

La distribución más usada para los tiempos de servicio es la exponencial, aunque es común encontrar la distribución degenerada o determinística (tiempos de servicio constantes) o la distribución Erlang (Gamma).

2.1.1.6. Descripción de un Sistema de Colas

Existen dos clases básicas de tiempo entre llegadas:

Determinístico, en el cual clientes sucesivos llegan en un mismo intervalo de tiempo, fijo y conocido. Un ejemplo clásico es el de una línea de ensamble, en donde los artículos llegan a una estación en intervalos invariables de tiempo (conocido como ciclos de tiempo) (Ferreira, 2004).

Probabilístico, en el cual el tiempo entre llegadas sucesivas es incierto y variable. Los tiempos entre llegadas probabilísticos se describen mediante una distribución de probabilidad (Ferreira, 2004).

En el caso probabilístico, la determinación de la distribución real, a menudo, resulta difícil. Sin embargo, la distribución exponencial, ha probado ser confiable en muchos de los problemas prácticos. La función de densidad, para una distribución exponencial depende de un parámetro, λ (letra griega lambda), y está dada por:

$$f(t) = \frac{1}{\lambda} e^{-\lambda t} \quad (1)$$

en donde λ (lambda) es el número promedio de llegadas en una unidad de tiempo.

El proceso de servicio define cómo serán atendidos los clientes. En algunos casos, puede existir más de un servidor en el sistema en el cual se proporcione el servicio que se requiere. Los supermercados y los bancos, son ejemplos de este proceso. Cada ventanilla y cada registradora son servidores que proporcionan el mismo servicio. A tales estructuras se les conoce como sistemas de colas de canal múltiple. En dichos sistemas, los servidores pueden ser idénticos, en el sentido en que proporcionan la misma clase de servicio con igual rapidez, o pueda que no sean idénticos. Por ejemplo, si todos los trabajadores en las ventanillas de un banco tienen la misma experiencia, se pueden considerarse como idénticos.

Un caso contrario de un sistema de canal múltiple, es un proceso de producción con una estación de trabajo que proporciona el servicio requerido, donde todos los productos deben pasar por esa estación de trabajo; en este caso se trata de un sistema de colas de canal sencillo; es importante hacer notar que incluso en un sistema de canal sencillo pueden existir muchos servidores que en conjunto llevan a cabo la tarea necesaria. Por ejemplo, en un negocio de lavado a mano de automóviles, que es una sola estación, puede tener dos empleados que trabajan en un auto de manera simultánea (Ferreira, 2004)

Otra característica del proceso de servicio es el número de clientes atendidos al mismo tiempo en una estación. En los bancos, solamente un cliente es atendido a la vez por cada servidor; por el contrario, los pasajeros que esperan en una estación de tren son atendidos en grupo, según la capacidad del tren que llegue.

Con un tiempo de servicio determinístico, cada cliente requiere precisamente de la misma cantidad conocida de tiempo para ser atendido (Kaufmann, 1978). Con un tiempo de servicio probabilístico, cada cliente requiere una cantidad distinta e incierta de tiempo de servicio. Los tiempos de servicio probabilísticos se describen matemáticamente mediante una distribución de probabilidad. En la práctica resulta difícil determinar cuál es la distribución real, sin embargo, una distribución que ha resultado confiable en muchas aplicaciones, es la distribución exponencial. En este caso, su función de densidad depende de un parámetro, digamos (la letra griega mu) y está dada por:

$$s(t) = \frac{1}{\mu} e^{-\mu t} \quad (2)$$

donde:

μ = número promedio de clientes atendidos por unidad de tiempo,
de modo que:

$1/\mu$ = tiempo promedio invertido en atender a un cliente

Por general el tiempo de servicio puede seguir cualquier distribución, sin embargo antes de que pueda analizar el sistema, se necesita identificar la distribución de probabilidad.

Preguntas relacionadas con el tiempo, centradas en el cliente, como:

- a. ¿Cuál es el tiempo promedio que un cliente recién llegado tiene que esperar en la fila antes de ser atendido?. La medida de rendimiento asociada es el tiempo promedio de espera, representado con **Wq**
- b. ¿Cuál es el tiempo que un cliente invierte en el sistema entero, incluyendo el tiempo de espera y de servicio?. La medida de rendimiento asociada es el tiempo promedio en el sistema, denotado por **W**

Preguntas cuantitativas relacionadas al número de clientes, como:

- a. En promedio ¿cuántos clientes están esperando en la cola para ser atendidos?. La medida de rendimiento asociada es la longitud media de la cola, representada con **Lq**
- b. ¿Cuál es el número promedio de clientes en el sistema?. La medida de rendimiento asociada es el número medio en el sistema, representado con **L**

Preguntas probabilísticas que implican tanto a los clientes como a los servidores, por ejemplo:

- a. ¿Cuál es la probabilidad de que un cliente tenga que esperar a ser atendido?. La medida de rendimiento asociada es la probabilidad de de espera, que se representa por **Pw**

- b. En cualquier tiempo particular, ¿cuál es la probabilidad de que un servidor esté ocupado?. La medida de rendimiento asociada es la utilización, denotada con **U**. Esta medida indica también la fracción de tiempo que un servidor está ocupado.

- c. ¿Cuál es la probabilidad de que existan n clientes en el sistema?. La medida de rendimiento asociada se obtiene calculando la probabilidad **P₀** de que no haya clientes en el sistema, la probabilidad **P₁** de que haya un cliente en el sistema, y así sucesivamente. Esto tiene como resultado la distribución de probabilidad de estado, representada por P_n , $n = 0, 1, \dots$

- d. Si el espacio de espera es finito, ¿Cuál es la probabilidad de que la cola esté llena y que un cliente que llega no sea atendido?. La medida de rendimiento asociada es la probabilidad de negación del servicio, representada por P_d

Preguntas relacionadas con los costos, como:

- a. ¿Cuál es el costo por unidad de tiempo por operar el sistema?
- b. ¿Cuántos servidores se necesitan para lograr mayor efectividad en los costos?

El cálculo específico de estas medidas de rendimiento depende de la clase de sistema de colas. Algunas de estas medidas están relacionadas entre sí. Conocer el valor de una medida nos permita encontrar el valor de una medida relacionada.

Relaciones entre medidas de rendimiento

El cálculo de muchas de las medidas de rendimiento depende de los procesos de llegadas y de servicio del sistema de colas en específico. Estos procesos son descritos matemáticamente mediante distribuciones de llegada y de servicio. Incluso sin conocer la distribución específica, las relaciones entre algunas de las medidas de rendimiento pueden obtenerse para ciertos sistemas de colas, únicamente mediante el uso de los siguientes parámetros de los procesos de llegada y de servicio.

λ = número promedio de llegadas por unidad de tiempo.

μ = número promedio de clientes atendidos por unidad de tiempo en una sección.

Sea una población de clientes infinita y una cantidad limitada de espacio de espera en la fila, el tiempo total que un cliente invierte en el sistema es la cantidad de tiempo invertido en la fila más el tiempo durante el cual es atendido:

Tiempo promedio en el sistema = Tiempo de espera + Tiempo de servicio

Como ya indicamos anteriormente, el tiempo promedio en el sistema y el tiempo promedio de espera están representados por las cantidades W y W_q , respectivamente. El tiempo promedio de servicio puede expresarse en términos del parámetro λ , así por ejemplo, si λ es 6 clientes por hora, entonces, en promedio cada cliente requiere $1/6$ de hora para ser atendido. Por lo general, el tiempo de servicio es $1/\mu$, lo cual nos conduce a la siguiente relación:

$$W = W_q + \frac{1}{\mu} \quad (3)$$

Consideremos ahora la relación entre el número promedio de clientes en el sistema y el tiempo promedio que cada cliente pasa en el sistema. Imaginemos que un cliente acaba de llegar y se espera que permanezca en el sistema un promedio de 15 minutos, durante este cuarto de hora, otros clientes siguen llegando a una tasa de supongamos 12 por hora. Cuando el cliente en cuestión abandona el sistema, después de 15 minutos, deja tras de sí un promedio de $\frac{1}{4} \times 12 = 3$ clientes nuevos.

Es decir, en promedio, existen tres clientes en el sistema en cualquier tiempo dado.

N° promedio de clientes en el sistema = N° de llegadas x Tiempo promedio en el sistema.

de modo que:
$$L = \lambda \times W \quad (4)$$

Utilizando una lógica parecida se obtiene la relación entre el número promedio de clientes que esperan en la cola y el tiempo promedio de espera en la cola:

N° promedio de clientes en la cola = N° de llegadas x Tiempo promedio en la cola

de manera que:
$$L_q = \lambda \times W_q \quad (5)$$

2.1.1.7. Notación de Kendall

Por convención los modelos que se aplican en teoría de colas se etiquetan de acuerdo a la distribución que se utilizan tanto en los tiempos de llegada como de servicio, esto es:

- M: Distribución exponencial (markoviana)
- D : Distribución degenerada (tiempos constantes)
- E_k : Distribución Erlang
- G : Distribución general

M / M / S : Modelo donde tanto los tiempos entre llegada como los tiempos de servicio son exponenciales y cuenta con S servidores.

M / G / 1: Tiempos entre llegada exponencial, tiempos de servicio general y cuenta con un sólo servidor

Terminología

Usualmente siempre es común utilizar la siguiente terminología estándar:

- **Estado del sistema:** Número de clientes en el sistema.
- **Longitud de la cola:** Número de clientes que esperan servicio (están en la cola).
- **N(t)** : Número de clientes en el sistema de colas en tiempo t ($t \geq 0$).
- **P_n(t)**: Probabilidad de que exactamente n clientes estén en el sistema en el tiempo t, dado el número en el tiempo cero.
- **S** : Número de servidores en el sistema de colas.
- λ_n : Tasa media de llegadas (número esperado de llegadas por unidad de tiempo) de nuevos clientes cuando hay n clientes en el sistema.
- μ_n : Tasa media de servicio para todo el sistema (número esperado clientes que completan su servicio por unidad de tiempo) cuando hay n clientes en el sistema.

Nota: μ_n representa la tasa combinada a la que todos los servidores ocupados logran terminar sus servicios

Cuando realizamos el estudio de un sistema de colas, es útil tener una forma compacta de referirse a las características más importantes del sistema. Esto se consigue a través de la notación $A/B/X/Y/H/Z$, que representa lo siguiente:

- A:** Indica cuál es la distribución que rige el tiempo entre llegadas.
- B:** Indica cuál es la distribución que rige el tiempo entre servicios.
- X:** Indica el número de servidores que trabajan en paralelo en el sistema.
- Y:** Indica la capacidad del sistema
- H:** Indica el tamaño de la población de los clientes potenciales.
- Z:** Indica la disciplina de la cola.

La siguiente tabla resume los posibles valores que puede tomar cada uno de los componentes anteriores.

Tabla 1
Posibles valores según la notación $A/B/X/Y/H/Z$

Característica	Valores típicos	Significado
A: Distribución del tiempo entre llegadas	M	Exponencial
	D	Determinista
	E_k	Erlang de parámetro k
B: Distribución del tiempo entre servicios	H_k	Mezcla de k exponenciales
	G	Una distribución general cualquiera
X: Número de servidores	1, 2, ..., ∞	
Y: Capacidad del sistema	1, 2, ..., ∞	
H: Población potencial		
	FIFO	Primero en entrar, primero en salir
	LIFO	Último en entrar, primero en salir
	RSS	Orden de servicio aleatorio
	PR	Orden de servicio priorizado
Z: Disciplina cola	GD	Una disciplina general cualquiera

Fuente: Tabla tomada del libro "Fundamentals of Queueing Theory"

Nota: Cuando estudiamos un sistema de colas, tenemos el valor G para el patrón de llegada o el patrón de servicio cuando no se asume alguna distribución concreta y que los resultados que se obtienen serán ciertos en general. Como se indicó que la función exponencial es la más usada en la práctica, se denote por M en vez de E , esto se debe a que E se reserva para las distribuciones de Erlang (que es un caso particular de la distribución gamma). En este caso se usa la letra M para hacer referencia a la propiedad de Markov (falta de memoria) de la distribución exponencial.

2.1.1.8. Cola M/M/1

Presentamos ahora resultados generales, que son ciertos para sistemas $G/G/1$, $G/G/s$ y $D/D/1$. Estos resultados tienen bastante interés en sí mismos, ya que proporcionan una primera información acerca del comportamiento de la cola y, además, serán útiles en el desarrollo de resultados para los modelos específicos.

2.1.1.9. Un poco más de notación

Los resultados que obtendremos en esta sección sólo dependerán de los tiempos medios de llegada y de servicio. La notación relacionada con estos tiempos medios será:

λ : Número medio de clientes que entran en el sistema por unidad de tiempo.

μ : Número medio de clientes que un servidor atiende por unidad de tiempo.

$\frac{1}{\lambda}$: Tiempo medio entre dos llegadas consecutivas al sistema.

$\frac{1}{\mu}$: Tiempo medio entre dos servicios consecutivos del mismo servidor.

Observamos que λ y μ denotan la frecuencia y $\frac{1}{\lambda}$, $\frac{1}{\mu}$ denotan el período.

Frecuencia: número de repeticiones por unidad de tiempo.

Período: tiempo entre dos repeticiones consecutivas.

Como veremos más adelante cuando estudiemos alguna distribución específica, esta relación frecuencia–período se encuentra también entre las distribuciones de Poisson (frecuencia) y exponencial (período).

S: Número total de servidores.

$\rho = \frac{\lambda}{S\mu}$: Intensidad de tráfico. Esto denota la fracción de servicio ($S\mu$)

usada por los clientes que llegan (λ).

El objetivo será estudiar el comportamiento a largo plazo del sistema de una cola, cuando la cola se “asiente”. Queremos ver las propiedades estacionarias del sistema:

- Si $\rho > 1$, la cola nunca se asienta. La tasa de llegadas es superior a la tasa de servicio y el número de clientes en la cola siempre aumenta, tendiendo a ∞ .
- Si $\rho = 1$, salvo que llegadas y servicios sean deterministas, el sistema tampoco se estabilizará
- Si $\rho < 1$, entonces el sistema se acabará estabilizando.

Si estamos diseñando un sistema y tenemos estimaciones fiables de λ y μ , entonces ya tenemos una condición necesaria para poder satisfacer la demanda: $\frac{\lambda}{S\mu} < 1$. Es decir, necesitaremos un número S de servidores tal que $\frac{\lambda}{\mu} < S$. Si estamos ante un sistema para el que está condición no se cumple todavía podemos intentar estudiar cosas como el ritmo al que se saturará, pero no podremos obtener ninguna propiedad estacionaria.

Habitualmente, lo que más nos interesa conocer al resolver modelos de colas es la distribución de probabilidad del número de clientes en el sistema en el instante t , que denotaremos por $N(t)$.

Este número es la suma del número de clientes en la cola $N_q(t)$ y el número de clientes en servicio, $N_s(t)$ (el uso de la letra q es por la palabra inglesa *queue* que significa cola). Debemos notar que, para cada t , $N(t)$, $N_q(t)$ y $N_s(t)$ son variables aleatorias (con una cierta distribución de probabilidad). De hecho, $N(t)$, $N_q(t)$ y $N_s(t)$ son procesos estocásticos en tiempo continuo y con espacio de estados discreto. De modo similar, usaremos las variables aleatorias N , N_q y N_s para referirnos al número de clientes en el estado estacionario. Usaremos la notación $p_n(t) = P(N(t) = n)$ y $p_n = P(N = n)$.

Como $N_s(t) = N(t) - N_q(t)$

$$\text{y } N_s = N - N_q, \tag{6}$$

Muchas veces no será posible o será muy complicado obtener las distribuciones de las variables aleatorias $N(t)$, $N_q(t)$, N y N_q , con lo que nos conformaremos con los valores medios. Dado un sistema con S servidores, hay dos valores medios de especial importancia:

Número medio de clientes en el sistema:

$$L = E(N) = \sum_{n=0}^{\infty} np_n \quad (7)$$

Número medio de clientes en la cola:

$$L_q = E(N_q) = \sum_{n=s+1}^{\infty} (n-s)p_n \quad (8)$$

Si en vez de preocuparnos por el número de clientes miramos a los tiempos de servicios, entonces nos interesarían las siguientes variables aleatorias asociadas a la distribución estacionaria:

T_q : tiempo que un cliente pasa en cola;

T : tiempo que un cliente pasa en el sistema;

T_S : tiempo que un servidor tarda en servir a un cliente.

En este caso, $T = T_q + T_S$.

A estas variables le podemos asignar los siguientes valores medios:

Tiempo medio en la cola:

$$W_q = E(T_q). \quad (\text{el uso de } W \text{ es por } \textit{Waiting time})$$

Tiempo medio en el sistema:

$$W = E(T).$$

2.1.1.10. Fórmulas de Little

Los siguientes resultados son ciertos para sistemas $G/G/1$ y $G/G/s$. En particular, esto quiere decir que también serán ciertos para colas deterministas. Las fórmulas de Little establecen una de las relaciones más fuertes en teoría de colas (Ramírez Lozano, 1995). Estos resultados fueron desarrollados por John D. C. Little a principios de los años 60. Esta fórmula relaciona los tamaños medios de ocupación del sistema con los tiempos de espera en el mismo:

$$L = \lambda W \quad \text{y} \quad L_q = \lambda W_q \quad (9)$$

$$\text{Además, como: } E(T) = E(T_q) + E(T_s) \quad \text{y} \quad E(T_s) = \frac{1}{\mu} \quad (10)$$

$$\text{Y como: } W = W_q + \frac{1}{\mu} \quad (11)$$

Por tanto, con conocer una de las cuatro medidas L , W , L_q y W_q se podrá determinar las otras tres. Es decir, aunque en la práctica pueda interesarnos más una medida que otra, a la hora de estudiarlas matemáticamente, podemos centrarnos en aquella que sea más sencilla de analizar en cada caso.

