UNIVERSIDAD MAYOR DE SAN ANDRÉS

FACULTAD DE CIENCIAS ECONÓMICAS Y FINANCIERAS

CARRERA DE CONTADURÍA PÚBLICA

PETAENG

Plan Excepcional de Titulación para Estudiantes Antiguos No Graduados

MODULO DE ACTUALIZACION

Para la obtención del Grado Académico de Licenciatura

"ANALISIS DE LA TEORIA DE COLAS Y SU APLICACIÓN AL SISTEMA FINANCIERO"

Autor: Flora Pamela Saire Conde

La Paz – Bolivia

2018

DEDICATORIA

Dedico este trabajo a Dios y a mis padres. A Dios porque ha estado conmigo a cada paso que doy, cuidándome y dándome fortaleza para continuar, a mis padres, quienes a lo largo de mi vida han velado por mi bienestar y educación siendo mi apoyo incondicional en todo momento. Depositando su entera confianza en cada reto que se me presentaba. Los amo con mi vida.

AGRADECIMIENTOS

Muchas gracias, a la Universidad Mayor de San Andrés y la Carrera de Contaduría Pública por darme la oportunidad de poder concluir mis estudios a través del programa P.E.T.A.E.N.G., asimismo, a todo el plantel docente quienes con su esfuerzo y dedicación hicieron posible este objetivo y a mis compañeros por su colaboración.

RESUMEN

La teoría de las colas es el estudio matemático de las colas o líneas de espera, donde la formación de colas es un fenómeno común en nuestro diario vivir que ocurre siempre que la demanda efectiva de un servicio excede a la oferta efectiva, por lo que se trata de una potente herramienta que puede ser de gran utilidad en una infinidad de situaciones tanto cotidianas como industriales y que se nos presentarán reiteradas veces en nuestro ambiente laboral.

Esto lo podemos notar claramente en las largas filas o colas que se forman a la espera de pagar un artículo en un almacén o para pagar cuentas en un banco. De manera que si estas filas se prolongan demasiado estos clientes podrían tomar la decisión de aproximarse hacia otro almacén de la competencia, por lo que el dueño del almacén obtendría menores ganancias, porque si estos clientes emigran hacia otros almacenes, las ventas disminuirían. Por otra parte si se controlarían más dependientes, significaría mayor tiempo ocioso, que es visto en pérdidas debido a una excesiva nómina del personal. Así que el propietario del almacén tendría que tomar la decisión de contratar una numerosa cantidad de empleados que sean capaces de reducir las pérdidas que han sido provocadas por las ventas no realizadas y por el tiempo ocioso de los dependientes.

La teoría de colas en si no resuelve este problema, se utiliza como ayuda a la toma de decisiones de la dirección de una empresa.

Con frecuencia, las empresas deben tomar decisiones respecto al caudal de servicios que debe estar preparada para ofrecer. Sin embargo, muchas veces es imposible predecir con exactitud cuándo llegarán los clientes que demandan el servicio y/o cuanto tiempo será necesario para dar ese servicio; es por eso que esas decisiones implican dilemas que hay que resolver con información escasa. Esto no es sencillo, ya que un cliente no llega a un horario fijo, osea que no se sabe con exactitud en que momento llegarán los clientes, así como también el tiempo de servicio no tiene un horario fijo. Pero, por otro lado, carecer de la capacidad de servicio suficiente causa colas excesivamente largas en ciertos momentos. Cuando los clientes tienen que esperar en una cola para recibir nuestros servicios, están pagando un coste, en tiempo, más alto del que esperaban. Las líneas de espera largas también son costosas por tanto para la empresa ya que producen pérdida de prestigio y pérdida de clientes.

El proceso básico por la mayor parte de los modelos de colas, es el siguiente:

Los clientes que requieren servicios, a través del tiempo, provienen de una fuente de entrada. Estos clientes arriban al sistema de servicios y se unen a una cola.

En un determinado tiempo se selecciona un miembro de la cola, mediante alguna regla conocida como disciplina de servicio.

Luego, se brinda el servicio requerido por el cliente en un mecanismo de servicio, después de lo cual el cliente sale del sistema de servicio.

Los objetivos de la teoría de colas, promueven la identificación del nivel óptimo de capacidad del sistema que minimiza el costo global. También consisten en evaluar el impacto de posibles alternativas de modificación de la capacidad del sistema, contribuye con la información vital que se requiere para tomar las decisiones concernientes prediciendo algunas características sobre la línea de espera: probabilidad de que se formen, el tiempo de espera promedio. Se basa en atender eficientemente las llegadas a las instalaciones que prestan algún tipo de servicio.

Es importante destacar que cuando una cola está compuesta de materiales (objetos inanimados) esperando algún tipo de procesamiento, el problema que tendrían es básicamente económico y cuánto equipo deben comprar. Por otro lado cuando una cola la forman personas esperando un servicio, podrían ser muy difíciles de cuantificar para ustedes.

La toma de decisiones es un proceso sumamente importante dentro de las empresas, cualquiera que sea su área de servicio. Esto se debe a que mediante de estas decisiones, se crean estrategias para mejorar el estado de la empresa.

Al presentarse un problema de colas, usted define primeramente cuál es el problema principal, es decir, el por qué se están llevando a cabo esas colas. Luego analiza esta problemática, definiendo las principales características que presenta. Después usted propone alternativas eligiendo la óptima para que aplique la decisión final y se pueda crear estrategias que mejoren o disminuyan este problema tan común en el diario vivir de las personas y claro, también de las empresas.

Básicamente la teoría de colas, estudia técnicas para solucionar problema que presentan separando así situaciones en las cuales se forman turnos de espera o colas que requieren la prestación un servicio.

INDICE

1. INTRODUCCION

Las colas son un aspecto de la vida moderna que nos encontramos continuamente en nuestras actividades diarias. Esto suele ocurrir, cuando la demanda real de un servicio es superior a la capacidad que existe para dar dicho servicio. Ejemplos reales de esa situación son: el pago de servicios en el sistema bancario, los semáforos mal ajustados, el pago de peaje en una autopista, los cajeros automáticos, el tráfico aéreo de un aeropuerto, la atención a clientes en un establecimiento comercial, la espera de los electrodomésticos dañados para ser reparados por un servicio técnico, la espera de pacientes para ser atendidos en un hospital, etc.

La teoría de colas o línea de espera, es una colección de modelos matemáticos que describen sistemas en donde los clientes esperan en una cola para recibir un servicio, estos clientes son elegidos de acuerdo a ciertos criterios del sistema de elección. Entre muchos usos, los modelos de colas pueden ser empleados para encontrar un "estado estable" en el sistema que genere un consumo óptimo de recursos, también se puede determinar la longitud promedio de la línea de espera (cola) y el tiempo de espera promedio para un sistema dado. El problema radica en determinar qué tamaño o tasa de servicio proporciona ese consumo óptimo, esto no es sencillo de determinar en sistemas de colas en bancos, dada la aleatoriedad en las llegadas de nuevos clientes; además el tiempo de servicio no es fijo, en algunos casos. Esta información, junto con los costos pertinentes, se usa para determinar la capacidad del sistema apropiada.

El análisis de teoría de colas es una de las herramientas utilizadas para la mejora de los diferentes sistemas de producción o servicios; particularmente en lo que concierne a los sistemas de servicios, el sistema bancario es una de las aplicaciones de esta teoría. Puesto que puede usarse para proporcionar respuestas aproximadas a muchas preguntas como las siguientes: ¿Cómo cambia el tiempo de espera en la cola si se agrega o se retira un canal de servicio (cajero)?, ¿Cuántos clientes son atendidos por servidor y cuánto tiempo desocupado tendrán los servidores?, ¿Cuántos canales de servicio o cajeros deberá tener un banco si se conoce el número de servicios por hora, de manera que el tiempo de espera sea aceptable tanto por el cliente, como para el servidor?. A todas las preguntas formuladas anteriormente se les dará respuesta en el desarrollo de nuestra investigación.

A continuación, definiremos algunos conceptos básicos que nos ayudara a entender la teoría de colas, desde una visión matemática.

2. ASPECTOS METODOLOGICOS DE ANALISIS

2.1. Objetivo general.

Conocer los diferentes conceptos, con el fin de profundizar y enriquecer el conocimiento a través de la investigación y recopilación de información acerca del tema. Mostrar de forma teórica los pasos a seguir para realizar un trabajo de forma específica respecto a la aplicación de la teoría de colas al sistema financiero. En donde se busca emitir una opinión independiente respecto al cumplimiento del sistema

2.2. Objetivos específicos.

- Explicar el desarrollo y aplicación de la teoría de colas
- Describir de los conceptos para el mejor entendimiento de la teoría de colas y su aplicación al sistema financiero.
- Identificar las consideraciones cuantitativas y cualitativas del servicio.

2.3. Justificación.

La realización de este trabajo está basada en la necesidad de conocer la aplicación de la teoría de colas que beneficia al sistema financiero en el control de espera, organización eficaz de distintas colas, reducción de tiempos de espera, asignación de atención en función al tipo de atención que quiere recibir desde la petición del ticket.

2.4. Alcance.

En el trabajo informe se describe la aplicación de la teoría de colas en el sistema financiero, identificando las operaciones que tienen a su cargo y cuál es el efecto de los resultados de la entidad financiera para lograr la atención eficiente al cliente.

2.5. Nivel de investigación.

El método de investigación utilizado para el presente trabajo será el método deductivo: en este método consiste en ir de lo general a lo particular, de forma que partiendo de enunciados de carácter general, se llegan a enunciados particulares, con el fin de establecer su estructura o comportamiento.

A través del tipo de estudio ya mencionado, se hará un análisis de los elementos que componen la temática ya abordada realizando una evaluación de su comportamiento.

El tipo de estudio del tema será:

• Descriptivo: Este tipo de estudio permitirá detallar, especificar y particularizar los hechos que se han suscitado en la ejecución para que nos permita sacar conclusiones que sean útiles para el presente trabajo.

2.6. Técnica de investigación.

Es una técnica que consiste en observar atentamente el fenómeno, hecho o caso, tomar información y registrarla para su posterior análisis. Para las técnicas de investigación se tomara en cuenta la siguiente bibliografía como ser:

- A. K. Erlang,(1996) "Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges", Birkhauser Boston.
- Cao Abad, Ricardo et. al. (2007-2008): Teoría de colas. Departamento de matemáticas (Universidad de A Coruña).
- Domingo Morales González (2003), Teoría de Colas. ISBN 84-605-1045-X.
- Gross, Donaldy CarlHarris (2008).Resumen traducido del libro Fundamentals of Queueing Theory, Wiley.
- Hillier F.S. y Lieberman G.J. (2010). Introducción a la investigación de operaciones. McGrawHill

3. MARCO PRÁCTICO

3.1. TEORIA DE COLAS

3.1.1. Definición.

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. Dentro de las matemáticas, la teoría de colas se engloba en la investigación de operaciones y es un complemento muy importante a la teoría de sistemas y la teoría de control. Se trata así de una teoría que encuentra aplicación en una amplia variedad de situaciones como negocios, comercio, industria, ingenierías, transporte y logística o telecomunicaciones.