Una ilustración gráfica de este resultado es, considerando un ejemplo de lo que podría ser el periodo de actividad de una cola desde que llega el primer cliente hasta que se vuelve a vaciar de nuevo:

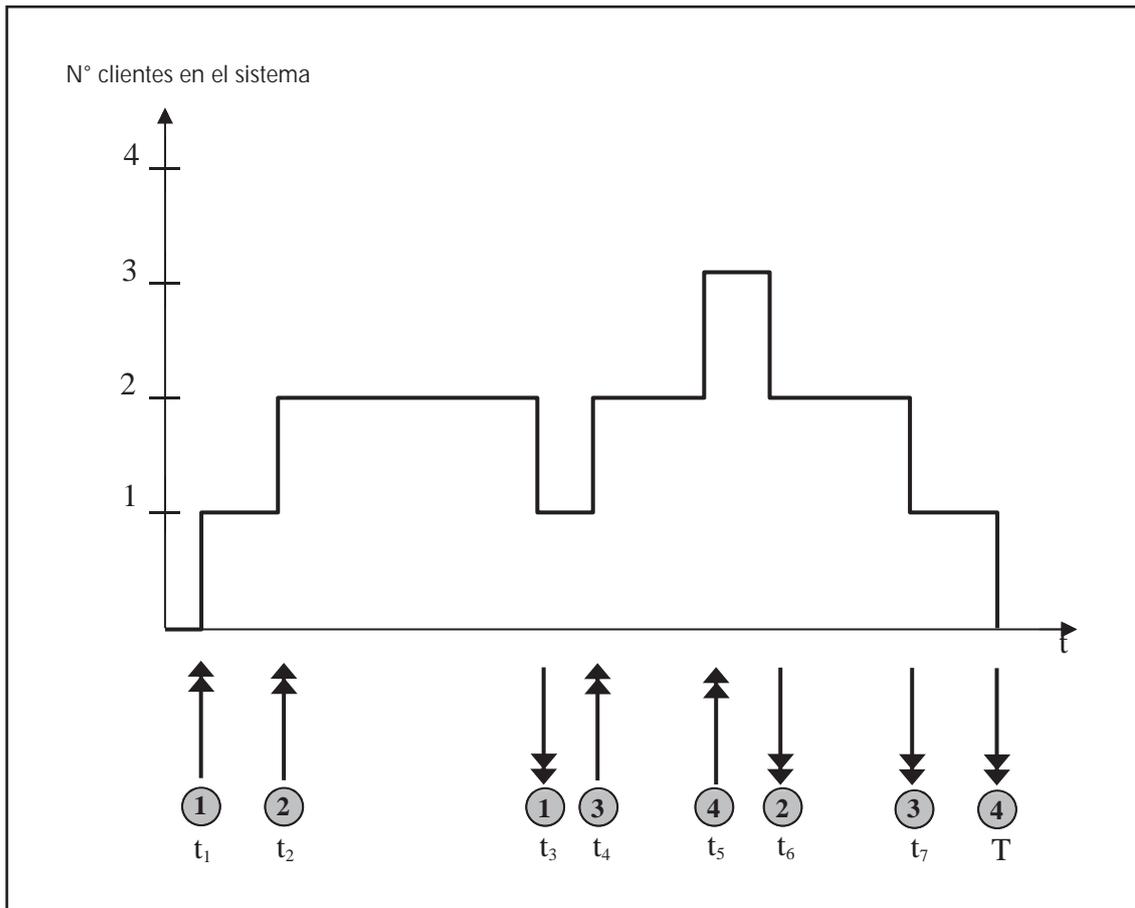


Figura 3. Ilustración de las fórmulas de Little.

Nota: Ilustración tomada del libro "Fundamentals of Queueing Theory"

Según la Figura 3, en el intervalo $\langle 0, T \rangle$ llegan

$N_c = 4$ clientes al sistema.

Veamos ahora cuanto valen L y W en este caso. El número medio de clientes en el sistema es:

$$L = \frac{1(t_2 - t_1) + 2(t_3 - t_2) + 1(t_4 - t_3) + 2(t_5 - t_4) + 3(t_6 - t_5) + 2(t_7 - t_6) + 1(T - t_7)}{T}$$

$$L = \frac{T + t_7 + t_6 - t_5 - t_4 + t_3 - t_2 - t_1}{T}$$

$$L = \frac{\text{área bajo la curva}}{T} \tag{12}$$

y el tiempo medio de espera es:

$$W = \frac{(t_3 - t_1) + (t_6 - t_2) + (t_7 - t_4) + (T - t_5)}{4}$$

$$W = \frac{T + t_7 + t_6 - t_5 - t_4 + t_3 - t_2 - t_1}{4}$$

$$W = \frac{\text{área bajo la curva}}{N_c} \quad (13)$$

De estas dos ecuaciones obtenemos que el área bajo la curva es:

$$LT = W N_c.$$

Despejando, obtenemos: $L = \frac{N_c}{T} W$ (14)

Pero la fracción $\frac{N_c}{T}$, es justamente el número de clientes por unidad de tiempo, es decir λ , de modo que $L = \lambda W$.

Realizando un procedimiento similar podríamos ilustrar también la fórmula $L_q = \lambda W_q$. Gracias a las fórmulas de Little podemos obtener de modo sencillo algunos resultados que nos ayuden a entender de mejor manera el comportamiento de modelos de colas bastante generales.

Por ejemplo, usando las ecuaciones (9), tenemos:

$$L - L_q = \lambda W - \lambda W_q = \lambda(W - W_q) = \lambda\left(\frac{1}{\mu}\right) = \frac{\lambda}{\mu} \quad (15)$$

Además, como $L - L_q = E(N) - E(N_q) = E(N - N_q) = E(N_s)$ (16)

Es decir que el número medio de personas que están en servicio en la distribución estacionaria es $\frac{\lambda}{\mu}$

En particular, para una cola $G/G/1$ esto coincide con la intensidad de tráfico. Notemos que esto nos indica que el número medio de personas en los servidores es independiente del número s de servidores, siempre que se cumpla el supuesto básico de que la intensidad de tráfico sea menor que 1, es decir: $(\rho = \frac{\lambda}{s\mu} < 1)$.

Supongamos ahora que tenemos un sistema $G/G/1$ y queremos conocer la probabilidad P_0 , es decir, la probabilidad de tener 0 clientes en el sistema (en el estado estacionario) y por tanto, el servidor estará inactivo:

$$\rho = \frac{\lambda}{\mu} = L - L_q = \sum_{n=1}^{\infty} np_n - \sum_{n=1}^{\infty} (n-1)p_n = \sum_{n=1}^{\infty} p_n = 1 - p_0 \quad (17)$$

Por tanto, $P_0 = 1 - \frac{\lambda}{\mu} = 1 - \rho$

Supongamos ahora que tenemos un sistema $G/G/s$. Por simetría, tendremos que los $\frac{\lambda}{\mu}$ clientes que, en media, estarán en servicio en la distribución estacionaria se repartirán equitativamente entre los S servidores. Por tanto, el número medio de clientes en un servidor será $\frac{\lambda}{s\mu}$

Denotemos por P_0 la probabilidad de que un determinado servidor esté ocupado. Entonces,

$$\begin{aligned} \rho &= \frac{\lambda}{s\mu} \quad \text{“Número medio de clientes en cada servidor”} \\ &= 0 \cdot (1 - P_0) + 1 \cdot P_0 = P_0 \end{aligned} \quad (18)$$

Tabla 2.

Resumen de resultados para sistemas G/G/s.

Fórmula	Descripción
$\rho = \frac{\lambda}{s\mu}$	Intensidad de tráfico; carga de trabajo por servidor
$L = \lambda W$	Fórmula de Little
$L_q = \lambda W_q$	Fórmula de Little
$W = W_q + \frac{1}{\mu}$	Inmediato usando que el tiempo medio del servidor es $\frac{1}{\mu}$
$L = L_q + \frac{\lambda}{\mu}$	Inmediato a partir de las fórmulas de Little y la relación anterior
$\frac{\lambda}{\mu}$	Número medio de clientes en servicio: carga de trabajo en el sistema
$1 - \rho$	Probabilidad de que un servidor cualquiera esté inactivo
$p_0 = 1 - \rho$	Probabilidad de tener el sistema vacío en un sistema G/G/1

Nota: Apuntes de Teoría de Colas Prof. Julio González Díaz

Una relación entre los tiempos de espera entre dos clientes consecutivos en un sistema G/G/1 será:

W_q^n : Tiempo en la cola del cliente n.

S^n : Tiempo en el servidor del cliente n.

T^n : Tiempo entre la llegada del cliente n y la llegada del cliente n+1.

Luego, trazamos la siguiente figura:

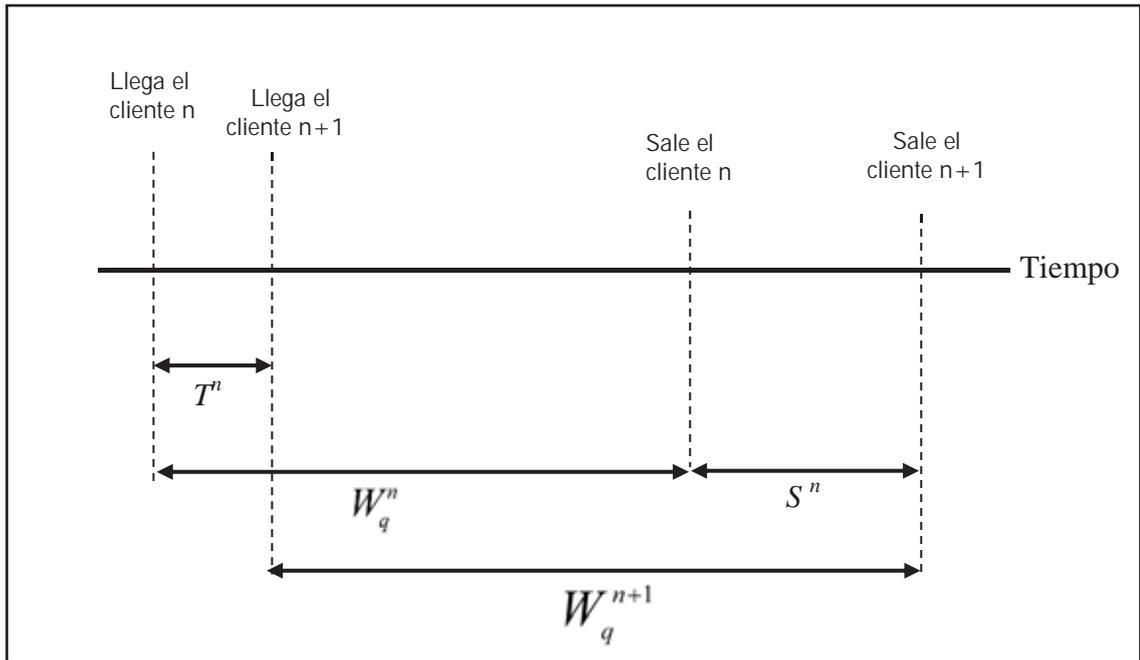


Figura 4. Tiempos de espera entre clientes consecutivos en un sistema G/G/1. Nota: Apuntes de Teoría de Colas Prof. Julio González Díaz

De esta figura, observamos que:

$$W_q^{n+1} = \text{máx}\{W_q^n + S^n - T^n, 0\} \quad (19)$$

Por tanto, si sabemos cuando ha llegado cada cliente y el tiempo que ha estado en el servidor, esta relación recurrente nos permite calcular de modo sencillo los tiempos de espera de todos los clientes

2.1.1.11. Un sistema D/D/1

Vamos a describir ahora algunas características de una cola D/D/1. (D: distribución determinista), los parámetros de interés para una cola de este tipo son los siguientes:

λ : Número de clientes que llegan al sistema por unidad de tiempo

μ : Número de clientes que un servidor sirve por unidad de tiempo.

$\frac{1}{\lambda}$: Tiempo (exacto) entre dos llegadas consecutivas al sistema.

$\frac{1}{\mu}$: Tiempo (exacto) entre dos servicios consecutivos del mismo

servidor.

$N(t)$: Número de clientes en el sistema en el instante t .

Vemos que estamos ante un número y no ante una variable aleatoria, donde $N(t)$ se puede descomponer como:

$N(t) = \text{“N° de llegadas en } \langle 0, t \rangle \text{”} - \text{“N° de servicios en } \langle 0, t \rangle \text{”}$.

Supongamos ahora que $\lambda > \mu$, es decir, $\rho = \frac{\lambda}{\mu} > 1$. Sabemos que

la cola no parará de crecer desde que entre el primer cliente, siempre habrá algún cliente en servicio. En este caso, podría interesarnos saber cómo crecerá el número de clientes en el sistema, $N(t)$ con el paso del tiempo, donde:

$$N(t) = \lambda t - \left\lfloor \mu \left(t - \frac{1}{\lambda} \right) \right\rfloor \quad (20)$$

$$N(t) = \lambda t - \left\lfloor \mu t - \frac{\mu}{\lambda} \right\rfloor \quad (21)$$

y observamos que esta fórmula no sirve cuando $\rho < 1$, puesto que en algunos instantes el servidor estará vacío, es decir si aplicamos la fórmula con $\mu > \lambda$, entonces para t suficientemente grande,

$\lambda t - \left\lfloor \mu t - \frac{\mu}{\lambda} \right\rfloor$ sería negativo.

Para ilustrar de mejor manera esta fórmula en el caso que $\rho > 1$, consideremos el caso que $\lambda = \frac{1}{3}$ (llega 1 cliente cada 3 minutos al

sistema) y $\mu = \frac{1}{5}$ (se atiende 1 cliente cada 5 minutos), luego:

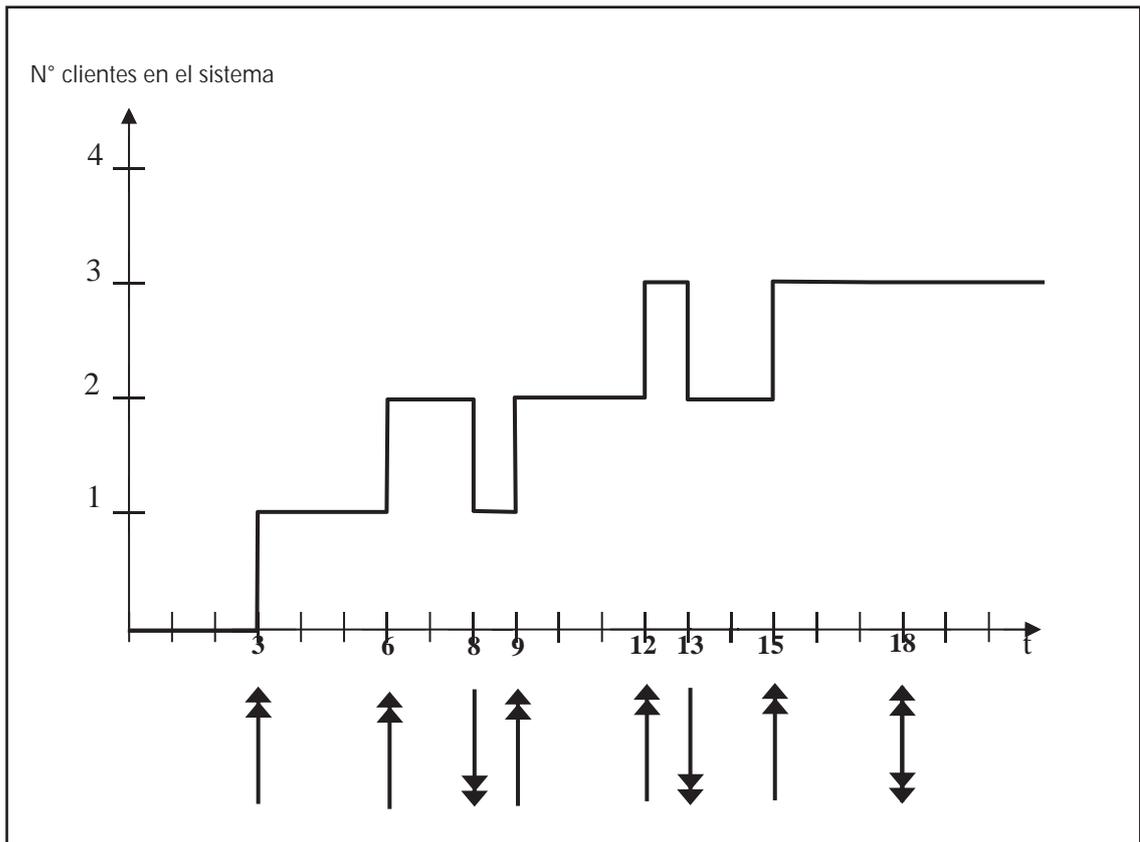


Figura 5. Evolución de un sistema D/D/1 con $\lambda = 1/3$ y $\mu = 1/5$

Nota: Apuntes de Teoría de Colas Prof. Julio González Díaz

Calculando el número de clientes en la cola para diferentes valores de t , aplicando la fórmula (21), tenemos:

$$t = 10 \rightarrow n(10) = \left\lceil \frac{10}{3} \right\rceil - \left\lceil \frac{10}{5} - \frac{3}{5} \right\rceil = 3 - 1 = 2$$

$$t = 18 \rightarrow n(18) = \left\lceil \frac{18}{3} \right\rceil - \left\lceil \frac{18}{5} - \frac{3}{5} \right\rceil = 6 - 3 = 3$$

$$t = 125 \rightarrow n(125) = \left\lceil \frac{125}{3} \right\rceil - \left\lceil \frac{125}{5} - \frac{3}{5} \right\rceil = 41 - 24 = 17$$

Observamos que la cola seguirá aumentando sin parar.

2.1.1.12. Herramientas Probabilísticas: Procesos Estocásticos

Ahora estudiaremos colas específicas, más concretamente, colas con tasas de llegada y servicio exponenciales. Para poder llevar a cabo este análisis es necesario tener unos conocimientos mínimos de procesos estocásticos, en particular, de los procesos de Poisson.

2.1.1.13. Procesos de Poisson

Para definir un proceso de Poisson, necesitamos recordar la definición de las distribuciones de probabilidad de Poisson y de la Exponencial, esto es:

Poisson (discreta): Mide el número de ocurrencias de un suceso por unidad de tiempo. Una distribución de Poisson de parámetro λ viene caracterizada por una función de distribución de probabilidad tal que, para todo entero $n \geq 0$,

$$P(n) = \frac{\lambda^n}{n!} e^{-\lambda} \quad (22)$$

Para esta distribución tanto la media como la varianza coinciden

con el parámetro λ , es decir: $\mu = \lambda$ (23)

$$\sigma^2 = \lambda \quad (24)$$

Exponencial (continua): Mide el tiempo entre dos ocurrencias consecutivas de un suceso. Una distribución exponencial de parámetro λ viene caracterizada por una función de densidad tal que:

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (25)$$

Para esta distribución la media y la varianza están dados por:

$$\mu = 1/\lambda \quad (26)$$

$$\sigma^2 = 1/\lambda^2. \quad (27)$$

2.1.1.14. Relación entre los procesos de Poisson y la distribución exponencial

Veamos ahora algunos resultados que clarifican la conexión existente entre la distribución exponencial y los procesos de Poisson.

Proposición 2.1 *El proceso de llegadas es un proceso de Poisson si y sólo si el tiempo entre dos llegadas consecutivas sigue una distribución exponencial.*

Supongamos que tenemos el sistema de una cola en el que la tasa de servicio viene dada por una $\exp(\mu)$. La propiedad anterior nos dice que si tengo k clientes en servicio entre mis s servidores, entonces el próximo cliente en salir servido lo hará en un tiempo que sigue una distribución $\exp\left(\sum_{i=1}^n \mu\right) = \exp(k\mu)$. En particular, si todos los servidores

están ocupados, este tiempo seguirá una distribución $\exp(s\mu)$ y el tiempo medio que el sistema requiere para atender a un cliente es $\frac{1}{s\mu}$.

Nótese que, debido a la propiedad de falta de memoria, estos tiempos son independientes del tiempo que puedan llevar los clientes en los distintos servidores.

Por último, para presentar la última propiedad de interés de la distribución exponencial, necesitamos definir la distribución de Erlang:

Distribución de Erlang: Una distribución de Erlang de parámetros $k \in \mathbb{N}$

y $\lambda > 0$, denotada por $E_k(\lambda)$, tiene como función de distribución:

$$F(t) = \begin{cases} 1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (28)$$

y como función de densidad,

$$f(t) = \begin{cases} \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (29)$$

La distribución Erlang tiene media $\frac{k}{\lambda}$ y varianza $\frac{k}{\lambda^2}$. Como caso particular, cuando $k = 1$, la distribución de Erlang se reduce a la exponencial; es decir, $E_1(\lambda) = \exp(\lambda)$. Del mismo modo que la distribución Erlang generaliza a la distribución exponencial, es un caso particular de la distribución Gamma; en el caso de esta última, la diferencia radica en que el primer parámetro puede tomar cualquier valor real positivo, no sólo números enteros (cuando toma un valor entero, se reduce a la distribución de Erlang).