3.1.2. Historia.

Históricamente, los primeros trabajos que comenzaron a dar cuerpo a la Teoría de Colas son los debidos al matemático danés A.K.Erlang[1], quien en 1909 publicó La teoría de probabilidades y las conversaciones telefónicas. Erlang era para ese entonces empleado de la Compañía Telefónica Danesa en Copenhage y su trabajo fue una aplicación de técnicas existentes en teoría de probabilidad al problema de determinar el número óptimo de líneas telefónicas en una centralita, teniendo en cuenta la frecuencia de las llamadas y su duración. Las aplicaciones de la Teoría de Colas a la telefonía continuaron después de Erlang. En 1927, E.C. Molina publicó Aplicación de la teoría de la probabilidad a problemas de líneas telefónicas; en 1928 T.C. Fry publicó Probabilidad y sus usos en Ingeniería. A principios de los años 30, F. Pollaczek publicó trabajos innovadores sobre el caso de llegadas poissonianas y servicios arbitrarios. También, por esa época, los matemáticos de la escuela Rusa A.N. Kolmogorov y A.Y. Khintchine, así como C.D. Crommelin, en Francia, y C. Palm, en Suecia, realizaron importantes aportaciones a la teoría de colas. A pesar de que a comienzos del estudio de la Teoría de Colas, las aportaciones fueron muy escasas, esta situación cambió notablemente a partir de los años 50, comenzando a publicarse gran número de trabajos sobre el tema. En la actualidad las aplicaciones de la Teoría de Colas en los campos de la Administración de empresas, Informática, las Telecomunicaciones y, en general, las nuevas tecnologías, abren aún un mayor porvenir a esta teoría matemática. Como bien sabemos la formación de colas o líneas de espera suele ocurrir cuando la demanda real de un servicio es superior a la capacidad que existe para dar dicho servicio. Ejemplos reales de esa situación son: los cruces de dos vías de circulación, los semáforos, el peaje de una autopista, los cajeros automáticos,

[1] Historia de la teoría de colas (https://es.wikipedia.org/wiki/Teor%C3%ADa_de_colas)

la atención a los clientes en un banco, etc. Fenómenos como los citados anteriormente y muchísimos otros tienen ciertas características comunes que dan lugar al desarrollo de modelos de sistemas de colas. La teoría de colas estudia el comportamiento de sistemas donde existe un conjunto limitado de recursos para atender las peticiones generadas por los clientes, de tal manera que cuando un cliente envía una tarea al sistema, ésta podrá tener que esperar para ser atendida por algún recurso del sistema, o, incluso, podrá ser rechazada si el sistema no tiene capacidad suficiente para que sea atendida. El estudio de estos sistemas implicará el modelado no sólo del sistema en sí, sino también el comportamiento aleatorio del tráfico ofrecido por los clientes al sistema. Este tráfico ofrecido por los usuarios se modela mediante dos procesos estocásticos: procesos de llegadas de las tareas al sistema y procesos de servicios en el sistema, usualmente considerados independientes entre sí. Con frecuencia, las empresas deben tomar decisiones respecto al caudal de servicios que debe estar preparada para ofrecer. Sin embargo, muchas veces es imposible predecir con exactitud cuándo llegarán los clientes que demandan el servicio y/o cuanto tiempo será necesario para dar ese servicio; es por eso que esas decisiones implican dilemas que hay que resolver con información escasa. Estar preparados para ofrecer todo servicio que se nos solicite en cualquier momento puede implicar mantener recursos ociosos y costos de operación excesivos. Pero, por otro lado, carecer de la capacidad de servicio suficiente causa colas excesivamente largas en ciertos momentos. Cuando los clientes tienen que esperar en una cola para recibir algún servicio, están pagando un coste, en tiempo, más alto del que esperaban. Las líneas de espera largas también son costosas para la empresa; ya que produce pérdida de prestigio y pérdida de clientes.

La eficiencia en el manejo de los tiempos direcciona a la entidad financiera en la prestación oportuna del servicio y permite una satisfacción en el cliente, representando su calidad de respuesta en lo que llamamos beneficio colectivo, basado basado en lo que concierne a una solución que contribuya a la toma de decisiones y el desarrollo del cumplimiento de los objetivos en comparación a sus competencias.

3.1.3. Disciplina de la cola.

La disciplina 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 colapsarse. Dentro de las matemáticas, la teoría de colas se engloba en la investigación de operaciones y es un complemento muy importante a la teoría de sistemas y la teoría de control. Se trata así de una teoría que encuentra aplicación en una amplia variedad de situaciones tomando en cuenta los siguientes ejemplos como negocios, comercio, industria, ingenierías, transporte y logística o también las telecomunicaciones.

En el caso concreto de la ingeniería, la teoría de colas permite modelar sistemas en los que varios agentes que demandan cierto servicio o prestación, confluyen en un mismo servidor y, por lo tanto, pueden registrarse esperas desde que un agente llega al sistema y el servidor atiende sus demandas. En este sentido, la teoría es muy útil para modelar procesos tales como la llegada de datos a una cola en ciencias de la computación, la congestión de red de computadoras o de telecomunicación, o la implementación de una cadena productiva en la ingeniería industrial.

En el contexto de la informática y de las tecnologías de la información y la comunicación las situaciones de espera dentro de una red son más frecuentes. Así, por ejemplo, los procesos enviados a un servidor para su ejecución forman colas de espera mientras no son atendidos; la información solicitada, a través de Internet, a un servidor Web puede recibirse con demora debido a la congestión en la red; también se puede recibir la señal de línea de la que depende nuestro teléfono móvil ocupada si la central está colapsada en ese momento, etc.

3.1.4. Mecanismos de servicio.

Se llama capacidad del servicio al número de clientes que pueden ser servidos simultáneamente. Si la capacidad es uno, se dice que hay un solo servidor (o que el sistema es mono canal) y si hay más de un servidor, multicanal. El tiempo que el servidor necesita para atender la demanda de un cliente (tiempo de servicio) puede ser constante o aleatorio; en este último caso se supone, por lo general, que los tiempos de servicio son vv.aa.ii.ii.dd. Además, que son independientes de los tiempos entre llegadas. A veces el servidor sólo está disponible durante una parte del tiempo de funcionamiento del sistema.

3.1.5. Estructuras típica.

El primer sistema que se muestra en la figura, se llama un sistema de un servidor y una cola. El segundo, una línea con múltiples servidores. El tercer sistema, aquél en que cada servidor tiene una línea de separación. El cuarto sistema, es una línea con servidores en serie. Este modelo puede aplicarse a trabajos ordenador que esperan tiempo de procesador.

Una cola se produce cuando la demanda de un servicio por parte de los clientes excede la capacidad del servicio.

- Se necesita conocer (predecir) el ritmo de entrada de los clientes y el tiempo de servicio con cada cliente.
- La teoría de colas es un conjunto de modelos matemáticos que describen sistemas de líneas de espera particulares.

Objetivo:

- Equilibrar los costos de capacidad del servicio y el “costo” de una espera larga.

3.1.6. Características.

Existen dos clases básicas de tiempos de llegada.

- **Determinístico**: El cual clientes sucesivos llegan en un mismo intervalo de tiempo, fijo y conocido.
- **Probabilístico:** El cual el tiempo entre llegadas sucesivas es incierto y variable. Los tiempos entre llegadas probabilísticas se describen mediante una distribución exponencial, ya que se ajusta a las características del sistema.

3.2. SISTEMA FINANCIERO.

3.2.1. El sistema financiero global.

Es el marco mundial de acuerdos legales, instituciones y agentes económicos tanto formales como informales que en conjunto facilitan flujos internacionales de capital financiero para propósitos de inversión y financiamiento comercial. Desde que emergió a finales del siglo XIX durante la primera ola moderna de globalización económica, su evolución es definida por el establecimiento de bancos centrales, tratados multilaterales, y organizaciones internacionales dirigidas a mejorar la transparencia, regulación y eficiencia de los mercados internacionales[2]. A finales del siglo XIX la migración mundial y las tecnologías para la comunicación facilitaron un crecimiento sin precedentes en el comercio internacional y la inversión. En el inicio de la Primera Guerra Mundial, el comercio se contrajo como resultado de que los mercados de divisas se paralizaron a causa de la liquidez del mercado monetario. Los países intentaban defenderse de las crisis externas con políticas proteccionistas y el comercio se detuvo virtualmente en 1933; empeorando los efectos de la Gran Depresión hasta que una serie de acuerdos comerciales recíprocos lentamente redujo los aranceles a nivel mundial. Los esfuerzos para renovar el sistema monetario internacional después de la Segunda Guerra Mundial mejoraron la estabilidad del tipo de cambio, promoviendo un crecimiento récord en las finanzas globales.

Una serie de devaluación de divisas y crisis petroleras en los 70's provocó que las monedas de la mayoría de los países fluctuaran. La economía mundial incremento su integración financiera en los 80's y 90's debido a la liberación de la cuenta de capital y la desregulación financiera. Una serie de crisis financieras en Europa, Asia, y América latina seguidas por efectos contagiosos debido a una

[2] James, Paul W.; Patomäki, Heikki (2007). *Globalization and Economy, Vol. 2: Globalizing Finance and the New Economy*. London, UK: Sage Publications. ISBN 978-1-4129-1952-4.

mayor exposición a flujos de capital volátiles. La crisis financiera global, que se originó en los Estados Unidos en el 2007, rápidamente se propagó entre las demás naciones y es conocida como el catalizador para la Gran Recesión a nivel mundial. Un ajuste de mercado al incumplimiento de Grecia con su unión monetaria en el 2009 originó una crisis de deuda soberana entre las naciones europeas conocida como la Crisis de la zona euro.

La decisión de un país de operar una economía abierta y globalizar su capital financiero lleva consigo implicaciones monetarias reflejadas en la balanza de pagos. También implica exposición a los riesgos de las finanzas internacionales, como deterioración política, cambios regulatorios, control de divisas, e incertidumbres legales de derechos de propiedad e inversiones. Tanto individuos como grupos participan en el sistema financiero global. Los consumidores y los negocios internacionales se encargan del consumo, producción e inversión. Los cuerpos gubernamentales e intergubernamentales actúan como promotores del comercio internacional, el desarrollo económico y la gestión de crisis. Los cuerpos regulators establecen regulaciones financieras y procedimientos legales, mientras que los cuerpos independientes facilitan la supervisión de la industria. Los institutos de investigación y otras asociaciones analizan información, publican reportes e informes normativos, y hacen discursos públicos sobre asuntos financieros globales.

Mientras que el sistema financiero global esta en el borde hacia una mayor estabilidad, los gobiernos deben lidiar con diversas necesidades regionales y nacionales. Algunas naciones están tratando de discontinuar ordenadamente las políticas poco convencionales instaladas para lograr recuperarse, mientras que otras se encuentran expandiendo su alcance y escala. Los legisladores de los mercados emergentes enfrentan un desafío de precisión ya que deben instituir cuidadosamente políticas macroeconómicas sustentables durante una sensibilidad de mercado extraordinaria sin provocar que los inversionistas retiren su capital hacia mercados más fuertes. La inhabilidad de las naciones para alinear sus intereses y lograr un consenso internacional en temas como regulación bancaria ha perpetuado el riesgo de futuras catástrofes financieras a nivel global.

3.2.2. El sistema financiero en Bolivia[3].

Si bien el ahorro es un elemento clave del bienestar individual, desde un punto de vista macroeconómico representa un bien común, esencial para el crecimiento económico de cualquier país, pues cuando éste se moviliza y se canaliza a la inversión, a través del sistema de intermediación financiera, posibilita el financiamiento de iniciativas que promueven una mayor

[3] El sistema financiero Boliviano (http://blogdocente.usfx.bo/ramiro-villegas/2017/06/05/el-sistema-financiero-boliviano/)

actividad económica. La intermediación financiera, es por tanto, un proceso por el cual las entidades financieras median entre el ahorro y la inversión. Es un proceso de confianza y administración de riesgos que es necesario preservar. En este sentido, velar por la confianza del público y la estabilidad del sistema de intermediación financiera, constituye una tarea fundamental que ningún Estado puede soslayar.

Las funciones y el rol que desempeña el Organismo Regulador y Supervisor del sistema de intermediación financiera, constituye un elemento esencial de la política pública en todos los países, por constituir este sector un componente básico de la economía. La necesidad de la regulación y supervisión financiera por parte del Estado responde a la naturaleza de la facultad soberana del Estado de velar por la cosa pública y el bienestar de la comunidad. En la mayoría de los países, la regulación y supervisión financiera ejercida por el Estado, obedece a dos propósitos interrelacionados entre sí: proteger los depósitos del público y mantener la solidez y estabilidad de sistema financiero, constituyéndose ambos en componentes vitales de la red de seguridad financiera. La acción reguladora del Estado en la actividad del sector financiero debe establecer un conjunto de principios y de normas relativas a la constitución de empresas que operan en este sector, ordenando el acceso al mercado y fijando los requisitos que deben cumplir para desarrollar sus operaciones. La actividad de intermediación financiera sólo puede ser ejercida por entidades debidamente autorizadas por el Estado. Por tanto, las personas o empresas que realicen la captación de depósitos, bajo cualquier modalidad, sin la debida autorización del Estado, infringen disposiciones legales y exponen a la población a la pérdida de sus ahorros.

3.2.3. Reseña histórica.

La historia de la supervisión y fiscalización de la actividad financiera en Bolivia está indisolublemente asociada a la evolución económica del país. Desde el origen de la República, y aún en períodos anteriores, el sistema financiero ha estado íntimamente ligado a la actividad económica y a los modelos o paradigmas de política económica. El sistema financiero formalizado tuvo su origen en los primeros bancos fundados a iniciativa privada, con atribuciones de entidades emisoras. El Estado autorizaba su funcionamiento y podía tomar las acciones para ejercer un derecho directo y mantener su control, además de utilizar los servicios de la banca para efectuar depósitos de las recaudaciones impositivas establecidas por Ley. Entre 1867 y 1871 el gobierno autorizó la creación de las tres primeras entidades financieras en el país: Banco Boliviano, Banco de Crédito Hipotecario de Bolivia y Banco Nacional de Bolivia.

Hasta 1871 los bancos en Bolivia desarrollaron sus actividades en virtud a disposiciones legales emitidas tanto por el H. Congreso Nacional como por el Poder Ejecutivo, autorizando el funcionamiento de cada banco mediante disposición legal expresa, debido a que no existía una norma especial para regular el sector en cuanto a su organización, funcionamiento y control. Posteriormente, mediante Ley de 17 de agosto de 1871 se estableció el primer fundamento de la legislación bancaria, disponiendo que el Estado era el único que podía autorizar la organización de los establecimientos de crédito con sujeción a los principios que rigen la materia, con atribuciones de emisión de moneda y, aunque en forma muy rudimentaria, se le facultaba para ejercer control sobre dicha emisión.

En septiembre de 1890, con la promulgación de la Ley de Bancos de Emisión y Comercio, se amplía las disposiciones de 1871 y se empezó a regular la creación y el funcionamiento de este tipo de instituciones. Esta Ley sistematizó una serie de medidas que venían aplicándose y estableció nuevas disposiciones que incidieron en el mercado financiero de la época. La Ley determinó capitales mínimos de constitución, depósitos de seriedad institucional, límites para la emisión de billetes con respecto al capital de la institución financiera, determinando que sería el Poder Ejecutivo quien conozca las solicitudes de establecimiento de Bancos de Emisión Asimismo, se nombró por primera vez el Inspector General, dependiente del Ministerio de Hacienda, con funciones permanentes de fiscalizar las entidades bancarias, incluyendo la fijación de su presupuesto, a partir de contribuciones de los bancos autorizados.

El 1° de noviembre de 1891 se promulgó una Ley que estableció los procedimientos para efectuar las inspecciones delimitando de esta manera las atribuciones del Inspector General. Otra disposición legal importante fue la Ley de 20 de noviembre de 1895, referida al procedimiento de liquidación administrativa cuando el activo no resultase suficiente para cumplir con sus obligaciones. Todas las disposiciones legales emitidas con carácter posterior, fueron complementando las ya existentes, entre ellas la Ley de 31 de marzo de 1900, emitida con el propósito de proteger la confianza del público que dejaba en manos de los bancos sus dineros en depósito, para lo cual determinó la constitución de reservas para casos de crisis, deficiencias de encaje, colocaciones e inversiones.

La dinámica del sistema financiero hizo necesario considerar la introducción de reformas al ordenamiento legal financiero. En 1928, el gobierno de Hernando Siles determinó la contratación de un grupo de expertos extranjeros que integraban la Misión Kemmerer, con el objeto de aportar en la creación de diversas leyes en el país tanto en materia financiera y tributaria, como de aduanas. En cuanto al sistema financiero, la Misión propuso tres leyes que incidieron en su funcionamiento: la Ley Monetaria, la Ley de Reorganización del Banco de la Nación Boliviana, que lo transformó en el Banco Central de la Nación Boliviana, y la Ley General de Bancos N° 608 de 11 de julio de 1928.

La Ley General de Bancos de 1928, mostraba características comunes con la actual legislación, enfocándose principalmente en la protección al ahorrista. Se consideraba que los individuos, al depositar dinero en las instituciones bancarias, tenían interés en el desenvolvimiento de éstas. Así, el desempeño financiero de los bancos era de interés público y debía ser regulado, motivo por el cual se creó la Superintendencia de Bancos, organismo que de acuerdo a esta Ley tendría los siguientes deberes:

- Hacer cumplir las leyes y decretos reglamentarios relativos a bancos.
- Vigilar e intervenir en la emisión e incineración de billetes y letras hipotecarias.

Se determinó además la atribución de la Superintendencia de realizar inspecciones en los bancos y las reglas a seguir en casos de liquidación voluntaria o forzosa. Durante la década de los cuarenta fue ampliándose el rango de acción de la Superintendencia de Bancos, habilitándole la regulación del sistema provisional. De esta manera, la Superintendencia regulaba el desempeño de las instituciones financieras de crédito, las compañías de seguro y las cajas jubilatorias.

En 1970 se aprobó la Ley del Sistema Financiero Nacional, con el propósito de unificar el sistema financiero nacional y contar con un instrumento que garantizase la ejecución coordinada de la política financiera y monetaria. Con la Ley del Sistema Financiero Nacional, la Superintendencia de Bancos, con todas las funciones y atribuciones, se incorporó al Banco Central de Bolivia, bajo la denominada División de Fiscalización.

Posteriormente, mediante Decreto Supremo N°21660, se dispuso que la Superintendencia de Bancos reasuma las funciones otorgadas en la Ley General de Bancos de 1928, como institución independiente del Banco Central de Bolivia. La disposición de restituir la Superintendencia de Bancos, separándola de la estructura orgánica del Banco Central de Bolivia, tuvo el objetivo de estructurar una supervisión más fortalecida. Hasta entonces, se había observado un rezago en el ámbito normativo prudencial y debilidades supervisoras que se reflejaron en diferentes crisis de instituciones bancarias. El marco legal vigente hasta 1993, aplicable a la actividad del sistema de intermediación financiera en Bolivia, estuvo contenido en la Ley General de Bancos de 1928 y una serie de disposiciones dispersas. El cambio de orientación de las disposiciones legales vigentes en el país hasta ese año, se efectivizó con la aprobación de la Ley N° 1488 de 14 de abril de 1993, reflejando una nueva perspectiva de modelo financiero, que contrasta con la antigua Ley General de Bancos de 1928.

3.2.4. La Autoridad de Supervisión Sistema Financiero – ASFI.

La Autoridad de Supervisión del Sistema Financiero (ASFI), es una institución de derecho público y de duración indefinida, con personalidad jurídica, patrimonio propio y autonomía de gestión administrativa, financiera, legal y técnica, con jurisdicción, competencia y estructura de alcance nacional, bajo tuición del Ministerio de Economía y Finanzas Públicas, y sujeta a control social.

El artículo 16° del Capítulo IV de la Ley N° 393 de Servicios Financieros (LSF) dispone que ASFI tiene por objeto regular, controlar y supervisar los servicios financieros en el marco de la Constitución Política del Estado, la Ley de Servicios Financieros y los Decretos Supremos reglamentarios, así como la actividad del mercado de valores, los intermediarios y entidades auxiliares del mismo. Las actividades financieras y la presentación de servicios financieros, serán realizadas únicamente por entidades autorizadas por la Autoridad de Supervisión del Sistema Financiero (ASFI), según los tipos de entidad financiera que la LSF define.

El artículo 137° del Decreto Supremo N° 29894 de 7 de febrero de 2009, que define la Estructura Orgánica del Órgano Ejecutivo del Estado Plurinacional, dispone: "… la Superintendencia de Bancos y Entidades Financieras se denominará Autoridad de Supervisión del Sistema Financiero de Bolivia y asumirá además las funciones y atribuciones de control y supervisión de las actividades económicas de valores y seguros…". Por su parte, el artículo 34° del Decreto Supremo N° 0071 de 9 de abril de 2009, que establece el proceso de extinción de las superintendencias generales y sectoriales, en su inciso b) manifiesta: "Las atribuciones, competencias, derechos y obligaciones en materia de valores y seguros de la Superintendencia de Pensiones, Valores y Seguros, establecidos en la norma vigente, serán asumidos por la Autoridad del Sistema Financiero, en todo lo que no contravenga a la CPE"

3.2.5. Banco Central De Bolivia.

El 20 de julio de 1928 constituye la fecha fundacional de lo que hoy es el Banco Central de Bolivia. En ese entonces, la Ley 632 del gobierno del Presidente Hernando Siles determinó la creación del Banco Central de la Nación Boliviana.

"Se declara Ley de la República el proyecto enviado por el Poder Ejecutivo, en fecha 4 del mes en curso, sobre el establecimiento del Banco Central de la Nación Boliviana, en sus noventa artículos, quedando así reorganizado el Banco de la Nación Boliviana" decía el texto del Artículo Único de la Resolución del Congreso Nacional, promulgada por el Poder Ejecutivo.

Pero desde el 20 de julio de 1928 pasarían aún varios meses hasta que el Banco Central inaugure sus actividades y adopte el nombre definitivo de Banco Central de Bolivia. A través de una modificación a la Ley de Bancos, el 20 de abril de 1929, el gobierno del Presidente Hernando Siles promulgo una Ley en la que se resolvía que, desde esa fecha en adelante, el nuevo Banco se denominaría Banco Central de Bolivia.