Proposición 2.2 (Suma de exponenciales). *Supongamos que tenemos las siguientes variables aleatorias independientes:*

$$T_1 = \exp(\lambda), T_2 = \exp(\lambda), \dots, T_n = \exp(\lambda).$$

Sea $T = T_1 + T_2 + \dots + T_n$. Entonces $T = E_n(\lambda)$

En el contexto de teoría de colas, éste resultado nos está diciendo que el tiempo desde que llega un cliente hasta que llegan n clientes más viene dado por una distribución Erlang $E_n(\lambda)$. Supongamos ahora que un cliente llega al sistema cuando ya hay $n > s$ clientes. En este caso, el resultado anterior relativo al mínimo de exponenciales indicaba que el tiempo de espera hasta que el siguiente cliente salga servido se distribuye según una $\exp(s\mu)$. El resultado que acabamos de ver indica que el tiempo de espera en la cola de este nuevo cliente será $E_{n-s+1}(\mu)$. Además, como será atendido también según una exponencial de parámetro μ , el tiempo total en el sistema estará dado por una $E_{n-s+2}(\mu)$.

2.1.1.15. Modelos de colas con tasas de llegada y servicio de Poisson

En este tema, basado fundamentalmente en el libro “Introducción a la Simulación y a la Teoría de Colas” escrito por Ricardo Cao Abad (Cao, 2002), presentaremos un análisis detallado de distintos sistemas de colas con llegadas y servicio de Poisson (tiempos entre llegadas y servicios exponenciales). Recordemos que, en general, el sistema de una cola viene representado por la notación $A/B/X/Y/H/Z$, donde:

A: Indica cuál es la distribución que rige el tiempo entre llegadas.

B: Indica cuál es la distribución que rige el tiempo entre servicios.

X: Indica el número de servidores en paralelo del sistema.

Y: Indica la capacidad del sistema.

H: Indica el tamaño de la población de clientes potenciales.

Z: Indica la disciplina de la cola.

2.1.1.16. El modelo M/M/1

Recordemos que la letra M representa que las llegadas y servicios siguen un proceso de Poisson. A modo de ilustración del procedimiento de resolución general de un modelo de colas, el modelo $M/M/1$ lo analizaremos de una forma más exhaustiva, derivando de forma explícita todas las expresiones de interés.

Tasa de llegadas $\rightarrow \lambda$

Tasa de servicio $\rightarrow \mu$

Número de servidores $\rightarrow 1$

Capacidad de la cola $\rightarrow \infty$

Población potencial $\rightarrow \infty$

Disciplina de la cola \rightarrow FIFO

Antes de pasar al análisis formal del modelo $M/M/1$ será conveniente recordar una nueva distribución de probabilidad.

Recordemos que, en general, a la hora de estudiar un modelo de colas, estaremos interesados en conocer la distribución de las variables aleatorias N y Nq , que cuentan el número de clientes en el estado estacionario. Más concretamente, estamos interesados en conocer las probabilidades p_n que caracterizan las distribuciones de probabilidad de dichas variables. En este modelo, las tasas de llegada y de servicio son completamente independientes del número de clientes que haya en el sistema. Por tanto, usando las fórmulas vistas

para los procesos generales de nacimiento y muerte, tenemos que,

para todo $n \geq 1$, $p_n = c_n p_0$, con $p_0 = 1 - \frac{\lambda}{\mu} = 1 - \rho$ y $c_n = \left(\frac{\lambda}{\mu}\right)^n$

Por tanto:

$$P(N = n) = p_n = (1 - \rho)\rho^n \quad (30)$$

es decir, la variable aleatoria N , que cuenta el número de clientes en el sistema en el estado estacionario viene dada por una distribución geométrica de parámetro $1 - \rho$ (recordemos que una condición necesaria y suficiente para que dicha distribución estacionaria existiese es que $\rho < 1$, con lo que $1 - \rho > 0$ y la correspondiente distribución geométrica está bien definida).

Usando la fórmula de la media de una distribución geométrica, el número medio de clientes en el sistema, $L = E(N)$ vendrá dado por:

$$L = E(N) = \frac{1 - (1 - \rho)}{1 - \rho} = \frac{\rho}{1 - \rho} = \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} = \frac{\lambda}{\mu - \lambda} \quad (31)$$

Para obtener el valor de L de forma directa, procedemos:

$$L = E(N) = \sum_{n=1}^{\infty} n(1 - \rho)\rho^n = (1 - \rho)\rho \sum_{n=1}^{\infty} n\rho^{n-1} \quad (32)$$

Para calcular $\sum_{n=1}^{\infty} n\rho^{n-1}$ basta tener en cuenta lo siguiente:

Si $f(x) = \sum_{n=1}^{\infty} x^n$, entonces (por la fórmula de la progresión geométrica)

tenemos que, para todo $x \in \langle -1, 1 \rangle$, $f(x) = \frac{x}{1 - x}$ Si ahora derivamos,

obtenemos que, para todo $x \in \langle -1, 1 \rangle$,

$$\sum_{n=1}^{\infty} nx^{n-1} = f'(x) = \frac{1(1-x) - (-1)x}{(1-x)^2} = \frac{1}{(1-x)^2} \quad (33)$$

Sustituyendo $x = \rho$, en esta expresión obtenemos que $\sum_{n=1}^{\infty} n\rho^{n-1} = \frac{1}{(1-\rho)^2}$

por tanto:

$$L = (1-\rho)\rho \sum_{n=1}^{\infty} n\rho^{n-1} = (1-\rho)\rho \frac{1}{(1-\rho)^2} = \frac{\rho}{1-\rho} \quad (34)$$

Una vez calculado L , podemos usar las fórmulas de Little para obtener el número medio de clientes en la cola, L_q , el tiempo medio de espera en el sistema W , y el tiempo medio de espera en la cola, W_q :

Fórmulas de Little:

$$L = \lambda W \quad y \quad L_q = \lambda W_q \quad (W = W_q + \frac{1}{\mu}) \quad (35)$$

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

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

$$W_q = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)} \quad (36)$$

$$L_q = \lambda W_q = \frac{\lambda^2}{\mu(\mu - \lambda)} \quad (37)$$

Del mismo modo que conocemos no sólo L y L_q sino que conocemos la distribución de probabilidad de N y N_q (caracterizada por los p_n), también podría interesarnos conocer cómo se distribuyen W y W_q , es decir, cómo se distribuyen las siguientes variables aleatorias:

ω : "Tiempo de espera en el sistema en la distribución estacionaria".

ω_q : "Tiempo de espera en la cola en la distribución estacionaria".

Las funciones de distribución de estas variables, las denotaremos por $W(t)$ y $W_q(t)$, respectivamente, que están dadas por:

$$W(t) = P(\omega \leq t) = \int_0^t (\mu - \lambda) e^{-(\mu - \lambda)x} dx \quad (38)$$

Donde $(\mu - \lambda)e^{-(\mu - \lambda)x}$ es la función de densidad de una exponencial de parámetro $\mu - \lambda$.

Por tanto, $\omega \stackrel{d}{=} \exp(\mu - \lambda)$ (39)

Antes, al usar las fórmulas de Little, obtuvimos que $W = \frac{1}{\mu - \lambda}$ donde efectivamente, es la media de una distribución exponencial de parámetro $\mu - \lambda$.

En forma similar podríamos obtener la distribución para $W_q(t)$, descomponiendo de la siguiente forma:

$$W_q(t) = P(\omega_q \leq t) = P(\omega_q = 0) + P(0 < \omega_q \leq t). \quad (40)$$

donde, la probabilidad $P(\omega_q = 0)$ es la probabilidad de no tener que esperar nada en la cola o, lo que es lo mismo, $p_0 = 1 - \frac{\lambda}{\mu}$ y el segundo sumando se puede calcular replicando el desarrollo para $W(t)$. En este caso, obtendríamos una distribución de probabilidad discontinua en 0, cuyo salto valdría exactamente p_0 , esto es:

$$W_q(t) = \begin{cases} 1 - \frac{\lambda}{\mu} e^{-(\mu - \lambda)t} & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (41)$$

y el salto en 0 es $W_q(0) - W_q(0^-) = 1 - \frac{\lambda}{\mu} - 0 = p_0$ (42)

2.1.1.17. El modelo M/M/s

Veamos ahora el caso en el que el sistema de la cola tiene varios servidores para atender a los clientes ($s > 1$) (Karian & Dudewicz, 1999)

M/M/s

Tasa de llegadas $\rightarrow \lambda$

Tasa de servicio $\rightarrow \mu$

Número de servidores $\rightarrow s > 1$

Capacidad de la cola $\rightarrow \infty$

Población potencial $\rightarrow \infty$

Disciplina de la cola \rightarrow FIFO

Al pensar en el modelo M/M/s como un proceso de nacimiento y muerte, ya podemos observar una primera diferencia que hará su estudio algo más complicado que el caso M/M/1:

Tasas de llegada: Independientemente del número de gente que haya en nuestro sistema, la tasa de llegada λ es constante.

Tasas efectivas de servicio: Aunque todos los servidores tienen una tasa de servicio constante, dada por μ , la tasa efectiva de servicio del sistema sí que depende del número de clientes en el mismo. La tasa de salida de clientes será mayor cuando hay $s > 1$ servidores ocupados que cuando hay sólo 1 ocupado.

$$\mu_n = \begin{cases} n\mu & n = 1, 2, \dots, s \\ s\mu & n = s+1, s+2, \dots \end{cases} \quad (43)$$

La distribución del mínimo de s distribuciones exponenciales de parámetro μ será una distribución exponencial de parámetro $s\mu$. Por tanto, la tasa efectiva de servicio de nuestro sistema depende del nivel de ocupación del mismo.

Ahora debemos caracterizar la variable aleatoria N a través de las probabilidades p_0, p_1, \dots . Otra dificultad añadida sobre el modelo $M/M/1$ es que ahora ya no tenemos una fórmula explícita para p_0 (en el modelo $M/M/1$ sabíamos que $p_0 = 1 - \frac{\lambda}{\mu}$). En este caso, aplicaremos lo siguiente:

$$\begin{cases} p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} c_n} \\ p_n = c_n p_0 \end{cases} \quad \text{y} \quad c_n = \frac{\prod_{i=0}^{n-1} \lambda_i}{\prod_{i=1}^n \mu_i} = \begin{cases} \frac{\lambda^n}{n! \mu^n} & n = 1, 2, \dots, s \\ \frac{\lambda^n}{s! s^{n-s} \mu^n} & n = s+1, s+2, \dots \end{cases} \quad (44)$$

Un sistema con estas características sería estacionario si y sólo si $\rho = \frac{\lambda}{s\mu} < 1$. Sin embargo, en el argumento que usamos allí estábamos asumiendo implícitamente que la tasa de servicio sería $s\mu$ independientemente del número de clientes en el sistema. Intuitivamente, podemos observar que si $\frac{\lambda}{s\mu} < 1$, el sistema no crecerá indefinidamente. Siempre que haya más de s clientes en el sistema, la tasa efectiva de servicio será $s\mu > \lambda$, con lo que no podrá seguir

creciendo sin límite. Esta idea es precisamente lo que está detrás del

siguiente argumento que establece que $\sum_{n=1}^{\infty} c_n < \infty$, esto es:

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{s-1} \frac{\lambda^n}{n! \mu^n} + \sum_{n=s}^{\infty} \frac{\lambda^n}{s! s^{n-s} \mu^n} = \sum_{n=1}^{s-1} \frac{\lambda^n}{n! \mu^n} + \frac{\lambda^s}{s! \mu^s} \sum_{n=s}^{\infty} \rho^{n-s} \quad (45)$$

La convergencia de la progresión geométrica estará asegurada siempre

que $\rho < 1$. Además, en dicho caso, la suma de la progresión es $\frac{1}{1-\rho}$ (

no es $\frac{\rho}{1-\rho}$ porque el primer sumando de la progresión es $\rho^0 = 1$); por

tanto:

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

$$y \quad p_n = c_n p_0 = \begin{cases} \frac{\lambda^n}{n! \mu^n} p_0 & n = 1, 2, \dots, s \\ \frac{\lambda^s}{s! \mu^s} \rho^{n-s} p_0 & n = s+1, s+2, \dots \end{cases} \quad (46)$$

Observación. A la hora de implementar en un ordenador las fórmulas que veremos en este tema, hay que tener cierto cuidado. Los factoriales y las potencias crecen muy rápidamente y si no tenemos cuidado podríamos desbordar la capacidad del ordenador. La “recursividad” de las fórmulas de los modelos de colas permite ser bastante eficientes. Por ejemplo, al calcular $(s-1)!$ en la sumatoria del denominador, podemos obtenerlo como $(s-2)!(s-1)$, donde $(s-2)!$ se calculó en el sumando anterior. Este tipo de consideraciones

son importantes cuando se trabaja con sistemas de colas con un gran número de servidores. En este caso, la implementación podría hacerse de forma eficiente del siguiente modo:

$$p_0 = \frac{1}{\sum_{n=0}^{s-1} c_n + c_{\geq s}}, \text{ donde:}$$

$$c_0 = 1$$

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

$$c_{\geq s} := \sum_{n=s}^{\infty} c_n = \frac{\lambda^s}{(s-1)! \mu^{s-1} (s\mu - \lambda)} = c_{s-1} \frac{\lambda}{s\mu - \lambda} \quad (47)$$

Una observación análoga se aplica al cálculo del resto de p_n .

2.1.1.18. Series de tiempo

Una **serie temporal** o **cronológica** es una secuencia de datos, observaciones o valores, medidos en determinados momentos y ordenados cronológicamente. Los datos pueden estar espaciados a intervalos iguales (como la temperatura en un observatorio meteorológico en días sucesivos al mediodía) o desiguales (como el peso de una persona en sucesivas mediciones en el consultorio médico, la farmacia, etc.). Para el **análisis de las series temporales** se usan métodos que ayudan a interpretarlas y que permiten extraer información representativa sobre las relaciones subyacentes entre los datos de la serie o de diversas series y que permiten en diferente medida y con distinta confianza extrapolar o interpolar los datos y así predecir el comportamiento de la serie en

momentos no observados, sean en el futuro (extrapolación pronóstica), en el pasado (extrapolación retrógrada) o en momentos intermedios (interpolación)..

2.1.1.19. Procesos Estacionarios

Para definir la estacionariedad de un proceso estocástico puede utilizarse las funciones de distribución, o bien los momentos. En el primer caso se enuncia la propiedad de estacionariedad en sentido estricto o fuerte, mientras en el segundo caso se enuncia la propiedad de estacionariedad en sentido amplio o débil.

Se dice que un proceso estocástico es **estacionario en sentido estricto** cuando su función de distribución conjunta es invariante respecto de un desplazamiento en el tiempo, es decir, consideremos que t_1, t_2, \dots, t_n corresponden a periodos sucesivos que denominaremos como $t, t+1, \dots, t+k$, cuando:

$$F(x_t, x_{t+1}, \dots, x_{t+k}) = F(x_{t+m}, x_{t+1+m}, \dots, x_{t+k+m}), \forall t, k, m \quad (48)$$

Se dice que un proceso estocástico es estacionario en **sentido amplio o débil** cuando se derivan de (48), esto es:

Las esperanzas matemáticas de las variables aleatorias no dependen del tiempo; es decir son constantes:

$$E(x_t) = E(x_{t+m}) \quad \forall m$$

$$\mu_t = \mu \quad \forall t$$

2.1.1.20. Modelos estocásticos de Series temporales

El objetivo del análisis estocástico de series temporales es inferir a partir de una serie temporal las características de la estructura probabilística subyacente. De forma intuitiva se puede decir que se trata de inferir las propiedades de la población a partir de la muestra. En este sentido, se presentan casos particulares del proceso estocástico lineal discreto general, los modelos de medias móviles (MA), autorregresivos (AR) o mixtos (ARMA), que se definen como:

I. Modelo de medias móviles de orden q [MA(q)]:

Este modelo se puede escribir como:

$$x_t = \mu + \theta(L)\varepsilon_t \quad (49)$$

Donde $\theta(L)$ es el operador polinomial de retardos, definido como:

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q \quad (50)$$

Las características del modelo MA(q) son

- El modelo de medias móviles MA(q) es siempre estacionario
- Un modelo de medias móviles MA(q) es invertible; esto es que puede escribirse como un modelo AR(∞)
- Las autocovarianzas y consecuentemente las autocorrelaciones se anulan cuando el desfase temporal considerado es superior al orden del modelo.

II. Modelo autorregresivo de orden p [AR(p)]:

Este modelo se puede escribir como:

$$\phi(L)x_t = \delta + \varepsilon_t \quad (51)$$

Donde $\phi(L)$ es el operador polinomial de retardos, definido como:

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p \quad (52)$$

Las características del modelo AR(p) son:

- El modelo de medias móviles AR(p) es siempre invertible
- Un modelo de medias móviles AR(p) es estacionario; si las raíces de la ecuación (52) caen dentro del círculo unidad.
- En el modelo AR, a diferencia de lo que sucede con el MA presenta siempre una tendencia amortiguada a cero, sin llegar a anularse.

III. Modelo estocástico no estacionario homogéneo ARIMA

Un proceso estocástico no estacionario homogéneo de primer grado es:

$$x_t = x_{t-1} + \varepsilon_t \quad (53)$$

Este modelo es no estacionario, ya que (53) es un modelo AR(1) no estacionario, como consecuencia de que $\phi_1 = 1$ no cumple la condición de estacionariedad. Las autocovarianzas también varían en el tiempo, implicando coeficientes de auto correlación próximos a la unidad y con un decrecimiento muy lento conforme aumenta el lapso temporal k.

2.1.1.21. Sistemas, modelos y simulación.

Algunos problemas no se pueden resolver mediante métodos analíticos. Entre otras razones podrían citarse la existencia de patrones no normalizados de entrada y de servicio, una gran complejidad del sistema a modelar o la naturaleza de la disciplina de cola.

La simulación es el diseñar y desarrollar un modelo computarizado de un sistema o proceso y conducir experimentalmente con este modelo con el propósito de entender el comportamiento del sistema del mundo real o evaluar varias estrategias con los cuales puedan operar el sistema.

Cuando tenemos que estudiar un sistema de espera y éste resulta ser relativamente sencillo, debemos considerar las soluciones analíticas para dar respuesta a cuestiones sobre su funcionamiento. Sin embargo, muchos sistemas reales resultan ser lo suficientemente complejos como para permitir obtener un modelo matemático completo con el que conseguir los resultados analíticos para las medidas de eficiencia de interés. Es entonces cuando acudimos a la simulación con el fin de aproximar las soluciones. El punto de partida será un modelo matemático especificado para describir cómo se producen las llegadas y cómo los servicios, que tenga en cuenta siempre las características de funcionamiento del sistema.

La idea de la aproximación a través de simulación consiste en recrear el funcionamiento del sistema, una o varias veces, con el fin de captar su comportamiento medio, bien durante cierta franja de tiempo (estado transitorio), bien en el estado estacionario.

Simular un sistema de espera consiste en simular eventos discretos, describiendo la evolución del sistema a lo largo del tiempo a través de

las variables de estado que cambian instantáneamente en puntos temporales separados. El sistema sólo cambia de estado en un número contable de instantes de tiempo, a pesar de que el tiempo evoluciona de forma continua (Chatfield, 1975). Estos instantes temporales de cambio son aquellos en los que se da un suceso instantáneo de llegada o salida del sistema, que puede cambiar, aunque no necesariamente, el estado del sistema.