3.2.5.1. Funciones del Banco Central de Bolivia.

El Banco Central es una institución del Estado que ejecuta sus funciones de manera autárquica, con la potestad de tomar decisiones propias, dirigidas a mantener el poder adquisitivo de la moneda nacional.

Es la única autoridad que define la política monetaria y cambiaria del país. Estos dos instrumentos se complementan y son las herramientas principales del Banco.

La política monetaria se entiende como el instrumento que regula la circulación del dinero que requiere la política económica del gobierno. La política cambiaria, en tanto, es un instrumento que ayuda a mantener la estabilidad del poder adquisitivo de la moneda nacional y apoya al normal funcionamiento de los pagos internacionales de Bolivia.

- El BCB es la única autoridad monetaria.
- Administrar el Sistema de Pagos.
- Administrar las Reservas Internacionales.
- Definir el Régimen Cambiario.
- Funciones con relación al Sector Público.
- Agente Financiero del Gobierno.
- Funciones con relación al Sistema Financiero.

3.3. DESCRIPCIÓN DE UN SISTEMA DE COLAS.

Un sistema de colas se puede describir como sigue. Un conjunto de “clientes” llega a un sistema buscando un servicio, esperan si este no es inmediato, y abandonan el sistema una vez han sido atendidos. En algunos casos se puede admitir que los clientes abandonan el sistema si se cansan de esperar.

El término “cliente” se usa con un sentido general y no implica que sea un ser humano, puede significar piezas esperando su turno para ser procesadas o una lista de trabajo esperando para imprimir en una impresora en red.

La teoría de colas fue originariamente un trabajo práctico. La primera aplicación de la que se tiene noticia es del matemático danés Erlang sobre conversaciones telefónicas en 1909, para el cálculo de tamaño de centralitas. Después se convirtió en un concepto teórico que consiguió un gran desarrollo, y desde hace unos años se vuelve a hablar de un concepto aplicado aunque exige un importante trabajo de análisis para convertir las fórmulas en realidades, o viceversa.

Un sistema de cola básico[4] aunque la mayor parte de los sistemas se puedan representar como en la figura 1, debe quedar claro que una representación detallada exige definir un número elevado de parámetros y funciones.

3.3.1. Componentes de un sistema de colas.

Seis son los componentes básicos que se deben utilizar para describir adecuadamente un sistema de colas:

a) Patrón de llegada de los clientes.
b) Patrón de servicio de los servidores.
c) Disciplina de cola.
d) Capacidad del sistema.
e) Número de canales de servicio.
f) Número de etapas de servicio

Algunos autores incluyen una séptima característica que es la población de posibles clientes.

a) Patrón de llegada de los clientes:

En situaciones de cola habituales, la llegada es estocástica, es decir la llegada depende de una cierta variable aleatoria, en este caso es necesario conocer la distribución probabilística entre dos llegadas de cliente sucesivas. Además habría que tener en cuenta si los clientes llegan independiente o simultáneamente. En este segundo caso (es decir, si llegan lotes) habría que definir la distribución probabilística de éstos.

También es posible que los clientes sean “impacientes”. Es decir, que lleguen a la cola y si es demasiado larga se vayan, o que tras esperar mucho rato en la cola decidan abandonar.

Por último es posible que el patrón de llegada varíe con el tiempo. Si se mantiene constante le llamamos estacionario, si por ejemplo varía con las horas del día es no-estacionario.

[4] Anexos figura 1 sistema de colas basico

b) Patrones de servicio de los servidores:

Los servidores pueden tener un tiempo de servicio variable, en cuyo caso hay que asociarle, para definirlo, una función de probabilidad. También pueden atender en lotes o de modo individual.

El tiempo de servicio también puede variar con el número de clientes en la cola, trabajando más rápido o más lento, y en este caso se llama patrones de servicio dependientes. Al igual que el patrón de llegadas el patrón de servicio puede ser no estacionario, variando con el tiempo transcurrido.

c) Disciplina de cola:

La disciplina de cola es la manera en que los clientes se ordenan en el momento de ser servidos de entre los de la cola. Los clientes son seleccionados de la línea de espera mediante algún mecanismo, pudiendo ser el orden de selección alguno de los siguientes, que son los más habituales, o algún otro:

- FIFO: (first in first out): Primero en llegar primero en ser servido.
- LIFO: (Last in first out): Último en llegar primero en ser servido.
- SIRO: (Service in random order): Los clientes son servidos en orden aleatorio.
- PRI: (Priority): los clientes tienen distinta prioridad.

En cualquier caso dos son las situaciones generales en las que trabajar. En la primera, llamada en inglés “preemptive”, si un cliente llega a la cola con una orden de prioridad superior al cliente que está siendo atendido, este se retira dando paso al más importante. Dos nuevos subcasos aparecen: el cliente retirado ha de volver a empezar, o el cliente retorna donde se había quedado. La segunda situación es la denominada “no-preemptive” donde el cliente con mayor prioridad espera a que acabe el que está siendo atendido.

d) Capacidad del sistema:

En algunos sistemas existe una limitación respecto al número de clientes que pueden esperar en la cola. A estos casos se les denomina situaciones de cola finitas. Esta limitación puede ser considerada como una simplificación en la modelización de la impaciencia de los clientes.

e) Número de canales del servicio:

Es evidente que es preferible utilizar sistemas multiservidos con una única línea de espera para todos que con una cola por servidor. Por tanto, cuando se habla de canales de servicio paralelos, se habla generalmente de una cola que alimenta a varios servidores mientras que el caso de colas independientes se asemeja a múltiples sistemas con sólo un servidor.

A continuación se presenta dos variantes de sistema multicanal
Tiene una sola cola de espera[5].
Tiene una sola cola para cada canal[6].
Se asume que en cualquiera de los dos casos, los mecanismos de servicio operan de manera independiente.

f) Etapas de servicio:

Un sistema de colas puede ser unietapa o multietapa. En los sistemas multietapa el cliente puede pasar por un número de etapas mayor que uno. Una peluquería es un sistema unietapa, salvo que haya diferentes servicios (manicura, maquillaje) y cada uno de estos servicios sea desarrollado por un servidor diferente.

En algunos sistemas multietapa se puede admitir la vuelta atrás o "reciclado", esto es habitual en sistemas productivos como controles de calidad y reprocesos.

Sistema Multietapa con retroalimentación[7].

g) Resumen:

Las anteriores características bastan, de modo general, para describir cualquier proceso. Evidentemente se puede encontrar una gran cantidad de problemas distintos y, por tanto, antes de comenzar cualquier análisis matemático se debería describir adecuadamente el proceso atendiendo a las anteriores características.

Una elección equivocada del modelo lleva a unos resultados erróneos, y en muchos casos no analizar adecuadamente nos puede llevar a pensar que el sistema no es posible de modelar.

3.3. TERMINOLOGÍA Y NOTACIÓN.

En lo sucesivo utilizaremos las herramientas probabilísticas de los procesos de nacimiento y muerte para el estudio de colas con distribución del tiempo entre llegadas y distribución del tiempo de servicio exponencial. Antes hemos de fijar la notación que vamos a usar.

N(t): Denota el número de clientes en el sistema en el instante t. N(t) es un proceso estocástico en tiempo continuo y con espacios de estados discreto.

[5] Anexos Figura 2 sistema de colas de espera
[6] Anexos Figura 3 sistema de colas una cola para cada canal
[7] Anexos Figura 4 sistema de colas multietapa

Z(t): Denota el número de clientes que han salido del sistema hasta el instante t.

Nq(t): Representa el número de clientes en la cola en el instante t.

Pn(t): Es la probabilidad de que en el instante t, se encuentren n clientes en el sistema. Supongamos conocido el número de clientes en el instante cero (usualmente dicho número es cero).

C: Denota el número de servidores del mecanismo de servicio.

λn: Representa el número medio de llegadas de clientes al sistema, por unidad de tiempo, cuando ya hay n clientes en él. También se denomina tasa de llegadas (que se correspondería con la tasa de nacimientos si N(t) es un proceso de nacimiento y muerte).Cuando las tasas de llegada no dependen de n (es decir, todos los λn son constantes) suele denotarse como λ dicho valor constante.

µn: Es el número medio de clientes a los que se les completa el servicio, por unidad de tiempo, cuando hay n clientes en el sistema. Es frecuente referirse a los µn como tasas de complexión de servicio (o, simplemente, tasas de servicio). Si todos los servidores tienen la misma distribución del tiempo de servicio, suele denotarse por µ el número medio de clientes que puede atender cada servidor por unidad de tiempo. Como consecuencia se tiene que $\mu_n = n\mu$ si $n = 1,2,...,c$ y $\mu_n = c \cdot \mu$ para $n \geq c$.

ρ: Es la llamada constante de utilización del sistema o intensidad de tráfico. Se define, como ρ = λ c·µ Cuando los λn son constantes y todos los servidores tienen la misma distribución de tiempo de servicio, **λ:** es el número medio de clientes que entran en el sistema y c·µ es el número medio de clientes a los que pueden dar servicio los c servidores cuando todos están ocupados. En estas condiciones, ρ representa la fracción de recursos del sistema que es consumida por los clientes. Así, intuitivamente, parece necesario que se cumpla, en estos casos, que $\rho < 1$ y además cuanto más cercano a 1 sea ρ, más tráfico ha de soportar el sistema (o menos tiempo libre tendrán los servidores, o más espera habrán de sufrir los clientes, como se quiera expresar). Aunque es evidente que ρ no tiene unidades, es habitual medir la intensidad de tráfico en Erlangs, en honor a los trabajos pioneros de Erlang en la teoría de colas.

Los modelos de colas que estudiaremos en el siguiente capítulo son todos estacionarios. En ellos las distribuciones de probabilidad marginales de los procesos estocásticos $\{N(t)\}_{t\geq 0}$ y$\{Nq(t)\}_{t\geq 0}$ no cambian con el tiempo t. En tales condiciones tiene perfecto sentido definir los siguientes conceptos:

N: Es la variable aleatoria que contabiliza el número de clientes en el sistema.

Nq: Denota la variable aleatoria número de clientes en la cola.

S: Representa el tiempo de servicio.

Tq: Representa el tiempo que un cliente invierte en la cola.

T = Tq + S: El tiempo total que un cliente invierte en el sistema.

pn: Es la probabilidad de que se encuentren n clientes en el sistema para n = 0,1,....

L: Representa el número medio de clientes en el sistema, es decir L = E[N].

Lq: Es el número medio de clientes en la cola, o lo que es lo mismo, Lq = E[Nq].

W: Es la variable aleatoria que describe el tiempo que un cliente pasa en el sistema.

Wq: Es la variable aleatoria que representa el tiempo que un cliente espera en la cola.

W: Es el tiempo medio que un cliente está en el sistema, o simplemente, W = E[W].

Wq: Denota el tiempo medio de espera en la cola para un cliente genérico.

Matemáticamente, Wq = E[Wq]. Para clasificar los posibles tipos de sistemas de colas debemos especificar las características que determinan los elementos que lo componen. Así, Kendall introdujo en 1953 la notación A/B/c para indicar que la distribución del tiempo entre llegadas es de tipo A, que B es la distribución del tiempo de servicio y que c es el número de servidores. Posteriormente esta notación se extendió dando lugar a la más habitual en nuestros días, consistente en designar el sistema de una cola con la nomenclatura A/B/c/k/m/D, donde:

A: Distribución del tiempo entre llegadas consecutivas.

B: Alude al patrón del tiempo de servicio por parte de los servidores disponibles.

c: Es el número de canales de servicio.

k: Es la restricción en la capacidad de la cola.

m: Es el tamaño de la población potencial.

D: Es la disciplina de la cola.

Así, por ejemplo, la notación M/D/2/∞/∞/FIFO indica que se trata del sistema de una cola con tiempo entre llegadas exponenciales, tiempo de servicio determinístico (i.e. siempre se tarda el mismo tiempo en darle servicio a cada cliente), hay 2 servidores en el mecanismo de servicio, no existe límite para el número de clientes que pueden estar en la cola de espera, la población potencial se supone con infinitos clientes y los clientes son atendidos según una disciplina FIFO. Como los tres últimos valores ∞/∞/FIFO son precisamente los asignados por defecto, la notación anterior podría abreviarse como M/D/2. Obsérvese que este tipo de abreviaturas sólo se pueden realizar si todos los parámetros a partir de uno dado son iguales a los valores por defecto, ya que en caso contrario se produciría ambigüedad. Así, el modelo E2/U/3/∞/4/FIFO, no podría abreviarse como E2/U/3/4/FIFO (aunque sí como E2/U/3/∞/4), ya que, aunque es claro que $A = E2$, $B = U$, $c = 3$ y $Z = FIFO$, nunca sabríamos si con él pretendemos indicar $k = 4$ y $m = \infty$ ó bien $m = 4$ y $k = \infty$.

3.3.1. Fórmulas de Little.

En los modelos con distribución entre llegadas y distribución del servicio exponencial (así como en muchos otros modelos más generales llamados ergódicos) se verifican ciertas fórmulas que relacionan los números medios de clientes en el sistema o en la cola, con los tiempos medios de un cliente en el sistema o en la cola. Estas son las llamadas fórmulas de Little. Cuando las tasas de llegada son constantes (es decir; $\lambda n = \lambda$ para todo $n = 0,1,...$), la primera fórmula de Little establece la siguiente igualdad:

$$\mathbf{L = \lambda \cdot W}$$

Mientras que la segunda se expresa mediante

$$\mathbf{Lq = \lambda \cdot Wq}$$

Realmente sólo la primera de ellas fue obtenida por Little en 1961, pero es costumbre referirse a ambas con el término primera y segunda fórmula de Little. Una forma intuitiva de entender el porqué de la validez de las fórmulas de Little es la siguiente: considérese un cliente que llega al sistema justo ahora. Después de un tiempo, cuya media es W, ese cliente saldrá servido del sistema. Como el número medio de clientes que llegan al sistema por unidad de tiempo es λ, el número medio de clientes que habrán llegado desde que nuestro cliente en cuestión entró en el sistema hasta que salió de él es $\lambda \cdot W$. Por otra parte, es obvio que dicho número medio de clientes es precisamente el número medio de clientes que hay en el sistema justo en el momento que sale del sistema nuestro cliente particular, es decir; L. Un razonamiento análogo es válido para la segunda fórmula de Little. Obviamente, las fórmulas de Little no pueden ser válidas si los λn no son constantes (¿qué sería λ entonces?), pero si pueden generalizarse a esa situación mediante:

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

$$Lq = \bar{\lambda} \cdot Wq$$

Siendo:

$$\bar{\lambda} = \sum_{\eta = 0}^{\infty} \lambda\eta p\eta$$

Otra relación importante (en este caso para relacionar W y Wq) es la dada por:

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

3.3.2. Procesos de nacimiento y muerte.

El proceso de Poisson (y en general, los procesos de conteo) son útiles para modelar situaciones en las cuales el objetivo es contabilizar el número de ocurrencias de cierto fenómeno (nacimientos en una población) hasta un instante t. Sin embargo, existen otros procesos considerados más generales, llamados procesos de nacimiento y muerte los cuales contemplan la posibilidad de que el número de individuos en la población pueda disminuir. Por ejemplo: es posible considerar el caso de contabilizar el número de individuos en la población cada vez que se produzca una muerte. Entonces, los procesos de nacimiento y muerte, además de generalizar el proceso de Poisson, permite que las tasas de nacimientos y muertes puedan depender del número de individuos en la población.

3.3.2.1. Definición del proceso de nacimiento y muerte.

Considérese un proceso estocástico $\{N(t)/t \geq 0\}$ con espacio de estados discretos: $E = \{0,1,2,...\}$ y supóngase que el proceso describe un sistema que diremos que se encuentra en el estado E_n en el instante t cuando $N_t = n$. Se dirá que el proceso estocástico es de nacimiento y muerte si existen sucesiones de números no negativos $\{\lambda n;\ n = 0,1,2,...\}$ y $\{\mu n;\ n = 1,2,...\}$ (llamadas tasas de nacimiento y muerte, respectivamente), tales que se verifican las siguientes propiedades

1. Los cambios de estado permitidos son de E_0 a E_1 y desde E_n a E_{n-1} o a E_{n+1}, para $n \geq 1$[8]

[8] Anexo Figura 5 proceso de nacimiento y muerte de un sistema de colas

2. Si el sistema se encuentra en el estado En en el instante t, entonces, la probabilidad de que entre t y t + Δt pase al estado E $_{n+1}$ es $\lambda_n \Delta t + o(\Delta t)$, y si n ≥ 1, la probabilidad de que pase a E $_{n-1}$ es $\mu_n \Delta t + o(\Delta t)$.

3. La probabilidad de que ocurra más de un cambio en un intervalo de tiempo entre t y t+Δt es o (Δt).

La función o(Δt), es la que representa el error debido al tamaño del intervalo Δt, siendo tan pequeña, cuanto menor es el intervalo y además cumpliendo la propiedad:

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

De esta propiedad se deduce que la función de distribución de probabilidad de los tiempos entre llegadas será continua en cero, es decir; la probabilidad de que el tiempo entre llegadas toma el valor cero, es considerado infinitesimal. Entonces obtener el proceso de Poisson, a partir del proceso de nacimiento y muerte implica considerar tasas de muertes iguales a cero y tasas de nacimientos constantes e iguales a λ. A continuación se pretende determinar la distribución de probabilidades asociadas a un proceso de nacimiento y muerte. Sea $P_n(t)$, la probabilidad de que el sistema se encuentre en el estado En el instante t, matemáticamente, $P_n(t) = P(N_t = n)$. Es importante considerar que en los procesos de nacimiento y muerte la posibilidad de transitar entre diferentes estados se puede explicar de la siguiente forma: Teniendo en cuenta que la probabilidad de que se produzcan dos o más sucesos en un intervalo pequeño de tiempo es despreciable (o(Δt)), si en el instante t+Δt hay n individuos, sólo deben considerarse tres posibilidades:

- En el instante t habían n−1 individuos y entre t y t+Δt hubo un nacimiento y ninguna muerte.
- En el instante t habían n individuos y entre t y t+Δt no hubo ni nacimientos, ni muertes.
- En el instante t habían n + 1 individuos y entre t y t + Δt hubo una muerte y ningún nacimiento.

De acuerdo con la regla de las probabilidades totales y con ayuda de la figura 5, se logra expresar de manera sencilla, la probabilidad de que el sistema esté en el estado E_n en el instante t + ∆t, dado los posibles estados que puede tomar el sistema en el instante t. Así, si n ≥ 1, se tiene que:

$$
\begin{aligned}
P_n(t+\Delta t) &= P(N_{t+\Delta t}=n/N_t=n-1)P(N_t=n-1)+P(N_{t+\Delta t}=n/N_t=n)P(N_t=n)\\
&+ P(N_{t+\Delta t}=n/N_t=n+1)P(N_t=n+1)+o(\Delta t)\\
&= P_{n-1}(t)\left[\lambda_{n-1}\Delta t+o(\Delta t)\right]\left[1-\mu_{n-1}\Delta t+o(\Delta t)\right]\\
&+ P_n(t)\left[1-\lambda_n\Delta t+o(\Delta t)\right]\left[1-\mu_n\Delta t+o(\Delta t)\right]\\
&+ P_{n+1}(t)\left[1-\lambda_{n+1}\Delta t+o(\Delta t)\right]\left[\mu_{n+1}\Delta t+o(\Delta t)\right]+o(\Delta t)\\
&= P_{n-1}(t)\left[\lambda_{n-1}\Delta t-\lambda_{n-1}\mu_{n-1}(\Delta t)^2\right]\\
&+ P_n(t)\left[1-\mu_n\Delta t-\lambda_n\Delta t+\mu_n\lambda_n(\Delta t)^2\right]\\
&+ P_{n+1}(t)\left[\mu_{n+1}\Delta t-\mu_{n+1}\lambda_{n+1}(\Delta t)^2\right]+o(\Delta t)\\
&= P_{n-1}(t)\lambda_{n-1}\Delta t-P_{n-1}(t)\lambda_{n-1}\mu_{n-1}(\Delta t)^2+P_n(t)-P_n(t)\mu_n\Delta t-P_n(t)\lambda_n\Delta t\\
&+ P_n(t)\mu_n\lambda_n(\Delta t)^2+P_{n+1}(t)\mu_{n+1}\Delta t-P_{n+1}(t)\mu_{n+1}\lambda_{n+1}(\Delta t)^2+o(\Delta t)
\end{aligned}
$$

Si se resta Pn(t) y se divide por ∆t se tiene que,

$$
\begin{aligned}
\frac{P_n(t+\Delta t)-P_n(t)}{\Delta t} &= P_{n+1}(t)(\mu_{n+1}-\lambda_{n+1}\mu_{n+1}\Delta t)+P_{n-1}(t)(\lambda_{n-1}-\lambda_{n-1}\mu_{n-1}\Delta t)\\
&- P_n(t)(\lambda_n+\mu_n-\lambda_n\mu_n\Delta t)+\frac{o(\Delta t)}{\Delta t}
\end{aligned}
$$

Aplicando límite cuando $\Delta t \to 0^+$

$$P'_n(t)=\lambda_{n-1}P_{n-1}(t)-(\lambda_n+\mu_n)P_n(t)+\mu_{n+1}P_{n+1}(t)$$

y

$$P'_0(t)=-\lambda_0P_0(t)+\mu_1P_1(t)$$

Supongamos que al inicio del tiempo el sistema se encuentra en el estado E_0 (es decir; en t = 0, no hay individuos en la población), se tienen condiciones iniciales $P_0(0) = 1$ y $P_n(0) = 0, \forall n \geq 1$. En general, las ecuaciones de balance son difíciles de resolver. De todas formas hay algunos casos particulares en los que la resolución es más sencilla.