¿Qué intenta la simulación?

1. Descubrir el comportamiento de un sistema
2. Postular teorías o hipótesis que expliquen el comportamiento observado
3. Usar esas teorías para predecir el comportamiento futuro del sistema, es decir mirar los efectos que se producirían en el sistema mediante los cambios dentro de él o en su método de operación (tiempo en minutos)

Por lo general el modelo nos ayuda a entender y mejorar un sistema

2.1.1.22. Funciones del modelo

- Comparar
- Predecir

Propiedades de los modelos

1) Componentes:

Son las partes de un conjunto que forman el sistema

2) Variables:

Pueden ser de dos tipos (Exógenas, Endógenas)

- Exógenas: Las entradas son originadas por causas externas al sistema.
- Endógenas: Las entradas son producidas dentro del sistema que resultan de causas internas, las cuales pueden ser de Estado o de Salida

Estadísticamente a las variables exógenas se las denomina como variables independientes.

3) Parámetros:

Son cantidades a las cuales el operador del modelo puede asignarle valores arbitrarios lo cual se diferencia de las variables.

Los parámetros una vez establecidos se convierten en constantes.

4) Relaciones Funcionales:

Describen a los parámetros de tal manera que muestran su comportamiento dentro de un componente o entre componentes de un sistema.

Las relaciones funcionales pueden ser de tipo determinísticos o estocásticos.

- Determinísticos: Sus definiciones que relacionan ciertas variables o parámetros donde una salida del proceso es singularmente determinada por una entrada dada.
- Estocásticos: Cuando el proceso tiene una salida indefinida, para una entrada determinada las relaciones funcionales se

representan por ecuaciones matemáticas y salen del análisis estadístico matemático.

5) Restricciones:

Estas son limitaciones impuestas a valores de las variables las cuales pueden ser de dos formas:

- Autoimpuestas: O sea asignadas por el mismo operador o
- Impuestas: O sea cuando son asignadas manualmente por el mismo sistema

2.1.1.23. Ventajas y consideraciones de la simulación

Las ventajas básicas del uso de la simulación en la resolución de un sistema de espera son:

1. La mayoría de los sistemas de la vida real con elementos estocásticos no pueden ser descritos con precisión a través de un modelo matemático evaluable analíticamente. En esos casos, la simulación es el único modo de investigación posible.
2. Permite estudiar el comportamiento de un sistema bajo un conjunto de condiciones predeterminadas.
3. Facilitan la comparación de diseños alternativos para el funcionamiento de un sistema y elegir el que mejor logra los objetivos marcados para su funcionamiento.
4. Se mantiene mejor el control sobre las condiciones del sistema que cuando se realiza la observación in situ del mismo.
5. Permite delimitar la franja temporal de interés, y estudiar la evolución temporal a partir de ciertas condiciones de inicio.

Sin embargo, también tiene una serie de consideraciones a tener en cuenta:

1. Cada vez que repliquemos el comportamiento de un sistema, estamos involucrando al azar, con lo cual hay que tenerlo en cuenta y tratar de paliar su influencia sobre las conclusiones. Es decir, hay que echar mano de la inferencia estadística para estimar y controlar el error que podamos estar cometiendo.
2. Para controlar el error y garantizar cierta confiabilidad de las conclusiones, será preciso determinar a priori la longitud de las réplicas (franja temporal de observación del sistema, especialmente cuando se estudie el estado estacionario, pues será preciso asegurar la convergencia), así como el número de réplicas.
3. La validez de las conclusiones dependerá de la validez del modelo estadístico que hayamos propuesto para describir el comportamiento de las llegadas y de los servicios. Es preciso siempre, validar el modelo de generación de datos con observaciones disponibles.
4. Los tiempos computacionales constituyen un serio obstáculo en bastantes ocasiones.

2.1.1.24. Elementos de un Modelo de Simulación

Cuatro son los elementos a tener en cuenta al abordar un modelo de simulación de teoría de colas, supuesto diseñado el modelo “físico”:

- a) Selección de los datos de entrada
- b) Simulación.

c) Análisis de los resultados

d) Validación del modelo.

Dado que estamos interesados en modelizar sistemas estocásticos, los datos de entrada deben representar del modo más fiable posible la realidad. En ocasiones se usan datos extraídos de la misma para reconocer la estructura de los datos de entrada. La ejecución de la simulación actualmente se realiza mediante paquetes informáticos avanzados. El análisis de resultados tiene que ver con el cálculo de la efectividad del sistema mediante las técnicas estadísticas apropiadas. Además la validación del modelo es una exigencia que muchas veces se olvida al realizar modelos, que consiste en comprobar que el sistema reacciona como lo hace la realidad (Karian & Dudewicz, 1999)

2.1.1.25. Modelización de las Entradas

La Modelización de las Entradas es un requerimiento no sólo de la simulación, sino de cualquier tipo de análisis probabilístico y numérico.

Los dos mayores problemas en la modelización de los datos de entrada son la selección de la familia de distribuciones estadísticas y una vez estimada la familia estimar los parámetros que definen la función de las diferentes entradas.

El primero de los dos problemas es evidentemente el más complicado mientras que el segundo sólo es abordable una vez se ha resuelto la selección de la familia de distribuciones estadísticas.

En muchos casos los paquetes de simulación suelen llevar una herramienta de ajuste estadístico. Cuando esto no ocurra deberemos recurrir a las diferentes técnicas estadísticas para definir tanto las familias como los parámetros.

2.1.1.26. Análisis de Resultados

Alcanzar conclusiones válidas a partir de los resultados requiere un gran y cuidadoso esfuerzo. Cuando se simulan sistemas estocásticos (Sobczyk, 2014), no es posible extraer conclusiones a partir de una única simulación que por naturaleza es estadística.

Por tanto para obtener conclusiones es necesario diseñar y ejecutar experimentos de una manera lógica y comprensiva.

Existen dos tipos de modelos de simulación: continuos o interrumpidos.

Un modelo interrumpido simularía por ejemplo un banco que abre a las 8:00 y cierra a las 14:00, vaciando la cola al final del servicio. Sin embargo un modelo continuo se podría asociar a un sistema productivo donde el trabajo con el que se acaba un día, es con el que se comienza al día siguiente. En este último caso es cuando interesan los resultados en el estado estacionario.

En los sistemas que conducen a modelos interrumpidos el estado estacionario es generalmente irrelevante. Lo que importa es el valor medio calculable al recoger un cierto número de resultados, admitiendo siempre que lo que se obtiene es un valor medio estimado en un intervalo de confianza.

En los sistemas que conducen a modelos continuos el problema es un poco más complicado, porque hay que eliminar de las muestras el estado transitorio, aunque la definición de estado transitorio exigiría el reconocimiento del estado estable y por tanto del estado transitorio.

Los más importantes paquetes de simulación llevan incorporado herramientas que realizan estos ejercicios mediante los cuales se pueden calcular los anteriormente citados intervalos de confianza.

2.1.1.27. Validación del Modelo

La validación de los modelos es probablemente el paso más importante, y probablemente también el paso más obviado por aquellos que modelizan.

Antes de iniciar el proceso de realizar un modelo de simulación es necesario que el modelizador se familiarice con el sistema que tiene que estudiar. Para ello es necesario involucrar a todos los niveles al personal implicados en el proceso que va a ser simulado. En ese caso uno de los problemas que aparece es el exceso de detalles en el modelo que lo convierten en improductivo.

El primer y fundamental paso en la validación es verificar que el programa hace lo que está previsto que haga. Otro paso es definir el

grado de credibilidad, es decir hasta que punto los que van a usar el modelo consideran que el mismo tiene una utilidad y representa la realidad en la medida que nos interesa. Para ello es necesario que los objetivos del estudio, las medidas de rendimiento y el nivel de detalle deba pactarse y mantenerse en el nivel más simple posible (Pardo & Valdés, 1987).

Cuando sea posible, los resultados de las simulaciones se deben comprobar con la realidad. Si ésta no estuviera disponible habría que intentar reproducir modelos teóricos con soluciones conocidas mediante métodos analíticos.

2.1.1.28. Programación de la simulación

Con el fin de agilizar la generalmente costosa simulación, es preciso estructurar lo mejor posible el proceso y tener en cuenta una serie de elementos cruciales para la elaboración del código.

Elección del lenguaje de programación:

- Lenguajes de programación general: C, C++, Visual C, JAVA, Pascal, Fortran, R
- Lenguajes específicos de simulación: SIMULA, SIMSCRIPT, GPSS, PROMODEL, SLAM, AutoMod, ModL, SLX

Estado del sistema.- Debemos considerar la colección de variables necesarias para describir el estado del sistema en un instante concreto de tiempo.

Reloj de simulación.- Da el valor actual del instante temporal simulado. Hay dos aproximaciones para hacer avanzar el reloj de simulación: adelanto del tiempo del próximo suceso y adelanto de tiempo a incrementos fijos. La primera aproximación es la utilizada por la mayoría de software de simulación y por la mayoría de la gente que programa con un lenguaje general (Taha & González Pozo, 2004)

Lista de sucesos.- Una lista que contiene información sobre los sucesos futuros y los instantes en que ocurrirán.

Contadores estadísticos.- Variables utilizadas para almacenar la información auxiliar para la medición de las magnitudes de interés.

La estructura del programa de simulación ha de constar de los siguientes pasos:

- 1. Rutina de inicialización.-** Un subprograma para inicializar el modelo de simulación en el instante 0.
- 2. Rutina de temporización.-** Un subprograma que determina el próximo suceso de la lista de sucesos y después adelanta el reloj de simulación al instante en que dicho suceso va a ocurrir.
- 3. Rutina de estado.-** Un subprograma que actualiza el estado del sistema cuando ocurre cualquier tipo de suceso y actualiza los contadores estadísticos.

4. **Rutina de librerías.-** Un conjunto de subprogramas utilizado para generar observaciones aleatorias de las distribuciones de probabilidad especificadas en el modelo de simulación.
5. **Generador de informe.-** Un subprograma que calcula estimaciones de las medidas de eficiencia deseadas y produce un informe cuando la simulación finaliza.
6. **Programa principal.-** Que engloba todas las rutinas anteriores: Inicialización, búsqueda del próximo suceso, actualización del estado del sistema. El programa principal puede también realizar el chequeo para la terminación de la simulación y hacer un llamado al generador de informes cuando la simulación terminó.

2.1.1.29. Simulando con PROMODEL

Descripción.

ProModel es un programa de simulación para computadora que permite simular cualquier tipo de sistemas de manufactura, sistemas logísticos, manejo de materiales, servicios, entre otros. Es uno de varios software de simulación más usados en el mercado. Cuenta con herramientas de análisis y diseños que unidas a la animación de los modelos bajo estudio, permiten al usuario conocer mejor el problema y alcanzar mejores resultados, siendo más confiables respecto de las decisiones a tomar (García Dunna, García Reyes, & Cárdenas Barrón, 2010).

Este programa nos permite trabajar con procesos de uno o varios productos, líneas de ensamble y de transformación, entre otros. En teoría, cualquier sistema de procesos puede ser modelado en un computador; solo hay que saber plantear el modelo. ProModel nos permite simular una gran cantidad de situaciones como: “Justo a Tiempo”, “Teoría de Restricciones”, logística, colas y muchas otras más. Además de permitir el simulado de acciones, nos permite optimizar los procesos, y así obtener los mejores resultados con el consumo mínimo de recursos; para dicha tarea, el sistema cuenta con 2 optimizadores (García Dunna, García Reyes, & Cárdenas Barrón, 2013).

Para comenzar a utilizar ProModel es necesario conocer las cuatro principales funciones del Software, las cuales son mencionadas a continuación:

1.- Entities (Entidades): Es la materia que será procesada en la simulación del sistema, es decir, son aquellas personas, partes, insumos, documentos, productos, etc. que ingresan al sistema para ser transformados en productos finales o clientes atendidos. Estas entidades son altamente dinámicas, ya que pasan de una estación de servicio o máquina, a otra.

2.- Locations (Locaciones): Representan las máquinas o personas que atienden, procesan, transforman, etc. a las entidades. Estas son estáticas dentro del sistema, ya que es de esperarse que

una máquina no se mueva de un lugar a otro dentro de un proceso.

3.- Processing (Procesos): Define la forma cómo se moverán las entidades entre las locaciones, más aún, se encarga de proveer las reglas que determinan cómo procesará cada máquina una entidad y el tiempo de ese procesamiento.

4.- Arrivals (Llegadas): Define cómo será alimentado el sistema con entidades; es decir define parámetros tales como la cantidad, tipo, frecuencia y lugar de arribo de las entidades. Cabe destacar que además el Software nos da la opción de agregar personal de mantenimiento, electricidad, gas, etc. Además a cada operación nos da la posibilidad de agregar costos de tal modo que se puede determinar el tiempo y el costo de un producto terminado.

Los especificadores técnicos, guías, y soporte asociado se encuentran publicados en el portal de ProModel Corporation (ProModel Corporation, s. f.)

Características Técnicas:

La presente guía de utilización es de autoría propia, según las especificaciones técnicas asociadas al proyecto de investigación. El software utilizado fue el ProModel 2016 con el fin de simular el modelo.

El modelo utiliza una distribución exponencial para definir la distribución de tiempo de llegada de los clientes.

El modelo utiliza una distribución Poisson, uniforme, exponencial para definir distribución de tiempo en la atención de los clientes.

2.2. Marco Conceptual

Simulación: Una definición más formal, formulada por (R. E. Shannon), es: "La simulación es el proceso de diseñar un modelo de un sistema real y llevar a término experiencias con él, con la finalidad de comprender el comportamiento del sistema o evaluar nuevas estrategias -dentro de los límites impuestos por un cierto criterio o un conjunto de ellos - para el funcionamiento del sistema". En las ciencias, la simulación es el artificio contextual que referencia la investigación de una hipótesis de trabajo utilizando modelos

Sistemas dinámicos: Es un sistema cuyo estado evoluciona con el tiempo. El comportamiento en dicho estado se puede caracterizar determinando los límites del sistema, los elementos y sus relaciones. (Wikipedia)

Líneas de espera: Una línea de espera es el efecto resultante en un sistema cuando la demanda de un servicio supera la capacidad de proporcionar dicho servicio. Este sistema está formado por un conjunto de entidades en paralelo que proporcionan un servicio a las transacciones que aleatoriamente entran al sistema. (Victoria Cauch Ucan)

Teoría de colas: La teoría de colas es el estudio matemático de las colas o líneas de espera dentro de un sistema. Esta teoría estudia factores como el tiempo de espera medio en las colas o la capacidad de trabajo del sistema sin que llegue a colapsar. (Victoria Cauich Ucan)

2.3. Antecedentes

La teoría de colas surgió alrededor de 1878 para analizar la congestión de tráfico telefónico con el objeto de cumplir la demanda del servicio en el sistema telefónico de Copenhague. A partir de la fecha se ha utilizado la teoría de colas en múltiples aplicaciones.

Un primer estudio basados en el algoritmo de R. Larson, desarrollan un algoritmo para reducir las longitudes de colas transitorias, así como tiempos de espera de cada cliente en el periodo ocupado. También desarrollan un algoritmo de línea para actualizar dinámicamente las estimaciones actuales de la longitud de la coda después de cada salida. Además, generalizan el algoritmo para el caso de un proceso de Poisson que varía en el tiempo. Informando que los resultados computaciones de la velocidad y la precisiones de los algoritmos pueden ser optimizados computacionalmente basado en procesos (Bertsimas & Servi, 1992).

Simular un sistema (teoría de colas) consiste en simular eventos discretos, haciendo que el sistema evolucione a lo largo del tiempo a través de las variables de estado que cambian consecutivamente. El sistema no cambia

de estado a pesar que el tiempo evoluciona permanentemente (Martínez et al., 2007)

También Clemente (2007) analiza el tema de las colas originadas en las oficinas de una entidad bancaria producto de la configuración propia del sistema encargado de administrarlas, con el fin de optimizar el tiempo de espera de los clientes. Se concentra en las colas generadas únicamente en las ventanillas. Para ello recopiló la información necesaria de la base de datos del sistema administrador de colas y diseño un modelo que replicase la situación actual mediante simulación. El programa aquí utilizado fue ARENA 9.0. Mostrando el funcionamiento actual de las colas en el banco, así como los parámetros necesarios para su correcta administración. Concluyendo que el modelo de simulación se ajustó a la situación actual de la empresa en forma adecuada (Clemente, 2008)

Respecto a la optimización Monzón (2011), con la finalidad de optimizar las facilidades del servicio en una entidad bancaria, y calibrando el modelo basado en teoría de colas utilizando el programa ARENA. Obteniendo que los resultados del trabajo permiten establecer una metodología factible y aplicable a distintos servicios basados en colas. Sus resultados sugieren que el concepto de servicio y estrategia para prever un servicio con facilidades, permiten diferenciarse de la competencia; y que toda empresa necesita someterse a un proceso de mejora continua de sus niveles competitivos (Monzón, 2011).

Según Madeiro (20012), una cola o Línea de espera se forma cuando un conjunto de entidades como personas, productos, etc. Demandan al mismo

tiempo un producto o servicio en un momento dado, excediendo su capacidad para abastecer a toda la cola en ese instante. Además, en su trabajo menciona que los objetivos de la teoría de colas son: (1) Identificar el nivel óptimo de un sistema que minimice el costo global y no colapsar, y (2) evaluar el impacto del sistema (líneas de espera) a distintas alternativas considerando un óptimo servicio. (Madeiro, 2012).

Así Arista (2016), con el objeto de dar solución al problema de optimización del número de cajeros en ventanilla que minimice los costos de espera y los de servicios en organización BCP. Utiliza el programa SIMIO para simular la optimización del número de cajeros en las ventanillas. Encontrando evidencias que el costo óptimo en un día fue de $8 \times S/.37.0864 = 296.7$ soles. Este costo esperado es la suma de los costos esperados de servir más los costos esperados por el tiempo que pasa un cliente en el sistema de espera. Además, el costo para cinco cajeros es menor comparando con cuatro y seis cajeros (Arista, 2016).

Otro antecedente es de Bauzoun & Muñoz (2016) en su estudio muestran que el servicio de atención bancaria ha experimentado cambios en los procesos de atención, siendo la principal causa "El avance tecnológico". Con el objetivo de proponer una mejora al sistema de atención de los usuarios de ventanillas preferencial del Banco de la Nación de la agencia "3" del distrito de José Leonardo Ortiz en la ciudad de Chiclayo. Utilizando la teoría de colas y la implementación en el simulador SIMIO. Encontrando evidencia que los usuarios percibían 2.54 en términos nominales de calificación, considerándose como "de acuerdo con la calidad". Mientras que

la cantidad de ventanillas (4 ventanillas) que son necesarias para atender es el menos al 60 por ciento de lo usuarios (Bazoun & Muñoz, 2016)

Por último, Urcia & Valera (2017), en su trabajo de investigación dirigido a determinar la relación directa entre los tiempos de espera para la atención en ventanilla y la calidad de servicio del área de operaciones del BCP. Consideraron un tamaño de muestra de 122 clientes de una población de 604, y como instrumento de recolección de información se utilizó encuestas con 22 preguntas. Utilizan el modelo SERVQUAL para encontrar la relación que existe entre la calidad de servicio y la atención en ventanillas. Este modelo fue calibrado utilizando los datos de las encuestas basados en 22 items relacionados con: elementos tangibles, fiabilidad, capacidad de respuesta, seguridad y empatía. Concluyendo que existe una correlación positiva moderada entre las variables de estudio con un $r = 0.611$. Además, especifica que 89.3 por ciento de clientes señalan que rara vez existe rapidez en el tiempo de espera, presentando en desacuerdo por la calidad de servicio. Por lo que, se acepta la hipótesis alterna, que, la dimensión rapidez en el tiempo de espera si influye en la calidad de servicio del Banco de Crédito de la oficina de El Porvenir en el 2017 (Urcia & Varela, 2017).