Así, si $\mu_n = 0$ ∀n = 1,2,... y $\lambda_n = \lambda$ ∀n = 0,1,... (Es decir; para el proceso de Poisson), las ecuaciones de balance resultan especialmente sencillas:

$$P'_n(t) = \lambda P_{n-1}(t) - \lambda P_n(t),\ n \geq 1$$

$$P'_0(t) = -\lambda P_0(t)$$

Y puede probarse sin excesiva dificultad que la solución es:

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!},\ \forall t \geq 0.$$

De hecho, para n = 0, la ecuación diferencial $P_0(t) = -\lambda P0(t)$ esta expresión es muy fácil de resolver, siendo su solución general, de la forma $P_0(t) = Ce^{-\lambda t}$ pero bajo la condición inicial $P_0(0) = 1$ y procediendo, en el caso general por inducción en n. Cuando el proceso estocástico de nacimiento y muerte es estacionario, las funciones $P_n(t)$ son constantes pn, que no dependen de t, y por tanto, el sistema de ecuaciones diferencias de balance se convierte en un sistema de infinitas ecuaciones lineales:

$$0 = \lambda_{n-1}p_{n-1} - (\lambda_n + \mu_n)p_n + \mu_{n+1}p_{n+1} \quad si\ n \geq 1$$

y

$$0 = -\lambda_0 p_0 + \mu_1 p_1$$

Que también pueden expresarse como:

$$(\lambda_n + \mu_n)p_n = \lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1} \quad si\ n \geq 1$$

y

$$\lambda_0 p_0 = \mu_1 p_1$$

Una forma intuitiva de interpretar estas ecuaciones es la siguiente: para cada posible estado, n, el miembro de la izquierda representa la probabilidad de dicho estado multiplicada por la suma de las tasas correspondientes a las formas de salir de este estado hacia otro distinto. Los términos de la derecha de cada ecuación de balance expresan la suma de las probabilidades de aquellos estados desde los cuales se puede llegar al estado n en una sola transición, multiplicadas por las tasas correspondientes a dicha transición.

Así, si n ≥ 1, en el término de la izquierda se multiplica a p_n por la suma de las tasas λ_n, que corresponde al hecho de que se produzca un nacimiento cuando el sistema está en el estado n (pasando, por tanto a n + 1) y para el caso de μ_n, corresponde a que se produzca una muerte cuando hay n individuos en la población (pasando por consiguiente, a una población con n−1 individuos). Cuando n = 0 el razonamiento anterior es válido salvo en lo referente a μ_0, el cual no aparece, porque no es lógico considerar de que se produzca una muerte cuando no hay individuos en la población. Para los términos de la derecha, cuando n ≥ 1, se puede llegar a un estado n procedente del estado n−1 (en cuyo caso debe haber un nacimiento, con tasa λ_{n-1}) o bien del estado n + 1 (siempre que haya una muerte, con tasa μ_{n+1}). Obviamente, si n = 0, no aparecerá el término correspondiente a n−1.

En este caso el proceso de nacimiento y muerte es estacionario, resulta bastante sencillo resolver las ecuaciones de balance. Así, tomando la ecuación correspondiente a n = 0 se puede despejar p_1, obteniendo

$$p_1 = \frac{\lambda_0}{\mu_1} p_0$$

Además, puede probarse por inducción una generalización de esta expresión para cualquier índice n:

$$p_n = \frac{\lambda_{n-1}}{\mu_n} p_{n-1}.$$

En efecto, la expresión es cierta para n = 1. Supongamos que también es cierta para n y probemos para n + 1. Utilizando la ecuación de balance n-ésima, se tiene que

$$\frac{(\lambda_n + \mu_n)p_n}{\mu_n} = \frac{\lambda_{n-1} p_{n-1}}{\mu_n} + \frac{\mu_{n+1} p_{n+1}}{\mu_n}$$

Y gracias a las hipótesis de inducción, puede escribirse como:

$$\left(\frac{\lambda_n}{\mu_n} + 1\right) p_n = p_n + \frac{\mu_{n+1} p_{n+1}}{\mu_n},$$

Entonces,

$$\frac{\lambda_n}{\mu_n} p_n = \frac{\mu_{n+1} p_{n+1}}{\mu_n},$$

Si simplificamos μn y despejamos p_n+1, llegamos a lo que se quería demostrar,

$$p_{n+1} = \frac{\lambda_n}{\mu_{n+1}} p_n.$$

De esta forma se tiene:

$$p_1 = \frac{\lambda_0}{\mu_1} p_0$$

$$p_2 = \frac{\lambda_1}{\mu_2} p_1 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} p_0$$

Y en general:

$$p_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n \mu_{n-1}\cdots\mu_1} p_0, \ \forall n = 1, 2, \ldots$$

Si denotamos al cociente como c_n, entonces:

$$c_n = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\mu_n \mu_{n-1}\cdots\mu_1}, \ \forall n = 1, 2, \ldots$$

Se tiene que $p_n = c_n p_0$. Ahora, como las p_n son probabilidades, se puede verificar que:

$$\begin{aligned}
\sum_{n=0}^{\infty} p_n &= 1 \\
&= p_0 + \sum_{n=1}^{\infty} p_n \\
&= p_0 + \sum_{n=1}^{\infty} c_n p_0 \\
1 &= \left(1 + \sum_{n=1}^{\infty} c_n\right) p_0
\end{aligned}$$

Por tanto podemos calcular p_0, de la siguiente manera:

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

Ahora, siempre que $\sum_{n=1}^{\infty} cn < \infty$. Esta última expresión es llamada condición de estado estacionario, esto viene a significar que la existencia de un proceso de nacimiento y muerte estacionario con tasas de nacimiento $\{\lambda_n\}_{n\geq 0}$ y tasas de muerte $\{\mu n\}_{n\geq 1}$ debe de cumplirse siempre que $\sum_{n=1}^{\infty} cn < \infty$.

3.3.3. El modelo de colas determinista d/d/1.

Los problemas más simples en teoría de colas son aquellos que no necesitan de distribuciones de probabilidad para ser descritos. Este es el caso de la cola determinística D/D/1, en la que las llegadas y los tiempos de servicio son constantes y existe un único servidor. Dado el cliente que hace el número k de los que han entrado en el sistema definimos:

t_k = Instante de llegada del cliente k-ésimo.

T_k = Tiempo trancurrido entre dos llegadas ($t_k - t_k - 1$).

W_k = Tiempo de espera en cola.

S_k = Tiempo de servicio.

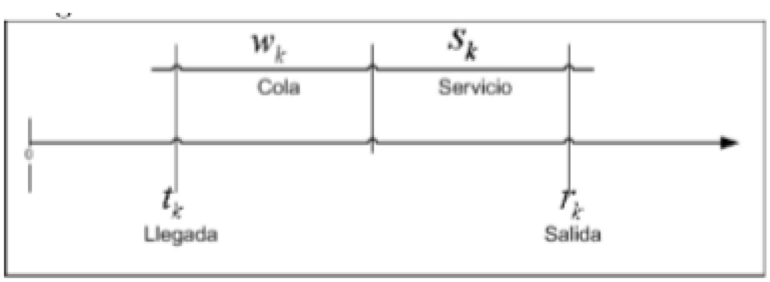
r_k = Tiempo de partida.

De lo anterior podemos verificar claramente que:

$$r_k = t_k + w_k + s_k \quad (5.1)$$

Lo cual podemos ver claramente en la figura 5.1.

Figura 5.1: Ilustración de la salida del k-ésimo cliente.



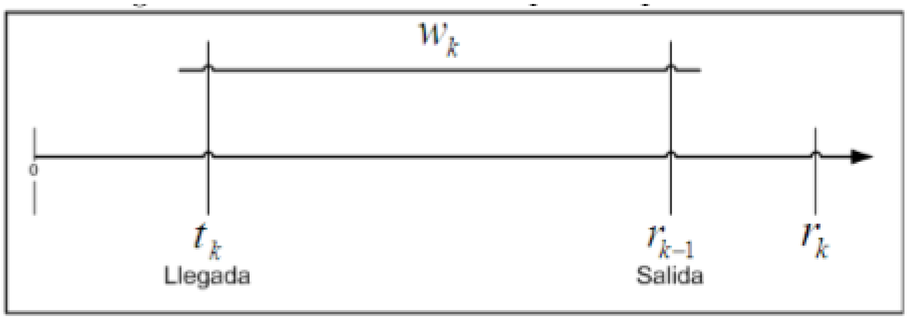
Si el sistema tiene un único canal c = 1, se tiene:

$$w_k = \begin{cases} r_{k-1} - t_k & , n_k > 0 \\ 0 & , n_k = 0 \end{cases} \quad (5.2)$$

Donde n_k es el número de clientes en el sistema cuando llega el cliente k-ésimo, es decir si $n_k = 0$ no hay clientes en el sistema, el tiempo de espera para el k-ésimo cliente será 0, de lo contrario el tiempo en cola del k-ésimo será igual al tiempo que tarde el cliente k−1 en salir del sistema menos el instante de llegada del k-ésimo cliente.

Una observación importante es que conocidos los tk y los sk, podemos conocer los wk por un proceso que reconstruye la historia de la cola, el cual veremos en el transcurso de esta sección. Supondremos que en el instante inicial t = 0 hay i clientes en cola. Si los clientes llegan a un único canal, a intervalos regulares de tiempo de longitud b (Tasa de servicio 1/b), las variables que definen el problema son:

Figura 5.2: Ilustración del tiempo de espera en cola.



$$\begin{aligned} \tau_k &= \text{a} \\ s_k &= \text{b}, \forall k \\ c &= 1 \end{aligned}$$

$$t_k = \begin{cases} 0 & ; k \leq i \\ (k-i)a & ; k > i \end{cases}$$

En primer lugar determinaremos el tiempo de espera en cola.

Proposición 1. Si se tienen i clientes en cola cuando llega el cliente k-ésimo, y si el tiempo entre llegadas de dos clientes consecutivos es a y el tiempo de servicio de dichos clientes es b, entonces el tiempo en cola del cliente k-ésimo es:

$$w_k = \begin{cases} (k-1)b & ; k \leq i \\ 0 & ; k > i, n_k = 0 \\ (k-1)b - (k-i)a & ; k > i, n_k > 0 \end{cases}$$

Demostración. Sabemos que si $n_k > 0$ entonces $w_k = r_{k-1} - t_k$, siendo n_k el número de clientes en el sistema cuando llega el cliente k-ésimo, t_k el instante de llegada del cliente k y r_k el tiempo de partida del cliente (k−1)-ésimo.

Si $k \leq i$ (5.3)

$$r_{k-1} = \sum_{j=1}^{k-1} s_j = \sum_{j=1}^{k-1} b = (k-1)b,\ \ t_k = 0$$

Si $k > i$ (5.4)

$$r_{k-1} = \sum_{j=1}^{k-1} s_j = \sum_{j=1}^{k-1} b = (k-1)b,\ \ t_k = (k-i)a$$

Proposición 2. La relación entre los tiempos de espera en cola de los clientes k-ésimo y (k−1)-ésimo viene dada por:

$$w_k = w_{k-1} + b - a,\ \ n_k > 0 \quad (5.5)$$

Demostración.

$$\begin{aligned} w_k &= r_{k-1} - t_k \\ &= (t_{k-1} + w_{k-1} + s_{k-1}) - t_k \\ &= -(t_k - t_{k-1}) + s_{k-1} + w_{k-1} \\ &= w_{k-1} + b - a \end{aligned} \quad (5.6)$$

ya que:

$$\begin{aligned} a &= t_k - t_{k-1} \\ b &= s_{k-1} \end{aligned}$$

A partir de aquí distinguiremos dos casos:

Primer caso b > a.

Segundo caso b < a.

Si b = a, no habrá espera si el sistema comienza con una cola vacía, si no es así, la cola se mantendrá con longitud constante.

Primer caso b > a. En este caso el tiempo de espera es cada vez mayor y la cola crece indefinidamente.

Corolario. El tiempo de espera en cola cuando el tiempo de servicio es mayor que el tiempo entre llegadas viene dado por:

$$w_k = \begin{cases} (k-1)b & ; k \leq i \\ (k-1)(b-a) + (i-1)a & ; k > i \end{cases} \tag{5.7}$$

Donde (k−1) (b−a) crece si k crece.

Analicemos las variables asociadas a un instante[9].

Proposición 3. Sean X,Y y Z variables asociadas a un instante t.

$$i) \quad X(t) = i + \left[\frac{t}{a}\right] \quad = \textit{Número de clientes que han llegado al sistema hasta el instante } t.$$

$$ii) \quad Y(t) = \left[\frac{t}{b}\right] + 1 \quad = \textit{Número de clientes que han entrado en servicio hasta el instante } t.$$

$$iii) \quad Z(t) = \left[\frac{t}{b}\right] \quad = \textit{Número de clientes que han salido del sistema hasta el instante } t.$$

Demostración.

i) El número de clientes que han entrado en el sistema hasta el instante t, será igual a los i que hay al principio, más los que han llegado hasta el instante t sabiendo que lo hacen a intervalos de tiempo de longitud constante a.

ii) El número de clientes que han entrado en servicio hasta el instante t será igual a la tasa de servicio ($\frac{1}{b}$) por el tiempo t más el último cliente que ha entrado en servicio y todavía no ha salido.

iii) El número de clientes que han salido del sistema hasta el instante t será igual a los que han sido servidos Y (t) menos el último que entró y está siendo servido: Z(t) = Y (t)−1.

Proposición 4. El número de clientes que se encuentran en el sistema en el instante t viene dado por la siguiente expresión:

---

[9] Los corchetes [·] indican el mayor entero que es menor que la fracción.

$$N(t) = i + \left[\frac{t}{a}\right] - \left[\frac{t}{b}\right], \tag{5.8}$$

que crece con t.

Demostración.

$$\begin{aligned} N(t) &= X(t) - Z(t) \\ &= i + \left[\frac{t}{a}\right] - \left[\frac{t}{b}\right] \end{aligned}$$

Donde N(t) es el número de clientes en el sistema en ese momento.

Segundo caso b < a. En este caso el servicio es más rápido que la afluencia de clientes a la cola, con lo cual el tiempo de espera w_k, irá disminuyendo hasta hacerse $w_k = 0$. A partir de ese instante los clientes no tendrán que esperar en cola; es decir serán servidos directamente.

Proposición 5. El último cliente tal que $w_k > 0$ (es decir, tiene que esperar en cola) viene dado por:

$$K = \begin{cases} \left[\dfrac{ia-b}{a-b}\right] & ; ai - b \neq m(a-b),\ m \in \mathbb{N} \\ \left[\dfrac{ia-b}{a-b}\right] - 1 & ; caso\ contrario \end{cases} \tag{5.9}$$

Donde K es el k-ésimo cliente que tiene que esperar. Demostración. Retomando el resultado obtenido de la proposición 1.4.1 para la primera ecuación en, tenemos:

$$\begin{gathered} w_k = (k-1)b - (k-i)a > 0 \\ (k-1)b - (k-i)a > 0 \\ kb - b - ka + ia > 0 \\ k(b-a) - b + ia > 0 \\ k(a-b) + b - ia < 0 \\ k(a-b) < ai - b \\ k < \frac{ai-b}{a-b} \end{gathered} \tag{5.10}$$

(5.11)

Tomando ahora:

$$
\begin{aligned}
w_{k+1} = kb - (k+1-i)a \leq 0 \\
kb - (k+1-i)a \leq 0 \\
kb - ka - a + ia \leq 0 \\
k(b-a) + (i-1)a \leq 0 \\
k(a-b) - (i-1)a \geq 0 \\
k(a-b) \geq a(i-1)
\end{aligned} \tag{5.12}
$$

$$
\begin{aligned}
k \geq \frac{a(i-1)}{a-b} &= \frac{ai - a + b - b}{a-b} \\
&= \frac{ai-b}{a-b} - \frac{a-b}{a-b} \\
&= \frac{ai-b}{a-b} - 1
\end{aligned}
$$

$$
k \geq \frac{ai-b}{a-b} - 1
$$

$$
\begin{aligned}
&Así\ pues,\ K = \left[\frac{ai-b}{a-b}\right]\ si\ ai - b \neq m(a-b)\ \ con\ m \in \{1,2,3,...\} \\
&Si\ ai - b = m(a-b)\ \ con\ m \in \{1,2,3,...\}\ \ entonces\ K = \left[\frac{ai-b}{a-b}\right] - 1
\end{aligned} \tag{5.13}
$$

Corolario. El tiempo de espera en cola será:

$$
w_k = \begin{cases} (k-1)b & ; k \leq i \\ (k-1)b - (k-i)a & ; i < k \leq K \\ 0 & ; k > K \end{cases} \tag{5.14}
$$

A continuación calcularemos el instante T, en que desaparece la cola.

Proposición 6. Así pues el instante en que la cola desaparece está dado por la siguiente expresión:

$$
T = (K-1)b = \left[\frac{ia-b}{a-b} - 1\right]\ \ ; ai - b \neq m(a-b),\ \ m \in \mathbb{N} \tag{5.15}
$$

Demostración:

$$\begin{aligned} T &= t_K + w_K \\ &= (K-i)a + (K-1)b - (K-i)a = (K-1)b \\ &= \left[\frac{ia-b}{a-b} - 1\right] b \end{aligned}$$

Las variables asociadas a un instante verifican:

Proposición 7. El número de clientes que han entrado al sistema, que están en servicio y que han salido de él hasta el instante t son respectivamente:

$$X(t) = i + \left[\frac{t}{a}\right] \tag{5.16}$$

$$Y(t) = \begin{cases} \left[\frac{t}{b}\right] + 1 & ; t \leq T \\ K + \left[\frac{t-t_k}{a}\right] & ; t > T \end{cases} \tag{5.17}$$

$$Z(t) = \begin{cases} \left[\frac{t}{b}\right] & ; t \leq T \\ K + \left[\frac{t-t_k}{a}\right] - 1 & ; t > T, t \in [t_k + ja, t_k + ja + b) \\ K + \left[\frac{t-t_k}{a}\right] & ; t > T, t \in [t_k + ja + b, t_k + (j+1)a) \end{cases} \tag{5.18}$$

$Donde\ j = 0, 1, 2, \ldots$

Demostración: En el caso de la ecuación (5.16) se resuelve como en el caso i) de la proposición 1.4.3. Si $t < T$, (5.17) y (5.18) se razona de igual forma que la proposición 3, pero si $t > T$, razonamos sobre un ejemplo: Sea $t > T$ y $K = 5 \Rightarrow T = (K-1)b = 4b$

En $T = 4b$ han salido $K-1 = 4$ clientes y el K-ésimo cliente entra en servicio. A partir de T los clientes entran en servicio según llegan y el sistema permanece vacío hasta que llega el siguiente cliente.

Así pues, para $t > T$*, habrá entrado en servicio* $Y(t) = K + \left[\frac{t-t_k}{a}\right] = 5 + 3 = 8.$

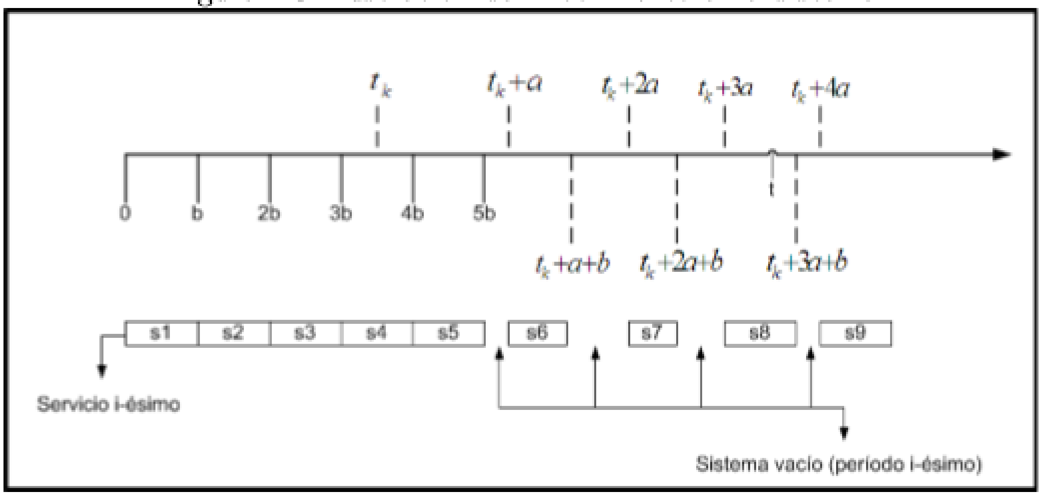
Además:

si $t \in [t_k + ja, t_k + ja + b) \Rightarrow Z(t) = Y(t) - 1 = 7$*, en el ejemplo.*

si $t \in [t_k + ja + b, t_k + (j+1)a) \Rightarrow Z(t) = Y(t).$

Es decir, si t está en un período vacío no hay que restar 1 porque no hay nadie en servicio.

Figura 1.3: Ilustración del movimiento en el sistema.



3.3.4. El modelo M/M/1.

Este modelo indica que el tiempo entre llegadas consecutivas de clientes al sistema es una distribución exponencial con parámetro λ, independiente del número de clientes que haya dentro del sistema, además que los tiempos entre servicio de los clientes también están distribuidos exponencialmenteconparámetrosμyunsoloservidor.Losvaloresdelosúltimostresparámetros según la notación Kendall son por defecto los siguientes:

- No hay restricciones respecto al número de clientes en cola.
- La población potencial es infinita.
- La disciplina de la cola es FIFO.