CAPÍTULO III

HIPOTESIS

3.1. Hipótesis

a. Hipótesis general

Un sistema con dos servidores obtenidos mediante la simulación para las líneas de espera aplicadas al tráfico de las ventanillas de una agencia bancaria presenta optimiza el tiempo de atención a los clientes

b. Hipótesis específicas

1. El comportamiento de los tiempos entre llegadas de los clientes en las ventanillas del banco se distribuyen en forma exponencial.
2. El modelo de teoría de colas que se ajusta al sistema de atención en el tráfico de las ventanillas del banco es $M/M/2$

3.2. Identificación de variables

En el conjunto de variables asociadas a un sistema de colas, se consideran las siguientes variables:

- Tráfico en la ventanilla de un banco

3.3. Operacionalización de variables

VARIABLE NOMINAL	Definición conceptual	Definición operacional	Indicadores
Tráfico en las Ventanillas de un banco	Conjunto de personas que asisten a una ventanilla de un banco para realizar transacciones financieras, pagos de servicios, depósitos, etc.	<p>Indica el número de clientes en la cola más en los servidores</p> <p>Indica el número de clientes que esperan a ser atendidos</p> <p>Tiempo de llegada entre un cliente y otro</p> <p>Tiempo en el que el cliente es atendido en ventanilla</p> <p>Tiempo en el que el cliente espera para ser atendido</p>	<p>Número de clientes en el sistema</p> <p>Número de clientes en la cola</p> <p>Tiempo de llegada</p> <p>Tiempo de servicio</p> <p>Tiempo de espera</p>

CAPÍTULO IV

METODOLOGÍA

4.1. Ámbito de estudio

El lugar de estudio se realiza en una agencia bancaria ubicada al interior de la ciudad Cusco.

4.2. Tipo y Nivel de Investigación

La presente investigación es de tipo aplicado cuyo propósito es solucionar un problema real y/o aplicado a fenómenos existentes (Hernández, Fernández, & Baptista, 2014). La metodología considera un nivel descriptivo. De tal manera que se pretende, especificar propiedades, características, y rasgos importantes del fenómeno de estudio (Bernal, 2010).

4.3. Unidad de Análisis

Cliente que llega a ser atendido en las ventanillas de la agencia bancaria.

4.4. Población de estudio

La población está conformada por todos los clientes que llegan a ser atendidos en las ventanillas de la agencia bancaria y realizan una transacción financiera.

4.5. Tamaño de muestra

Para el presente estudio se realizó un muestreo no probabilístico por conveniencia, porque se cuenta con todos los usuarios que ingresan al interior de la agencia bancaria y realizan una transacción financiera. La muestra considerada es:

Día de la semana	N° de clientes
Lunes	337
Martes	332
Miércoles	254
Jueves	251
Viernes	312
Tamaño de la muestra	n = 1486

Fuente: propia

4.6. Técnicas de recolección de muestra

La técnica de recolección de muestra fue por observación, es decir, se registró a todas las personas que ingresaban a la entidad bancaria, recolectando la información cuantitativa.

4.7. Técnicas de recolección de información

4.7.1. Recolección de información:

Se elaboró y se aplicó un instrumento para registrar los datos de campo a todas las personas que ingresaban a la entidad bancaria, este instrumento consistía en la observación directa de la variable

4.8. Técnicas de análisis e interpretación de la información

El estudio tiene varias fases, cada una con una metodología distinta de trabajo que ha de ser planificada. Primeramente aplicaremos la estadística descriptiva para observar los resultados obtenidos de la información recolectada, luego la estadística inferencial para tomar decisiones sobre los resultados obtenidos.

Una vez que se disponga de los datos, se hará un análisis de las series que nos permita caracterizar mediante modelos algunas de las variables involucradas en el problema (tiempo entre llegadas, tiempos de servicios); este análisis se hará utilizando el Software libre R (V 3.5.1);

Por último, una vez modeladas las variables y hechas las hipótesis apropiadas se desarrollará un modelo de teoría de colas que permita estudiar la eficiencia del sistema, así como variaciones en ésta ante cambios en los parámetros de entrada aplicando una simulación con el Software libre ProModel (V 2016).

4.9. Técnicas para demostrar la verdad o falsedad de hipótesis planteadas

Se utilizará las pruebas de bondad de ajuste de Kolmogorov – Smirnof, para observar la significancia del ajuste al modelo empírico y del modelo teórico.

CAPÍTULO V

RESULTADOS

5.1. Procesamiento, análisis, interpretación y discusión de resultados.

5.1.1. Análisis de datos estadísticos descriptivos.

El resultado obtenido al tomar la muestra de los clientes atendidos en el periodo del 09 al 13 de octubre del 2017 en las ventanillas de una agencia de un banco de la ciudad del Cusco fueron las siguientes:

Tabla 3.
Análisis descriptivos de datos

Día de la semana	Tiempo de Llegada TLL	Tiempo de Servicio TA	Tiempo de Espera TE
Lunes	$\bar{X} = 76.99$	$\bar{X} = 77.03$	$\bar{X} = 1215.41$
Martes	$\bar{X} = 75.27$	$\bar{X} = 80.56$	$\bar{X} = 1138.30$
Miércoles	$\bar{X} = 97.71$	$\bar{X} = 109.61$	$\bar{X} = 373.29$
Jueves	$\bar{X} = 100.02$	$\bar{X} = 101.31$	$\bar{X} = 860.71$
Viernes	$\bar{X} = 85.56$	$\bar{X} = 103.2$	$\bar{X} = 706.70$

Fuente: Resultados obtenidos en este estudio

Especificación de variables:

TLL: Tiempo de llegada al sistema entre un cliente y otro (en segundos)

TA: Tiempo de servicio de un cliente en el servidor (en segundos)

TE: Tiempo de espera de un cliente en la cola (en segundos)

Interpretación: En la tabla 3 podemos observar que los días donde la media de los tiempos de llegada resultan aproximados son los lunes, martes y viernes, por lo que podemos indicar que formarán un posible grupo homogéneo, mientras que los días miércoles y jueves menos afluencia de clientes que formarán otro posible grupo homogéneo.

Realizamos ahora la prueba de homogeneidad aplicando el software R, tomando las series de dos en dos para el tiempo de llegada (**TLL**) en los 5 días de la semana, aplicando una diferencia de medias, obteniéndose los siguientes resultados:

1. Comparación entre las series de los tiempos de llegadas de los días lunes y martes.

Planteamos las hipótesis nula y alterna, donde μ_x es la media de TLL1

y μ_y es la media de TLL2

Ho: $\mu_x = \mu_y$

Ha: $\mu_x \neq \mu_y$

$\alpha = 0.05$

```
> t.test(TLL1,TLL2)
Welch Two Sample t-test
data:  TLL1 and TLL2
t = 0.21585, df = 636.01, p-value = 0.8292
95 percent confidence interval:
-13.87889  17.30687
sample estimates:
mean of x   mean of y
76.98507    75.27108
```

Como $p = 0.8292 > 0.05$, aceptamos la hipótesis nula con un nivel de confianza del 95%, que indica que las medias $\mu_x = \mu_y$, es decir que

sus grupos son homogéneos, por lo que los tiempos de llegadas de lunes y martes tienen el mismo comportamiento.

2. Comparación entre las series de los tiempos de llegadas de los días lunes y miércoles.

Planteamos las hipótesis nula y alterna, donde μ_x es la media de TLL1 y μ_y es la media de TLL3

Ho: $\mu_x = \mu_y$

Ha: $\mu_x \neq \mu_y$

$\alpha = 0.05$

```
> t.test(TLL1, TLL3)
Welch Two Sample t-test
data:  TLL1 and TLL3
t = -2.0671, df = 518.49, p-value = 0.03922
95 percent confidence interval:
-40.418427  -1.027766
sample estimates:
mean of x mean of y
76.98507  97.70817
```

Interpretación: En el script de salida del programa R se observa que el p-valor = 0.03922 < 0.05, rechazamos la hipótesis nula con un nivel de confianza del 95%, que indica que las medias $\mu_x \neq \mu_y$, es decir que sus grupos no son homogéneos, por lo que los tiempos de llegadas de lunes y miércoles no tienen el mismo comportamiento.

Las comparaciones de las demás series de datos se completan en el ANEXO 2 de éste informe; desarrollando una tabla resumen de las comparaciones correspondientes a las medias de las 5 series para los tiempos de llegada, determinando el **p-valor**, se obtiene:

Tabla 4.
Resultados de las comparaciones para los tiempos de llegada

	TLL1	TLL2	TLL3	TLL4	TLL5
TLL1		-	-	-	-
TLL2	0.8292		-	-	-
TLL3	0.0392	0.0164		-	-
TLL4	0.0104	0.0026	0.8205		-
TLL5	0.6733	0.4831	0.0745	0.0223	

Fuente: Resultados obtenidos en este estudio

Especificación de variables:

TLL1: Tiempo de llegada al sistema entre un cliente y otro (día Lunes)

TLL2: Tiempo de llegada al sistema entre un cliente y otro (día Martes)

TLL3: Tiempo de llegada al sistema entre un cliente y otro (día Miércoles)

TLL4: Tiempo de llegada al sistema entre un cliente y otro (día Jueves)

TLL5: Tiempo de llegada al sistema entre un cliente y otro (día Viernes)

Interpretación: De la tabla 4, tomando en cuenta estos resultados, podemos realizar la separación de la serie en dos grupos homogéneos, lo cual corrobora al resultado observado en el primer análisis descriptivo, estos grupos serán:

Grupo N° 1: TLL1, TLL2, TLL5

Grupo N° 2: TLL3, TLL4

Ahora realizamos la prueba de homogeneidad de los tiempos de servicio (TA) tomados de dos en dos en los 5 días de la semana, aplicando el programa R-Studio, obteniendo los siguientes resultados:

3. Comparación entre las series de los tiempos de servicio de los días lunes y martes.

Planteamos las hipótesis nula y alterna, donde μ_x es la media de TA1 y μ_y es la media de TA2

$$H_0: \mu_x = \mu_y$$

$$H_a: \mu_x \neq \mu_y$$

$$\alpha = 0.05$$

```
> t.test(TA1,TA2)
Welch Two Sample t-test
data: TA1 and TA2
t = -0.70668, df = 664.97, p-value = 0.4800
95 percent confidence interval:
 -13.35099  6.28424
sample estimates:
mean of x mean of y
 77.02687  80.56024
```

Interpretación: Como el p-valor = 0.48 > 0.05 aceptamos la hipótesis nula, es decir a un nivel de confianza del 95% observamos que las series que corresponden a los días lunes y martes son homogéneas.

4. Comparación entre las series de los tiempos de servicio de los días lunes y viernes.

Planteamos las hipótesis nula y alterna, donde μ_x es la media de TA1 y μ_y es la media de TA5

$$H_0: \mu_x = \mu_y$$

$$H_a: \mu_x \neq \mu_y$$

$$\alpha = 0.05$$

```

> t.test(TA1,TA5)

Welch Two Sample t-test
data: TA1 and TA5
t = -2.2049, df = 616.07, p-value = 0.0278
95 percent confidence interval:
-18.734610 -1.083454
sample estimates:
mean of x mean of y
77.02687 86.93590

```

Interpretación: Como el p-valor = 0.48 > 0.05 aceptamos la hipótesis nula, es decir a un nivel de confianza del 95% observamos que las series que corresponden a los días lunes y viernes son homogéneos.

Las comparaciones de las demás series de datos se completan en el ANEXO 3 de éste informe. Desarrollando una tabla resumen de las comparaciones correspondientes a las medias de las 5 series referidas a los tiempos de servicio, determinando el **p-valor**, se obtiene:

Tabla 5.

Resumen de las comparaciones para los tiempos de servicio

	TA1	TA2	TA3	TA4	TA5
TA1		-	-	-	-
TA2	0.4800		-	-	-
TA3	3.695e-06	3.348e-05		-	-
TA4	1.913e-06	4.072e-05	0.2339		-
TA5	0.0278	0.1535	0.0006	0.0015	

Fuente: Resultados obtenidos en este estudio

Especificación de variables:

TA1: Tiempo de servicio de un cliente en ventanilla (día Lunes)

TA2: Tiempo de servicio de un cliente en ventanilla (día Martes)

TA3: Tiempo de servicio de un cliente en ventanilla (día Miércoles)

TA4: Tiempo de servicio de un cliente en ventanilla (día Jueves)

TA5: Tiempo de servicio de un cliente en ventanilla (día Viernes)

Interpretación: De los resultados de la tabla 5, observamos que hay homogeneidad en las series para TA1 con TA2, TA2 con TA5 y TA3 con TA4, por lo demás no encontramos homogeneidad en el tiempo de servicio debido a que el personal de atención puede variar de un día a otro o de turno mañana a turno tarde, por lo que para realizar la simulación optaremos por tomar el tiempo promedio de las 5 series; sin embargo será necesario encontrar el tipo de distribución de las series para cada día.

5.1.2. Determinación del tipo de distribución para cada serie

Primero realizaremos una prueba de series temporales para ver el comportamiento del número de clientes que llegan al banco, para esto realizamos el conteo del número de clientes por intervalos de 5 minutos durante el periodo de 7 horas que es el tiempo de atención al banco, obteniendo en total 84 periodos en cada día, obteniendo los siguientes resultados:

Tabla 6.
Número promedio de clientes que llegan al banco en intervalos de 5 minutos

	Lunes	Martes	Miércoles	Jueves	Viernes
Número de Clientes por día	335	332	254	251	312
Número de clientes que llegan cada 5 minutos (promedio)	3.99	3.95	3.02	2.99	3.71

Fuente: Resultados obtenidos en este estudio

Interpretación: En esta tabla observamos que los días Lunes, Martes y Viernes tienen valores promedios aproximados, es decir que no hay diferencia de medias entre las tres series de datos, mientras que los días miércoles y jueves también tienen valores medios próximos.

Para corroborar estos resultados, obtenemos los diagramas de cajas para ver la curtosis, simetría y sesgo de los datos para cada día distribuidos en intervalos de tiempo de 5 minutos.

En la figura 6 realizamos la comparación de los diagramas de cajas para las 5 series de datos donde podemos observar que los días lunes, martes y viernes tienen características similares, son asimétricas a la derecha y un tanto leptocúrticas, puesto que la concentración de los datos está entre 2 y 6 clientes que llegan cada 5 minutos, mientras que los días miércoles y jueves forman otro grupo homogéneo por tener similares características como asimetría a la derecha y la concentración de los datos está entre 1 y 4 clientes que llegan cada 5 minutos.

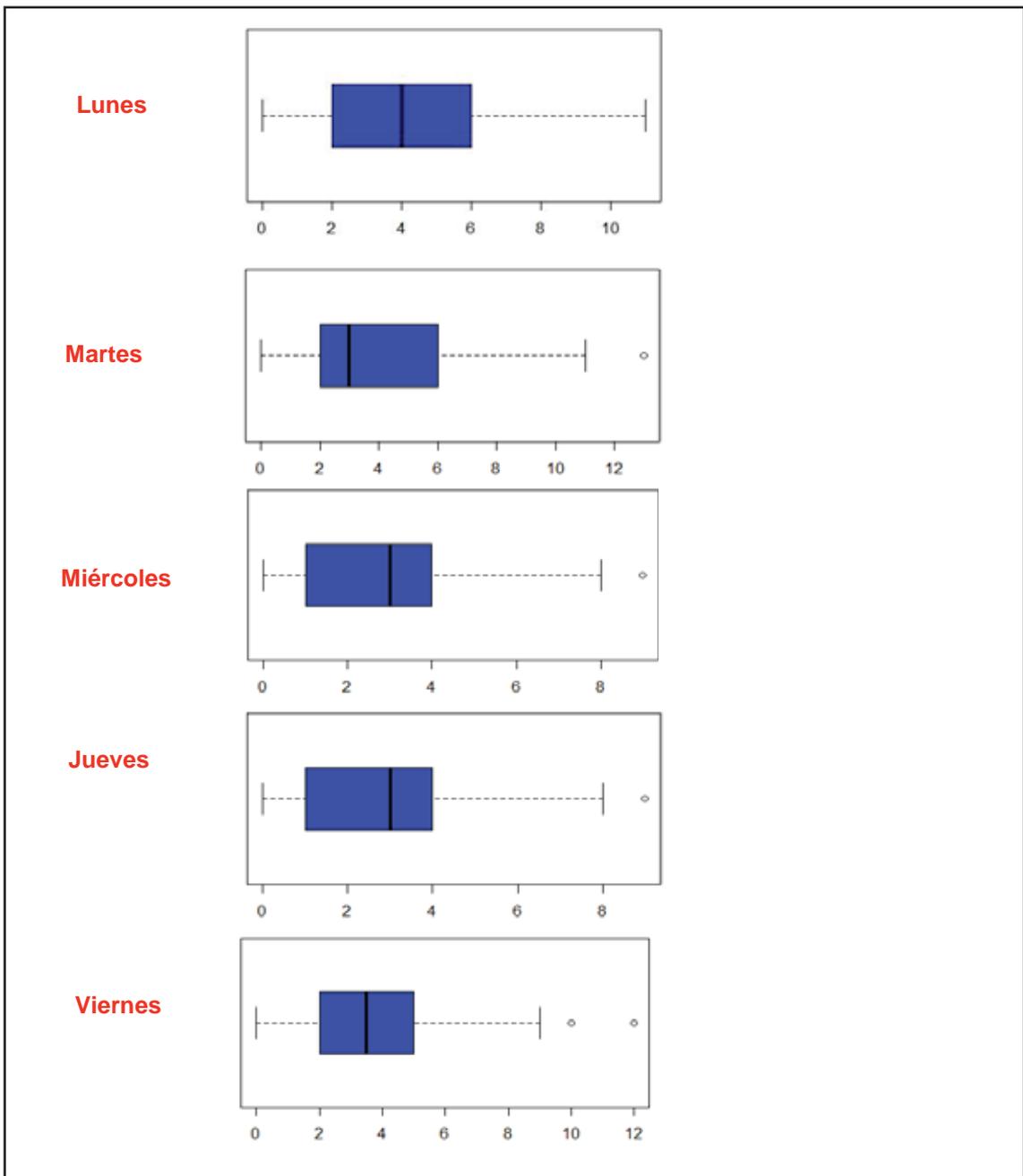


Figura 6. Diagrama de cajas en intervalos de 5 minutos de los datos de 5 días de la semana Nota: Resultados obtenidos en R

Realizaremos ahora las gráficas de dispersión para ver el comportamiento de las series durante las horas del día (7 horas) teniendo una serie de 84 intervalos de 5 minutos considerados desde

las 9am a 1pm y de 3pm a 6pm, tiempo en el que el banco presta atención a los clientes, determinaremos cuáles son las horas de mayor afluencia de clientes denominados (horas punta) y las horas de menor afluencia de clientes, donde los servidores no estarán saturados, esto es:

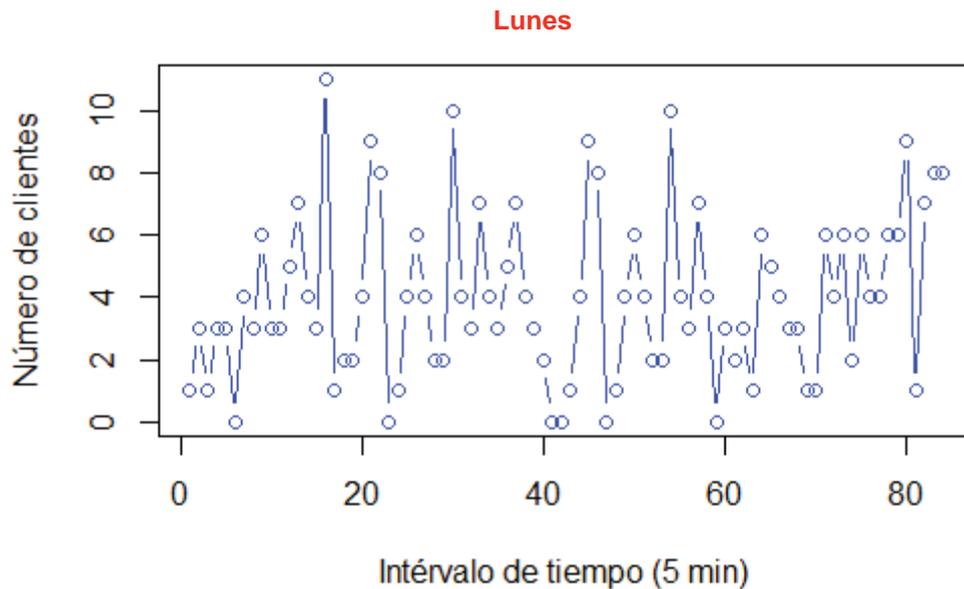


Figura 7. Dispersión de la serie de datos para el día lunes.

Nota: Resultados obtenidos en este estudio aplicando el programa R

Interpretación: Observamos que hay mayor afluencia de clientes entre las 10 y 11 de la mañana luego en el periodo de [40 – 60] que es las 12m y 1pm, así como también en la última hora del día es decir de 5 a 6pm. También es posible observar que las horas de menor afluencia de clientes son entre las 9am y 10am y por la tarde entre las 4pm y 5pm.

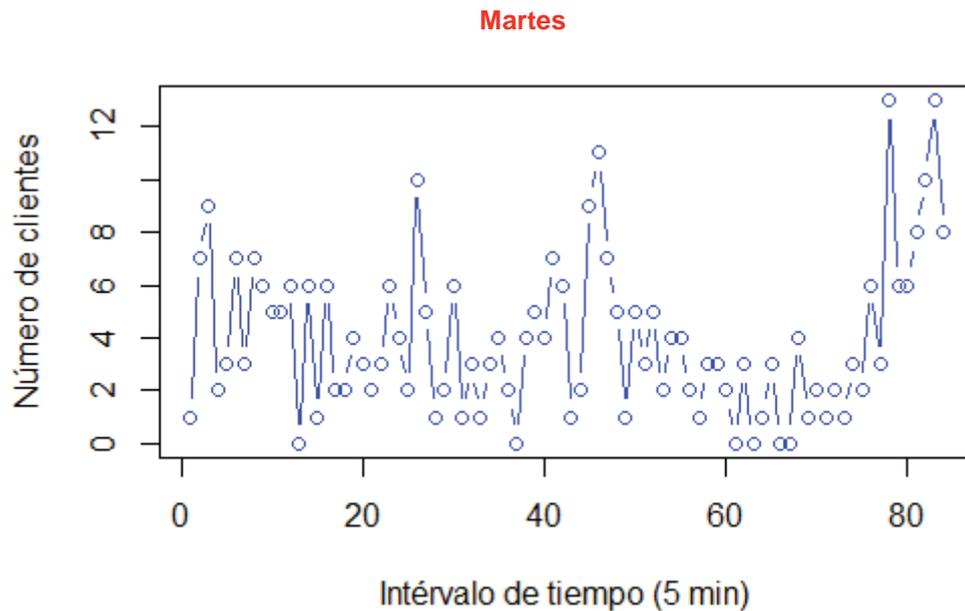


Figura 8. Dispersión de la serie de datos para el día martes.

Nota: Resultados obtenidos en este estudio aplicando el programa R

Interpretación: Observamos que hay mayor afluencia de clientes después del periodo [40-50] es decir entre las 12 y 1pm, así como también en la última hora del día es decir de 5 a 6pm. La menor afluencia de clientes es entre las primeras horas de la tarde, es decir entre las 3pm y 5pm.

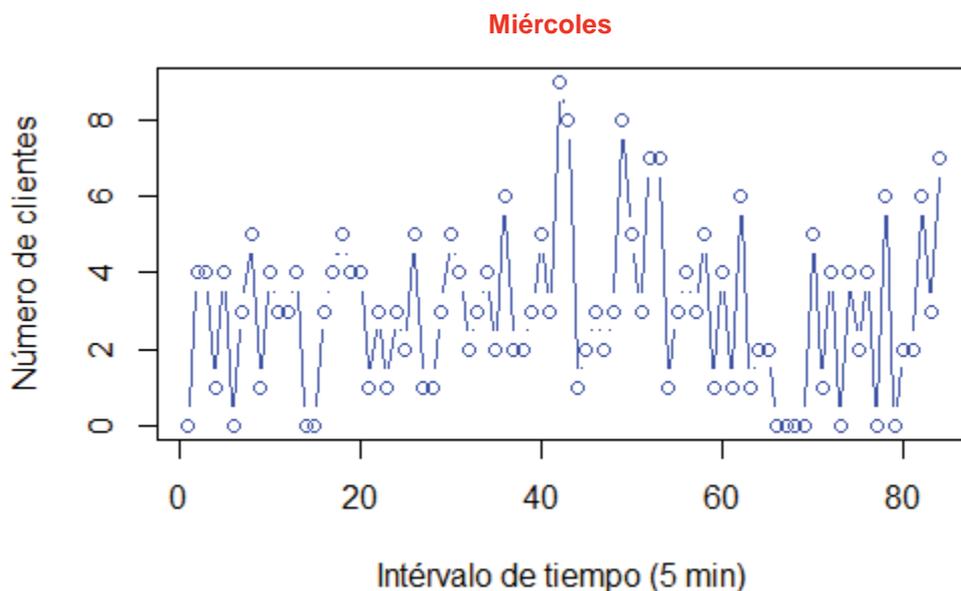


Figura 9. Dispersión de la serie de datos para el día miércoles

Nota: Resultados obtenidos en este estudio aplicando el programa R

Interpretación: Observamos que al inicio del día hay una regular afluencia de clientes al banco pero al promediar las 12 y 1pm hay gran afluencia de los clientes, así como también en las primeras horas de la tarde es decir entre las 3pm y 4pm además en la última hora del día es decir de 5 a 6pm. La menor afluencia de clientes es 9am las primeras horas de la mañana, es decir entre las 9am y 12m.

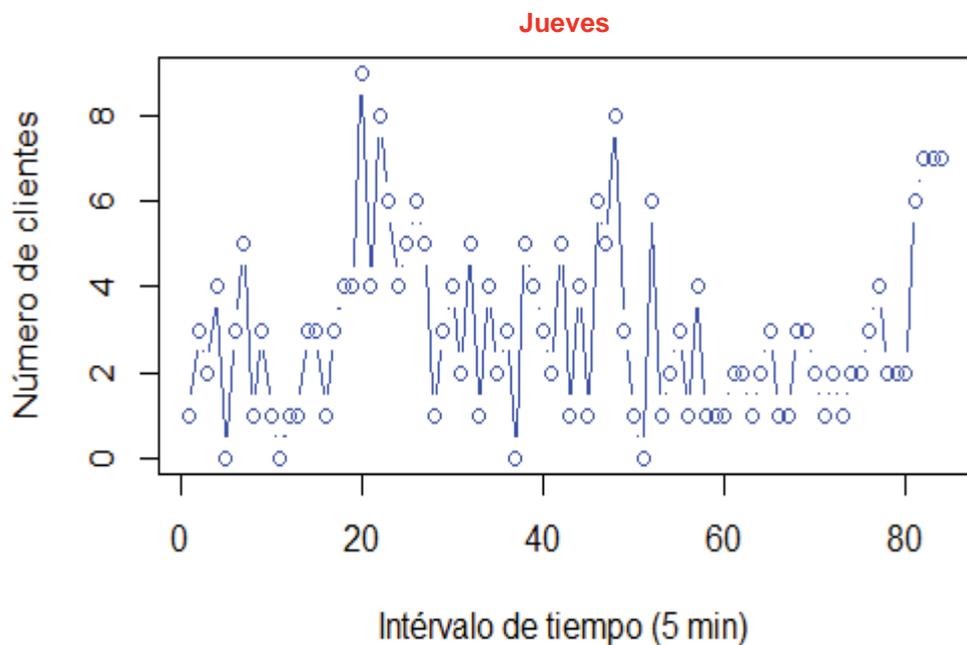


Figura 10. Dispersión de la serie de datos para el día jueves

Nota: Resultados obtenidos en este estudio aplicando el programa R

Interpretación: Observamos que hay mayor afluencia de clientes en el periodo [20-40] es decir entre las 10:40am y 12m, también entre las 12:30pm a 1pm, así como también en la última hora del día es decir de 5 a 6pm. La menor afluencia de clientes es entre las primeras horas de la mañana 9 a 10am y las primeras horas de la tarde, es decir entre las 3pm y 5pm.

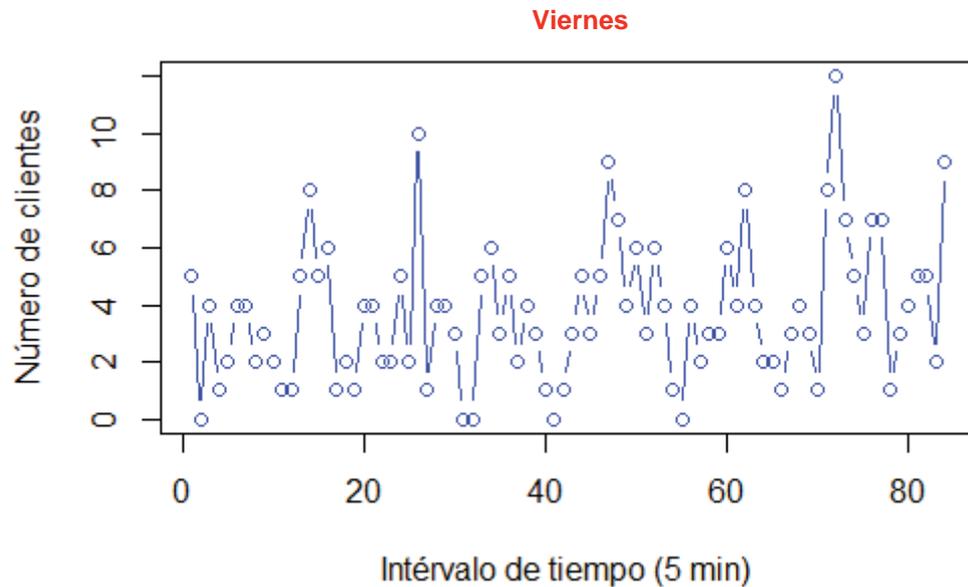


Figura 11. Dispersión de la serie de datos para el día viernes

Nota: Resultados obtenidos en este estudio aplicando el programa R

Interpretación: Observamos que hay mayor afluencia de clientes entre las 11 y 12m, así como también en las últimas horas de la tarde es decir de 4 a 6pm. La menor afluencia de clientes es entre las primeras horas de la mañana y al medio día

Determinaremos ahora la distribución de probabilidad que se ajuste más a las series de datos para cada uno de los 5 días de los tiempos de llegada aplicando el programa R, por lo que es necesario primeramente obtener las funciones de distribución acumulada, luego realizar un ajuste mediante la prueba de Kolmogorov – Smirnov, esto es:

1. **Determinamos la función de probabilidad para las series correspondientes a los tiempos de llegada (TLL)**

Graficando la función de densidad

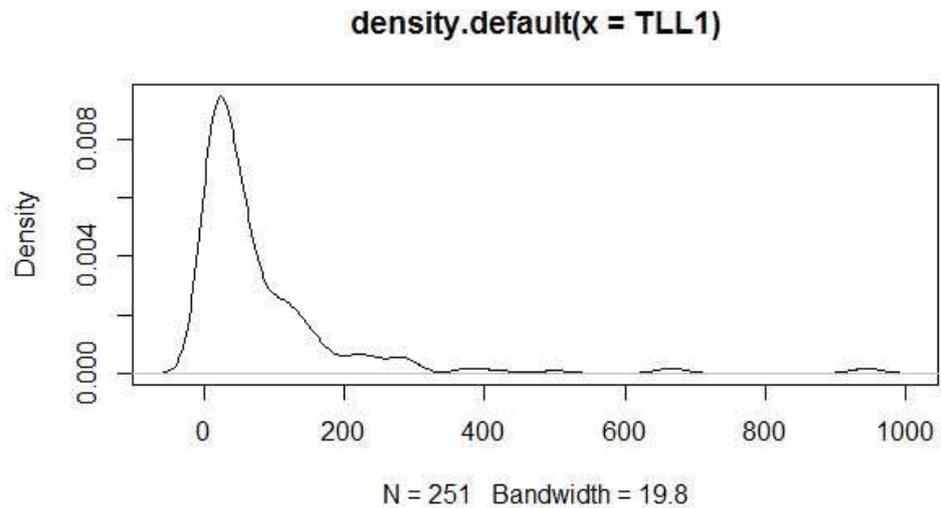


Figura 12. Función de densidad para la serie TLL1

Nota: Resultados obtenidos en este estudio aplicando el programa R

Ahora comparamos las funciones de distribución acumulada de la función empírica (TLL1) con la función exponencial, esto es:

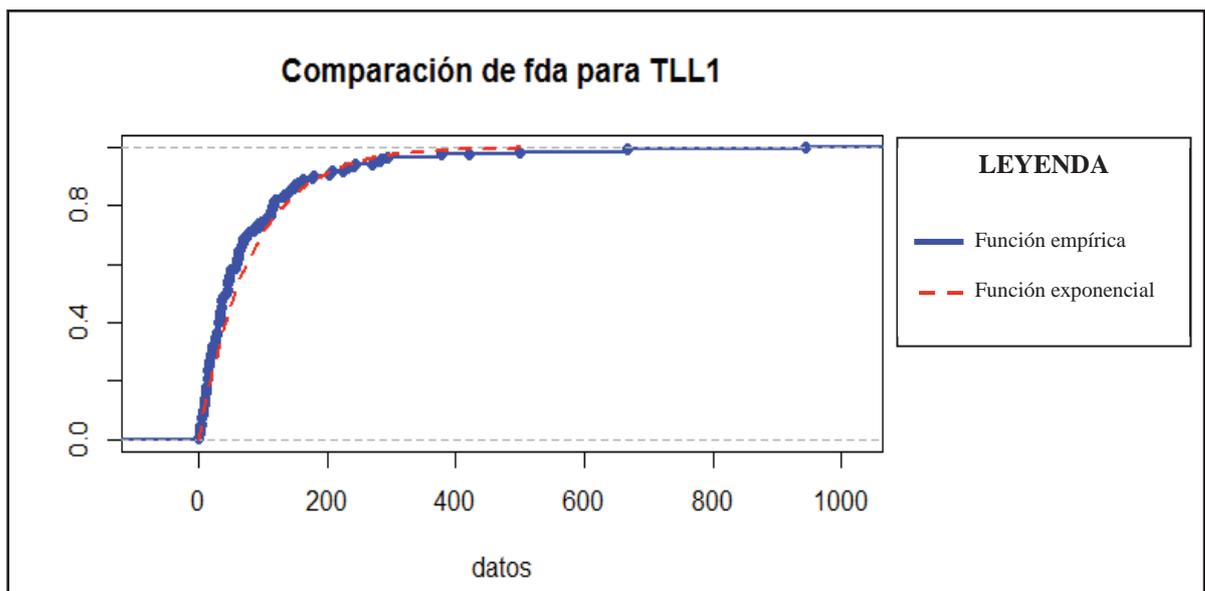


Figura 13. Comparación de función de distribución acumulada

Nota: Resultados obtenidos en este estudio aplicando el programa R

Comparando las gráficas de las funciones de distribución acumuladas tanto para la función empírica con la función exponencial, observamos gran similitud, por lo que aplicamos la prueba de bondad de ajuste para comprobar, si la serie para el tiempo de llegadas es exponencial, esto es:

Ho: $\text{fun_empírica}(\text{TLL1}) = \text{fun_exp}(\text{TLL2})$

Ha: $\text{fun_empírica}(\text{TLL1}) \neq \text{fun_exp}(\text{TLL2})$

$\alpha = 0.05$

```
> ks.test(fun_empirica(TLL1) and fun_exp(TLL2),
est, exact = FALSE, alternative = "greater")
Two-sample Kolmogorov - Smirnov test
data: fun_empírica(TLL1) and fun_exp(TLL1)
p-value = 0.9052
```

Interpretación: Con la prueba de bondad de ajuste se obtiene un p-valor = 0.9052 por lo que no se rechaza la hipótesis nula que nos indica un alto grado de semejanza de los datos de la serie con la distribución exponencial, por lo que concluimos que la serie TLL1 se ajusta a la distribución exponencial; de la misma forma para las demás series obtenemos:

```
> ks.test(fun_empirica(TLL2),fun_exp(TLL2),est,exact = FALSE, alternative = "greater")
Two-sample Kolmogorov-Smirnov test
data: fun_empirica(TLL2) and fun_exp(TLL2)
D^+ = 0.023904, p-value = 0.8664
alternative hypothesis: the CDF of x lies above that of y
```

Figura 14. Prueba de K-S para el día martes

```

> ks.test(fun_empirica(TLL3),fun_exp(TLL3),est,exact = FALSE, alternative = "greater")

Two-sample Kolmogorov-Smirnov test

data: fun_empirica(TLL3) and fun_exp(TLL3)
D^+ = 0.01992, p-value = 0.9052
alternative hypothesis: the CDF of x lies above that of y

```

Figura 15. Prueba de K-S para el día miércoles

```

> ks.test(fun_empirica(TLL4),fun_exp(TLL4),est,exact = FALSE, alternative = "greater")

Two-sample Kolmogorov-Smirnov test

data: fun_empirica(TLL4) and fun_exp(TLL4)
D^+ = 0.063745, p-value = 0.3606
alternative hypothesis: the CDF of x lies above that of y

```

Figura 16. Prueba de K-S para el día jueves

```

> ks.test(fun_empirica(TLL5),fun_exp(TLL5),est,exact = FALSE, alternative = "greater")

Two-sample Kolmogorov-Smirnov test

data: fun_empirica(TLL5) and fun_exp(TLL5)
D^+ = 0.023904, p-value = 0.8664
alternative hypothesis: the CDF of x lies above that of y

```

Figura 17. Prueba de Komogorov para el día viernes

Con estos resultados podemos verificar que los clientes ingresan al sistema con una distribución exponencial, que es fundamental para el proceso de simulación.

2. Determinamos la función de probabilidad para las series correspondientes a los tiempos de servicio (TA)

Determinando la distribución de probabilidad que se ajuste más a las series de datos para el tiempo de atención en cada uno de los 5 días

aplicando R, donde obtenemos las funciones de distribución acumulada, luego realizamos un ajuste mediante la prueba de Kolmogorov – Smirnov, esto es:

Determinando la función para la serie TA1, primeramente graficamos un histograma de frecuencias

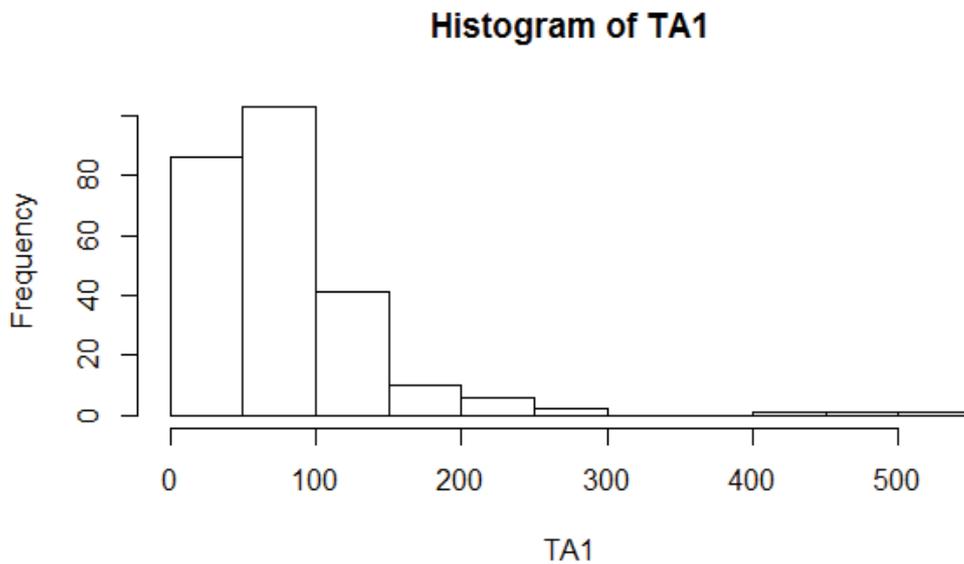


Figura 18. Histograma de frecuencias para las series del tiempo de servicio

Nota: Resultados obtenidos en este estudio aplicando el programa R

Interpretación: Del histograma observamos que hay mayor concentración de datos para valores de TA1 menores que 100 segundos, quiere decir que la serie es insesgada para la derecha y la mayor cantidad de datos tienen un tiempo bajo de servicio, es decir los clientes no pierden mucho tiempo en el momento de la atención.

Comparación de fda para TA1

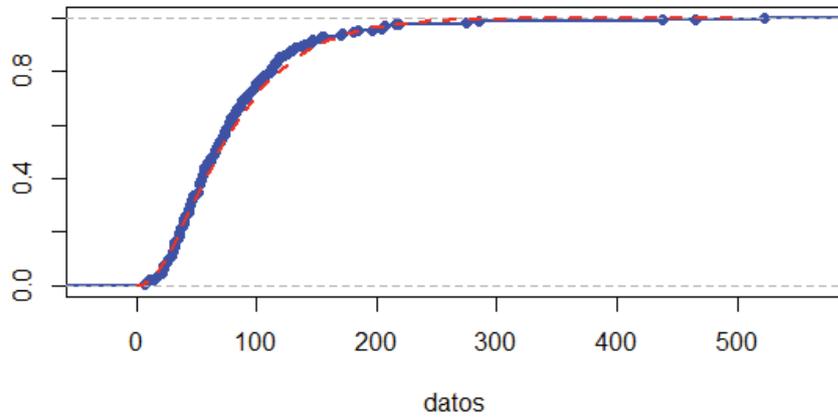


Figura 19. Comparación de función de distribución acumulada

Nota: Resultados obtenidos en este estudio aplicando el programa R

density.default(x = TA1)

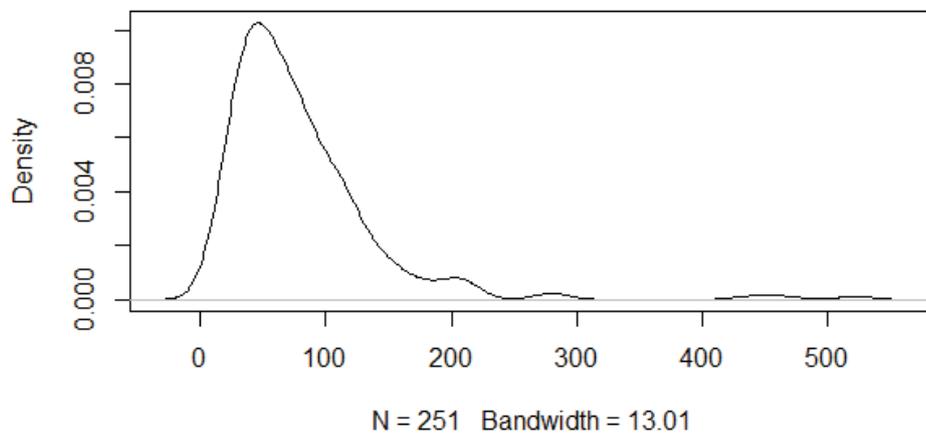


Figura 20. Función de densidad para el tiempo de servicio

Nota: Resultados obtenidos en este estudio aplicando el programa R

Comparación de fda para TA1

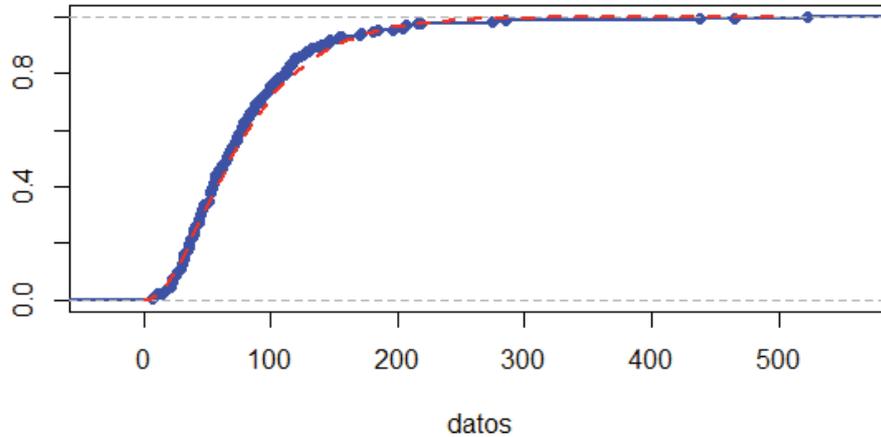


Figura 21. Comparación de FDA para TA1

Nota: Resultados obtenidos en este estudio aplicando el programa R

Interpretación: Comparando las gráficas de las funciones de distribución acumuladas tanto para la función empírica con la función gamma, observamos gran similitud, por lo que aplicamos la prueba de bondad de ajuste, para validar esta característica, esto es:

Ho: $\text{fun_empírica}(\text{TA1}) = \text{fun_exp}(\text{TA1})$

Ha: $\text{fun_empírica}(\text{TA1}) \neq \text{fun_exp}(\text{TA1})$

$\alpha = 0.05$

```
> ks.test(fun_empirica(TA1),fun_gamma(TA1),est,exact = FALSE, alternative = "greater")  
  
Two-sample Kolmogorov-Smirnov test  
  
data: fun_empirica(TA1) and fun_gamma(TA1)  
D+ = 0.035857, p-value = 0.7242  
alternative hypothesis: the CDF of x lies above that of y
```

Figura 22. Prueba de K-S para TA1

Con la prueba de bondad de ajuste se obtiene un $p\text{-value} = 0.7242$ al 95% de confianza, que nos indica un alto grado de semejanza de los

datos de la serie con la distribución gamma, por lo que concluimos que la serie TA1 se ajusta a la distribución gamma; de la misma forma para las demás series obtenemos:

```
> ks.test(fun_empirica(TA2),fun_gamma(TA2),est,exact = FALSE, alternative = "greater")  
  
Two-sample Kolmogorov-Smirnov test  
  
data: fun_empirica(TA2) and fun_gamma(TA2)  
D^+ = 0.035857, p-value = 0.7242  
alternative hypothesis: the CDF of x lies above that of y
```

Figura 23. Prueba de K-S para TA2

De estos resultados podemos afirmar que el modelo de teoría de colas que se presenta para el conjunto de datos es M/M/1, es decir que la distribución para los tiempos de llegada es exponencial y para los tiempos de servicio también consideraremos exponencial, el número de servidores del sistema se considera en 1 puesto que hay una sola ventanilla de atención en estudio.

5.2. Presentación de resultados

Simulación de los procesos de Atención

Aplicando el programa ProModel pasaremos a realizar la simulación de acuerdo a lo explicado en el capítulo II con los datos obtenidos y el ajuste de las distribuciones tanto para el tiempo de servicio y el tiempo de llegada. Tomamos los tiempos promedios para cada día generando dos grupos homogéneos:

Grupo N° 1 (G1): TLL1, TLL2, TLL5

Grupo N° 2 (G2): TLL3, TLL4

Completamos los tiempos promedios de llegada, servicio y espera (en minutos) de la tabla siguiente:

Tabla 7. Resumen descriptivo de las series en minutos

Día de la semana	N° de clientes n	Promedios		
		Tiempo de llegada	Tiempo de servicio	Tiempo de Espera
		TLL	TA	TE
Lunes	337	1.28	1.28	20.26
Martes	332	1.25	1.34	18.97
Miércoles	254	1.63	1.83	6.22
Jueves	251	1.67	1.69	14.35
Viernes	312	1.43	1.72	11.78

Fuente: Resultados obtenidos en este estudio

$$\text{Tiempo medio de llegadas para G1: } \frac{1}{\lambda} = \frac{1.28+1.25+1.43}{3}$$

$$\frac{1}{\lambda} = 1.32 \quad \Rightarrow \quad \lambda = 0.75$$

$$\text{Tiempo medio de servicios para G1: } \frac{1}{\mu} = \frac{1.28+1.34+1.72}{3}$$

$$\frac{1}{\mu} = 1.45 \quad \Rightarrow \quad \mu = 0.68$$

$$\text{Determinando } \rho = \frac{\lambda}{\mu} \quad \rho = \frac{\lambda}{\mu} = \frac{0.75}{0.68} = 1.10 > 1$$

por lo tanto, la cola no se asienta y será necesario aumentar un servidor más, luego:

$$\text{Sea } S = 2 \quad \Rightarrow \quad \rho = \frac{\lambda}{S\mu} = \frac{0.75}{2(0.68)} = 0.55 < 1$$

Por lo que la cola tiende a estabilizarse.

Tiempo medio de llegadas para G2: $\frac{1}{\lambda} = \frac{1.63+1.67}{2}$
 $\frac{1}{\lambda} = 1.65 \Rightarrow \lambda = 0.60$

Tiempo medio de servicios para G2: $\frac{1}{\mu} = \frac{1.83+1.69}{2}$
 $\frac{1}{\mu} = 1.76 \Rightarrow \mu = 0.56$

Determinando $\rho = \frac{\lambda}{\mu}$ $\rho = \frac{\lambda}{\mu} = \frac{0.60}{0.56} = 1.07 > 1$

Por lo tanto, la cola no se asienta y será necesario aumentar un servidor más, luego:

$$\text{Sea } S = 2 \Rightarrow \rho = \frac{\lambda}{S\mu} = \frac{0.60}{2(0.56)} = 0.53 < 1$$

Por lo que la cola tiende a estabilizarse.

Luego, simulando en ProModel para G1, se obtiene los siguientes resultados:



Figura 24. Captura de pantalla de ProModel, simulación

Tabla 8.
Resumen de resultados G1 (un servidor)

Nombre	Tiempo Programado (Min)	Capacidad	Total Entradas	Tiempo Por entrada Promedio (Min)	Contenido Máximo	Contenido Actual	% Utilización
FILA DE ESPERA	420	20	305	13.02	20	17	47.26
CAJERO	420	1	288	1.36	1	1	93.51
Loc1	420	1	0	0	0	0	0

Fuente: Resultados obtenidos en este estudio aplicando el programa Pro Model

Observamos que en las 7 horas de simulación que es el tiempo total que funciona el sistema por día, se obtuvo 305 llegadas de los cuales han sido atendidos 288 clientes, también se puede observar que 17 clientes abandonaron el sistema debido a que hubo saturación en la cola, lo cual indica una pérdida para el banco; además también se observa que el cajero estuvo funcionando al 93.51% de su capacidad que es relativamente sin descanso, por lo que concluimos que el sistema no satisface a los clientes por la pérdida de tiempo (en promedio 13.01 minutos de tiempo de espera), así mismo a los cajeros por exceso de trabajo.

Ahora simularemos el sistema con dos servidores, obteniendo los siguientes resultados:

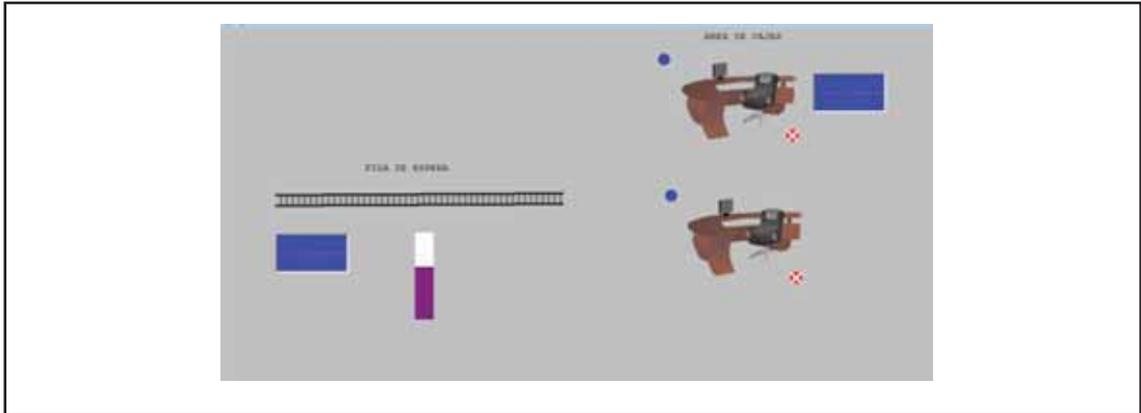


Figura 25. Captura de pantalla de ProModel, simulación

Tabla 9.
Resumen de resultados G1 (dos servidores)

Nombre	Tiempo Programado (Min)	Capacidad	Total Entradas	Tiempo Por entrada Promedio (Min)	Contenido Máximo	Contenido Actual	% Utilización
FILA DE ESPERA	420	20	327	0.52	6	0	1.99
CAJERO.1	420	1	183	1.42	1	1	61.89
CAJERO.2	420	1	144	1.32	1	1	45.31
CAJERO	840	2	327	1.37	2	2	53.60
Loc1	420	1	0	0	0	0	0

Fuente: Resultados obtenidos en este estudio aplicando el programa Pro Model

Para este caso observamos que ingresaron 327 clientes al banco, donde se pudo atender al total de clientes, teniendo al sistema trabajando en un 53.60% es decir a la mitad de su capacidad, sin embargo se encuentra satisfacción en los clientes debido a que solo esperan en promedio 2 minutos para ser atendidos. Por tanto concluimos que el sistema funciona mejor con dos servidores, es decir el modelo de cola que se ajusta al sistema del banco es M/M/2.

Ahora simularemos en Promodel para el grupo G2, donde se obtuvo los siguientes resultados:

Tabla 10.
Resumen de resultados G2 (un servidor)

Nombre	Tiempo Programado (Min)	Capacidad	Total Entradas	Tiempo Por entrada Promedio (Min)	Contenido Promedio	Contenido Actual	% Utilización
FILA DE ESPERA	420	20	285	21.81	14.80	17	73.99
CAJERO	420	1	268	1.56	0.99	1	99.29
Loc1	420	1	0	0	0	0	0

Fuente: Resultados obtenidos en este estudio aplicando el programa Pro Model

Observamos que con un solo cajero para el grupo G2, el sistema se satura, donde de los 285 clientes que ingresaron, 17 abandonaron el sistema debido a que hubo saturación en la cola, también podemos apreciar que el cajero se encuentra trabajando en el 99.29% de su capacidad lo cual puede implicar error en algunas de las transacciones por parte del cajero debido al cansancio originado por el exceso de trabajo. Entonces será necesario modelar el sistema con dos servidores, donde obtenemos los siguientes resultados:

Tabla 11.
Resumen de resultados G2 (dos servidores)

Nombre	Tiempo Programado (Min)	Capacidad	Total Entradas	Tiempo Por entrada Promedio (Min)	Contenido Promedio	Contenido Máximo	Contenido Actual	% Utilización
FILA DE ESPERA	420	20	333	0.98	0.78	10	0	3.91
CAJERO.1	420	1	182	1.56	0.67	1	0	67.75
CAJERO.2	420	1	151	1.45	0.52	1	0	52.40
CAJERO	840	2	333	1.51	0.60	2	0	60.07
Loc1	420	1	0	0	0	0	0	0

Fuente: Resultados obtenidos en este estudio aplicando el programa Pro Model

Para este caso observamos que ingresaron 333 clientes al banco, donde se pudo atender al total de clientes, teniendo al sistema trabajando en un 60.07% encontrando la satisfacción en los clientes debido a que solo esperan en promedio 1.56 minutos para ser atendidos. Por tanto nuevamente concluimos que el sistema funciona mejor con dos servidores, comprobando así que el modelo de cola que se ajusta de mejor manera es el M/M/2.

DISCUSIÓN

Al obtener los resultados de esta investigación aplicando la simulación con ProModel, sugiere una distribución de probabilidad para los datos obtenidos, encontramos que el sistema que emplea la agencia bancaria no es óptimo si sólo utiliza un servidor, pero si utiliza dos servidores el sistema funciona en toda su capacidad, obteniendo más ingresos a la empresa y teniendo más clientes satisfechos. La simulación con ProModel, no sólo ayuda a escoger un sistema adecuado de atención sino que también reduce los costos de investigación.

Urcia & Varela (2017) encontraron que un 89.3 por ciento de los clientes que hacen uso de las ventanillas de un banco están insatisfechos con el tiempo que emplean para hacer atendidos, éste resultado se obtuvo realizando una encuesta de satisfacción y sugieren que la atención en las ventanillas no debe superar los 3.5 minutos por cliente, sin embargo no realizaron un trabajo experimental y mucho menos de simulación para referenciar estos resultados.

Hernandez et al (2018) indica que la simulación ayuda a facilitar el cálculo de las variables y su proceso, así mismo permite reducir tiempo y dinero que en la práctica causarían errores costosos; sin embargo podemos acotar también que la simulación ayuda a resolver problemas sin necesidad de la presencia de los individuos implicados (clientes), teniendo solamente conocimiento de la distribución con que llegan los clientes a la cola.

Bazoun & Muñoz (2016) indica que los factores que dan calidad al servicio de la agencia bancaria son: seguridad, rapidez, tangibilidad y respeto; de ellos el único calificado nominalmente en desacuerdo como de calidad es la rapidez. Sin embargo se puede observar que para la existencia de rapidez no sólo depende del servicio si no también de la cantidad de servidores, por lo que consideró la necesidad de una simulación.

CONCLUSIONES

1. El sistema obtenido de la simulación con mayor eficiencia presenta como característica dos ventanillas, el promedio del tiempo de servicio TA es 96.1 segundos por cliente (aproximadamente 1:36 minutos por cliente) y el número promedio de clientes en la cola es 12 personas.
2. El comportamiento de tráfico en ventanillas con respecto al número de clientes que llegan es exponencial con $\lambda = 1.29$ minutos y con respecto al tiempo de servicio es exponencial con $\lambda = 1.60$ minutos
3. El modelo de la teoría de colas que garantiza mayor eficiencia del sistema es M/M/2, es decir la atención con 2 servidores en ventanilla, cuando las distribuciones de servicio y de llegada sea exponencial.

RECOMENDACIONES

1. Realizar la observación de las series por periodos más largos (años) para determinar las diferencias del comportamiento de las series entre días del mes, semanas del mes y meses del año.
2. Utilizar el software de simulación Promodel para modelar situaciones más complejas como el uso de cajeros automáticos, servicio de atención al cliente y otros servicios adicionales que tienen los clientes en el banco.
3. Aplicar la simulación mediante ProModel en la enseñanza universitaria puesto que simular genera ahorro de tiempo y dinero en la investigación con series que dependen del tiempo.

BIBLIOGRAFIA

- Arista, J. (2016). Aplicación de la teoría de colas al problema de atención al cliente para la optimización del número cajeros en ventanillas en la organización BCP. Universidad Nacional Mayor de San Marcos, Lima:Perú.
- Bazoun, A. H., & Muñoz, D. (2016). Propuesta de mejora para la atención de los usuarios de ventanilla preferencial del Banco de la Nación en la Agencia "3" José Leonardo Ortiz - Chiclayo. Universidad Católica Santo Toribio de Mogrovejo, Chiclayo: Perú.
- Bernal, C. A. (2010). Metodología de la investigación: administración, economía, humanidades y ciencias sociales. México: Pearson Educación, Prentice Hall.
- Bertsimas, D. J., & Servi, L. D. (1992). Deducing Queueing from Transactional Data: The Queue Inference Engine, Revisited. *Operations Research*, 40(3-supplement-2), S217–S228. <https://doi.org/10.1287/opre.40.3.S217>
- Cao, R. (2002). Introducción a la simulación y a la teoría de Colas. A Coruña: Netbiblo.
- Carro, R., & Gonzales-Gonzales, D. A. (2012). Modelos de Líneas de espera.
- Clemente, L. A. M. (2008). Mejora en el Nivel de Atención a los clientes de una entidad Bancaria usando Simulación. Pontificia Universidad Católica del Perú, Lima:Perú.
- Dominguez, G., & Bocanegra, A. R. (2016). Didactica y aplicación de la administración de operaciones. Contaduría y administración (1st ed.). IMCP (Instituto Mexicano de contadores Publicos).

- Hernández, R., Fernández, C., & Baptista, P. (2014). Metodología de la investigación. Mexico, D.F.: McGrawHill.
- Hiller, F. S., & Lieberman, G. J. (2002). Investigación de operaciones (McGraw-Hill/Interamericana Editores). SA.
- Karian, Z. A., & Dudewicz, E. J. (1999). Modern statistical systems and GPSS simulation (2nd ed). Boca Raton, Fla: CRC Press.
- Madeiro, M. C. (2012). Modelo de gestión de cajas de un supermercado utilizando datos transaccionales. Universidad de Chile, Santiago de Chile: Chile.
- Martínez, M. A., Morales, J., Barber, J. X., & Aparicio, J. (2007). Apuntes de teoría de colas: simulación y análisis de sistemas de espera. Elche: Universidad Miguel Hernández.
- Monzón, W. J. (2011). Optimización de las facilidades del servicio en una entidad bancaria. Universidad Nacional Mayor de San Marcos, Lima:Perú.
- Ramírez, J. A. (1995). Bata de cola: apuntes para una teoría de Sevilla. Madrid, España : Merida, Badajóz [Spain]: Ediciones Libertarias ; Editora Regional de Extremadura.
- Sobczyk, K. (2014). Stochastic Approach to Fatigue: Experiments, Modelling and Reliability Estimation. Retrieved from <http://public.ebib.com/choice/publicfullrecord.aspx?p=5595349>
- Taha, H. A., & González Pozo, V. (2004). Investigación de operaciones. México: Pearson/Educación.
- Urcia, V. M., & Varela, P. J. (2017). Influencia de los tiempos de espera para la atención en ventanilla en la calidad de servicio del Banco de Crédito Oficina El Perovenir en el 2017. Universidad Privada de norte, Trujillo:Perú.

ANEXOS

ANEXO 1

MATRIZ DE CONSISTENCIA

MARCO TEÓRICO	FORMULACION DEL PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES	DISEÑO METODOLÓGICO
<p>Antecedentes de investigación: Con el objeto de dar solución al problema de optimización del número de cajeros en ventanilla que minimice los costos de espera y los de servicios en organización BCP. Utiliza el programa SIMIO para simular la optimización del número de cajeros en las ventanillas. (Arista et al., 2016)</p>	<p>Problema general ¿Cuál es la eficiencia de las líneas de espera aplicadas al tráfico en las ventanillas de una agencia bancaria de la ciudad de Cusco?</p>	<p>Objetivo general Analizar la eficiencia de las líneas de espera aplicadas al tráfico en las ventanillas de una agencia bancaria de la ciudad de Cusco.</p>	<p>Hipótesis general Un sistema con dos servidores obtenidos de mediante la simulación para las líneas de espera aplicadas al tráfico de las ventanillas de una agencia bancaria presenta mayor eficiencia en la atención a los clientes</p>	<p>Variable:</p> <p>Tráfico en la ventanilla de un banco</p>	<p><u>Enfoque Investigación:</u> Cuantitativo</p> <p><u>Nivel de la Investigación:</u> Descriptivo</p> <p><u>Tipo de investigación:</u> Aplicada</p> <p><u>Diseño de investigación:</u> No experimental, transversal</p> <p><u>Unidad de Análisis</u> Entidad bancaria</p> <p><u>Población de estudio</u> Clientes del banco que realizan una transacción bancaria</p> <p><u>Técnica de selección de muestra</u> No Probabilístico</p> <p><u>Tamaño de muestra</u> n = 1486</p> <p><u>Ámbito de estudio (localización política y geográfica)</u> Una agencia de un banco en la ciudad del Cusco</p> <p><u>Técnicas de recolección de información</u> Fichas de registros</p>
<p>Bases teóricas 1. Teoría de colas 2. Series de Tiempo 3. Simulación de sistemas dinámicos</p> <p>Marco Conceptual (Palabras clave): 1. Simulación 2. Sistemas dinámicos 3. Líneas de espera 4. Teoría de colas</p>	<p>Problemas específicos 1. ¿Cómo es el comportamiento del tráfico en las ventanillas del banco? 2. ¿Cuál es el modelo de teoría de colas en el tráfico en las ventanillas del banco que garantice mayor eficiencia mediante la simulación?</p>	<p>Objetivos específicos 1. Describir el comportamiento del tráfico en las ventanillas de un banco 2. Determinar el modelo de teoría de colas en el tráfico en las ventanillas del banco mediante la simulación</p>	<p>Hipótesis específicas No Aplica No Aplica</p>		

ANEXO 2

RESULTADOS DE LAS COMPARACIONES DE MEDIAS DE LAS SERIES CORRESPONDIENTES A LOS TIEMPOS DE LLEGADAS

> t.test(TLL1,TLL2)

Welch Two Sample t-test

data: TLL1 and TLL2
t = 0.21585, df = 636.01, **p-value = 0.8292**
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-13.87889 17.30687
sample estimates:
mean of x mean of y
76.98507 75.27108

> t.test(TLL1,TLL5)

Welch Two Sample t-test

data: TLL1 and TLL5
t = -0.42178, df = 641.27, **p-value = 0.6733**
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-19.82488 12.81426
sample estimates:
mean of x mean of y
76.98507 80.49038

> t.test(TLL2,TLL5)

Welch Two Sample t-test

data: TLL2 and TLL5
t = -0.70171, df = 629.66, **p-value = 0.4831**
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-19.825496 9.386895
sample estimates:
mean of x mean of y
75.27108 80.49038

```
> t.test(TLL1,TLL3)
```

Welch Two Sample t-test

```
data: TLL1 and TLL3  
t = -2.0671, df = 518.49, p-value = 0.03922  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-40.418427 -1.027766  
sample estimates:  
mean of x mean of y  
76.98507 97.70817
```

```
> t.test(TLL1,TLL4)
```

Welch Two Sample t-test

```
data: TLL1 and TLL4  
t = -2.5699, df = 563.59, p-value = 0.01043  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-40.64780 -5.42986  
sample estimates:  
mean of x mean of y  
76.98507 100.02390
```

ANEXO 3
RESULTADOS DE LAS COMPARACIONES DE MEDIAS DE LAS SERIES
CORRESPONDIENTES A LOS TIEMPOS DE SERVICIO

> t.test(TA1,TA2)

Welch Two Sample t-test

data: TA1 and TA2
t = -0.70668, df = 664.97, **p-value = 0.4800**
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-13.35099 6.28424
sample estimates:
mean of x mean of y
77.02687 80.56024

> t.test(TA1,TA5)

Welch Two Sample t-test

data: TA1 and TA5
t = -2.2049, df = 616.07, **p-value = 0.0278**
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-18.734610 -1.083454
sample estimates:
mean of x mean of y
77.02687 86.93590

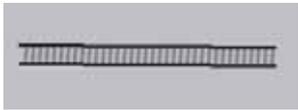
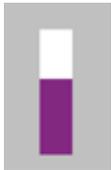
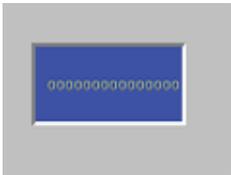
> t.test(TA2,TA5)

Welch Two Sample t-test

data: TA2 and TA5
t = -1.429, df = 614.83, **p-value = 0.1535**
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-15.137337 2.386024
sample estimates:
mean of x mean of y
80.56024 86.93590

ANEXO 4

CONFIGURACIÓN DE PROMODEL PASO A PASO.

Fila		Lugar donde los clientes esperan a ser atendidos
Mesa (Lugar)		Lugar donde clientes serán atendidos
Barra de estado		Muestra el nivel de saturacion de la fila
Contador		Acumula la cantidad de atenciones y llegadas al sistema
Cliente (Entidad)		Representa al Cliente que utiliza el sistema.

1. Se instaló el software de modelado de la web del fabricante, (ProModel Corporation, 2015). Utilizándose el navegador Google Chrome y mínimamente el sistema operativo Windows 7.
2. Para empezar a utilizar el modelador se ha definido previamente en el Capítulo II , las distribuciones de tiempo de llegada y de atención de los clientes a la organización, debido a que estos datos deben ser insertados en el modelador a fin de utilizar las variables correspondientes.

3. Se procedió a iniciar el software clickeando doble vez en el ícono  situado en el escritorio del equipo una vez finalizada la instalación del paso 1.
4. Una vez cargado el software, nos vamos a la sección Archivo, luego a Nuevo, con el fin de crear un nuevo documento de simulación.



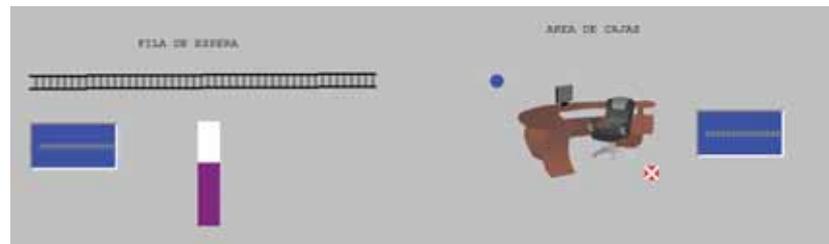
5. Una vez dentro, procedemos a nombrar nuestro aplicativo, dándole el nombre *Simulacion_Banco*, y según las unidades de mediciones, marcamos las casillas correspondientes a *Minutos* y *Metros* para el tiempo y la distancia respectivamente, posteriormente, clickeamos en *Ok*.



6. Seguidamente empezamos a construir el escenario de simulación, haciendo click en Construir y luego en Locaciones.



7. Utilizamos las locaciones necesarias, arrastrando de izquierda a derecha las locaciones, quedando con el modelo siguiente:



8. Renombramos las locaciones de la siguiente manera:

Ícono	Nombre	Cap.	Unidades
	FILA_DE_ESPERA	20	1
	CAJERO	1	1
	TEXTO_AREA_CAJAS	1	1

Siendo:

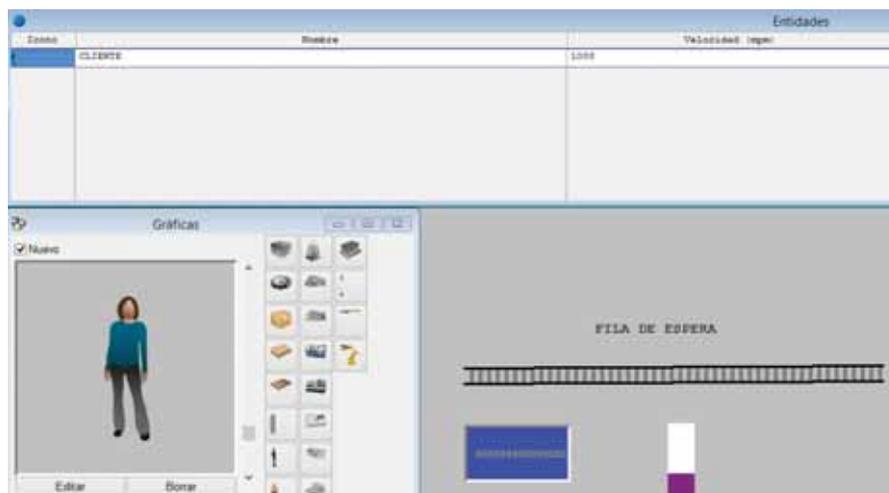
Ícono: La imagen de la locación propiamente dicha.

Nombre: El nombre (como variable) de la locación.

Cap: La capacidad de la locación, en este caso, la *FILA_DE_ESPERA*, tendrá una capacidad de 20 personas, si ingresan más clientes al sistema cuando su capacidad está al máximo, los arribos se consideran fallidos por falta de capacidad.

Unidades: La cantidad numérica de cada locación.

9. Las entidades: muestra la representación gráfica del cliente en la simulación.



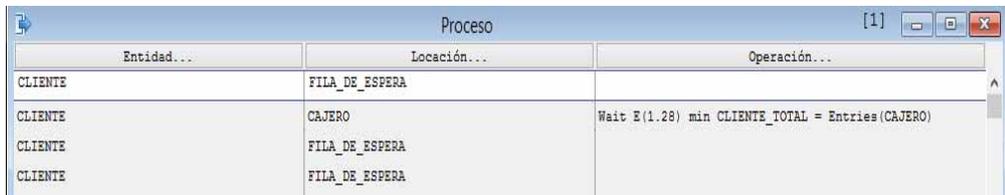
Siendo:

Ícono: El símbolo gráfico del cliente.

Nombre: El nombre (como variable) de la entidad.

Velocidad: Velocidad a la que se movera el icono por la simulación

10. El procesamiento:  Muestra la lógica del movimiento de la entidad por las locaciones.



Entidad...	Locación...	Operación...
CLIENTE	FILA_DE_ESPERA	
CLIENTE	CAJERO	Wait E(1.28) min CLIENTE_TOTAL = Entries(CAJERO)
CLIENTE	FILA_DE_ESPERA	
CLIENTE	FILA_DE_ESPERA	

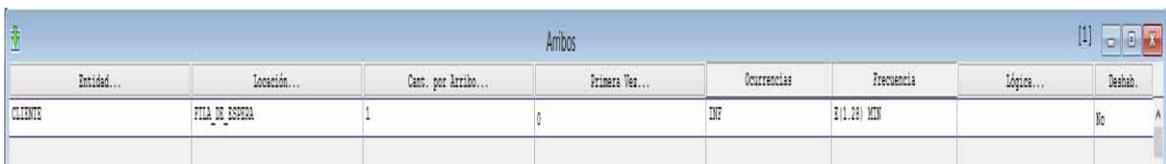
Siendo:

Entidad: El cliente que se mueve de (Inicio) a (Destino)

Para la primera repetición, el inicio se da en la *FILA_DE_ESPERA*, y el destino se muestra en la siguiente línea: *CAJERO*

Operación: Muestra la lógica de la atención, en la que la *ENTIDAD*, espera un tiempo según una distribución de *E(1.28) minutos*, mientras que *CLIENTE_TOTAL = Entries(CAJERO)*, acumula la cantidad de clientes que llegaron hasta el cajero para ser atendidos.

11. Los Arribos,  muestran la lógica de ingreso de entidades al sistema: mediante una distribución E(1.28 MIN)



Entidad...	Locación...	Cant. por Arribo...	Primera Vez...	Ocurrencias	Frecuencia	Lógica...	Destab.
CLIENTE	FILA_DE_ESPERA	1	0	INF	E(1.28) MIN		No

Siendo:

Entidad: El cliente que llega y usa el sistema.

Locación: *FILA_DE_ESPERA*, el lugar en el que el cliente iniciará el recorrido.

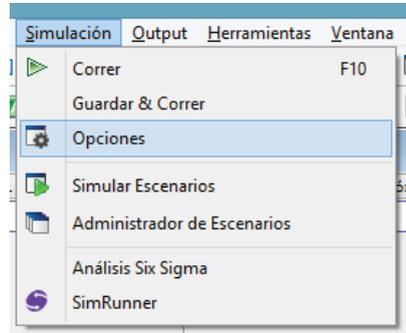
Cantidad por arribo: Muestra la cantidad de llegadas en una sola vez al sistema.

Primera vez: Muestra la cantidad de clientes acumulados, antes de iniciar el sistema.

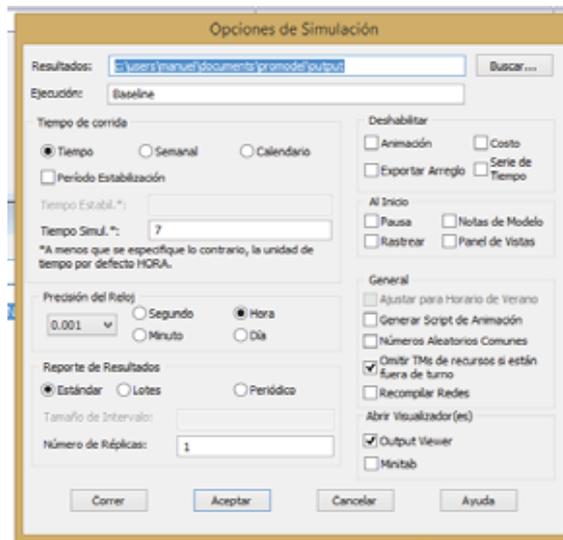
Ocurrencias: Es la cantidad de Entidades que ingresarán al sistema, por lo general se coloca en INF (infinito) para que el modelador permita definir la cantidad de arribos.

Frecuencia: Programa la lógica de distribución de llegadas.

12. Configuramos las opciones finales:



Y colocamos el tiempo de *utilización del sistema* en 7 horas.



Por último damos click en aceptar.

13. Iniciar la simulación. Hacemos click en