Por lo tanto es un sistema de espera de un solo recurso y denotamos por λ la tasa de llegadas al sistema y por μ la velocidad de servicio del recurso. Las tasas de llegada y de servicio son:

$$\lambda_n = \lambda, \forall n = 0,1,2,3,...$$

$$\mu_n = \mu, \forall n = 1,2,3,......$$

Nos interesa evaluar el sistema en régimen permanente, así que calculamos la distribución de probabilidad. Esta distribución se obtiene sustituyendo la tasa de llegadas y de servicios en la función:

$$p_n = \frac{\lambda_{n-1}}{\mu_n} p_{n-1} \tag{5.19}$$

Viene dada de la distribución de probabilidades en un proceso de nacimiento y muerte, al hacer iteraciones de p_n resulta:

$$p_n = p_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} \tag{5.20}$$

Además, siendo el sistema estable, p_n es una distribución de probabilidad, es decir se satisface que:

$$\sum_{n=0}^{\infty} p_n = 1 \tag{5.21}$$

Sea:

$$c_n = \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} \tag{5.22}$$

Dado la igualdad de los parámetros se tiene:

$$c_n = \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} = \frac{\lambda^n}{\mu^n} = \rho^n \tag{5.23}$$

Como $\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{\infty} \rho^n$ es una serie geométrica, será convergente si y sólo si $|\rho| < 1$, es decir; que $0 < \rho < 1$. Esta condición $0 < \rho < 1$, es por tanto equivalente a que el modelo sea estacionario. Otra forma de expresarla es $\lambda < \mu$, que tiene la interpretación adicional de que el número medio de clientes que entran en el sistema por unidad de tiempo sea menor que el número medio de clientes que podrían ser atendidos por el servidor por unidad de tiempo, en caso de que éste estuviese absolutamente todo el tiempo atendiendo a clientes (cosa que no ocurre siempre). En lo que sigue supondremos que el sistema de la cola es estacionario (es decir; que $\rho < 1$). Lo primero que debemos calcular es la suma de la serie c_n:

$$
\begin{aligned}
\sum_{n=1}^{\infty} c_n &= \sum_{n=1}^{\infty} \rho^n \\
\sum_{n=1}^{\infty} \rho^n &= \rho + \rho^2 + \rho^3 + \rho^4 + \cdots \\
&= \rho \left(1 + \rho + \rho^2 + \rho^3 + \rho^4 + \cdots\right) \\
\sum_{n=1}^{\infty} \rho^n &= \rho \left(1 + \sum_{n=1}^{\infty} \rho^n\right) \\
\sum_{n=1}^{\infty} \rho^n - \rho \sum_{n=1}^{\infty} \rho^n &= \rho \\
(1-\rho) \sum_{n=1}^{\infty} \rho^n &= \rho \\
\sum_{n=1}^{\infty} \rho^n &= \frac{\rho}{(1-\rho)}
\end{aligned}
\tag{5.24}
$$

Así de las ecuaciones anteriores podemos obtener el valor de p_0, donde:

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

$$
p_0 = \begin{cases} 0 & , \rho > 1 \Rightarrow p_n = 0, \forall n \text{ caso desbordado} \\ 1-\rho & , 0 < \rho < 1 \Rightarrow p_n = \rho^n(1-\rho), \forall\, n \end{cases}
\tag{5.25}
$$

En realidad, para que 0 < ρ < 1, lo que ha ocurrido es que λ < μ es decir, que el tiempo medio de llegadas sea menor que el tiempo medio de servicios. Esto es razonable si pensamos que para que exista una distribución estacionaria la longitud de la cola no puede tender a infinito, deduzcamos entonces p_0, para ello tenemos que:

$$
\begin{aligned}
1 &= \sum_{n=0}^{\infty} p_n = p_0 + \sum_{n=1}^{\infty} p_n \\
&= p_0 + \sum_{n=1}^{\infty} p_0\rho^n = p_0(1 + \sum_{n=1}^{\infty} \rho^n) \\
p_0 &= \frac{1}{1 + \sum_{n=1}^{\infty} \rho^n} = \frac{1}{1 + \frac{\rho}{1-\rho}} \\
p_0 &= \frac{1}{\frac{1-\rho+\rho}{1-\rho}} \\
p_0 &= 1 - \rho
\end{aligned}
\tag{5.26}
$$

De aquí se tiene la función de probabilidad:

$$
\begin{aligned}
p_n &= p_0\rho^n \\
p_n &= (1-\rho)\rho^n
\end{aligned}
$$

Resulta interesante observar que la probabilidad de encontrar k o más clientes en el sistema es[2]:

$$
\begin{aligned}
P(N \geq k) &= \sum_{i=k}^{\infty} p_i \\
&= \sum_{i=k}^{\infty} (1-\rho)\rho^i \\
&= (1-\rho)\sum_{i=k}^{\infty} \rho^i \\
&= (1-\rho)(\rho^k + \rho^{k+1} + \rho^{k+2} + \cdots) \\
&= (1-\rho)\rho^k(1 + \rho + \rho^2 + \cdots) \\
&= (\rho^k - \rho^{k+1})(1 + \rho + \rho^2 + \cdots) \\
&= \rho^k + \rho^{k+1} + \rho^{k+2} + \cdots - \rho^{k+1} - \rho^{k+2} \\
&= \rho^k
\end{aligned}
\tag{5.27}
$$

Una vez calculada la distribución de probabilidad del sistema, calcularemos los parámetros de interés desde el punto de vista del usuario:

a) Número medio de clientes en el sistema:

$$\begin{aligned}
L = E(N) &= \sum_{n=0}^{\infty} np_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n \\
&= (1-\rho)\sum_{n=0}^{\infty} n\rho^n \\
&= (1-\rho)\rho\sum_{n=0}^{\infty} n\rho^{n-1}; \quad \frac{d}{d\rho}(\rho^n) = n\rho^{n-1} \\
&= (1-\rho)\rho\sum_{n=0}^{\infty} \frac{d\rho^n}{d\rho} \\
&= (1-\rho)\rho\frac{d}{d\rho}\left(\sum_{n=0}^{\infty} \rho^n\right) \\
&= (1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) \\
L &= (1-\rho)\rho\frac{1}{(1-\rho)^2} \\
L &= \frac{\rho}{1-\rho}
\end{aligned} \tag{5.28}$$

b) Número medio de clientes en cola:

Puede obtenerse a partir de N_q de la siguiente manera:

$$\begin{aligned}
L_q = E[N_q] &= \sum_{n=1}^{\infty}(n-1)p_n \\
&= \sum_{n=1}^{\infty} np_n - \sum_{n=1}^{\infty} p_n = \sum_{n=0}^{\infty} np_n - (1-p_0) \\
&= L - \rho = \frac{\rho}{1-\rho} - \rho \\
L_q &= \frac{\rho^2}{1-\rho}
\end{aligned} \tag{5.29}$$

c) Tiempos medio de espera del cliente en el sistema:

Aplicando directamente la fórmula de Little tenemos:

$$W = \frac{L}{\lambda} = \frac{\frac{\rho}{1-\rho}}{\lambda} = \frac{\frac{\frac{\lambda}{\mu}}{1-\frac{\lambda}{\mu}}}{\lambda} = \left(\frac{1}{\mu-\lambda}\right)$$

(5.30)

d) Tiempo medio de clientes en cola:

Aplicando nuevamente la fórmula de Little encontramos:

$$W_q = \frac{L_q}{\lambda} = \frac{\frac{\rho^2}{1-\rho}}{\lambda} = \frac{\frac{(\frac{\lambda}{\mu})^2}{1-\frac{\lambda}{\mu}}}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)} \tag{5.31}$$

Además, puede comprobarse fácilmente que se verifica la relación W = W_q + $\frac{1}{\mu}$. De hecho esta relación junto con las dos fórmulas de Little permite calcular el valor de cualesquiera tres de las cantidades L, L_q, W y W_q, dada la cuarta. Si se desea tener más información sobre la espera de clientes en la cola o en el sistema, debe calcularse la distribución de probabilidad de las variables W y W_q. Estas distribuciones permitirán calcular la probabilidad de cualquier suceso relativo al tiempo de estancia en la cola o en el sistema.

Primeramente abordaremos el cálculo de la función de distribución de W, que denotaremos por W(t). Para ello aplicamos la regla de las probabilidades totales condicionando al número, N, de clientes que hay en el sistema cuando llega el cliente en cuestión y tenemos en cuenta que.

$$\mathbf{W}|_{N=n} \stackrel{d}{=} \Gamma(\mu, n+1)$$

De esta forma, para cada t ≥ 0, se tiene:

$$\begin{aligned}
\mathbf{W(t)} &= P[\mathbf{W} \leq t] = \sum_{n=0}^{\infty} P[\mathbf{W} \leq t|_{N=n}]P[N=n] \\
&= \sum_{n=0}^{\infty} \left(\int_0^t \frac{\mu^{n+1}}{n!} x^n e^{-\mu x} dx \right) p_n \\
&= \sum_{n=0}^{\infty} \left(\int_0^t \frac{\mu^{n+1}}{n!} x^n e^{-\mu x} dx \right) \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n \\
&= \int_0^t \mu \left(1 - \frac{\lambda}{\mu}\right) e^{-\mu x} \left[\sum_{n=0}^{\infty} \frac{(\lambda x)^n}{n!} \right] dx \\
&= \int_0^t \mu \left(1 - \frac{\lambda}{\mu}\right) e^{-\mu x} e^{\lambda x} dx \\
&= \int_0^t (\mu - \lambda) e^{-(\mu-\lambda)x} dx \\
&= \left[-e^{-(\mu-\lambda)x} \right]_{x=0}^{x=t} \\
&= 1 - e^{-(\mu-\lambda)t}
\end{aligned}$$

(5.32)

Que es la función de distribución de una exponencial de parámetro μ−λ. Así pues,

$$\mathbf{W} \stackrel{d}{=} exp(\mu - \lambda)$$

Es obvio, por tanto, que como conclusión de esto también se puede volver a obtener W = $\frac{1}{\mu-\lambda}$.

Dado que la distribución del tiempo de servicio para el sistema M/M/1 es exponencial y esta no tiene memoria, entonces el valor del tiempo que le queda por servir al cliente que está siendo atendido tendrá la misma distribución que el tiempo de servicio demandado.

Así W_q está condicionado a que hayan n tareas en el sistema por lo tanto, es una suma de n variables exponenciales independientes e idénticamente distribuidas; es decir es una variable aleatoria Erlang −n cuya función de distribución es:

$$\mathbf{W}_q|_n(t) = 1 - e^{-\mu t}\sum_{j=0}^{n-1}\frac{(\mu t)^j}{j!}, \quad \forall t \geq 0$$

Así la distribución de probabilidad marginal de W_q vendrá dada por:

$$\begin{aligned}
\mathbf{W}_q(t) &= P[\mathbf{W}_q \leq t] \\
&= \sum_{n=0}^{\infty} p_n P[\mathbf{W}_q \leq t|_{N=n}] \\
&= \sum_{n=0}^{\infty} p_n F_{\mathbf{W}_q|_n}(t)] \\
&= \sum_{n=0}^{\infty} p_n \left[1 - e^{-\mu t}\sum_{j=0}^{n-1}\frac{(\mu t)^j}{j!}\right] \\
&= p_0 \mathbf{W}_q|_0(t) + \sum_{n=1}^{\infty} p_n \left[1 - e^{-\mu t}\sum_{j=0}^{n-1}\frac{(\mu t)^j}{j!}\right] \\
&= (1-\rho) + \sum_{n=1}^{\infty} p_n - e^{-\mu t}\sum_{n=1}^{\infty} p_n \sum_{j=0}^{n-1}\frac{(\mu t)^j}{j!} \\
&= \sum_{n=0}^{\infty} p_n - e^{-\mu t}\sum_{n=1}^{\infty}(1-\rho)\rho^n \sum_{j=0}^{n-1}\frac{(\mu t)^j}{j!} \\
&= 1 - (1-\rho)e^{-\mu t}\sum_{n=1}^{\infty}\rho^n \sum_{j=0}^{n-1}\frac{(\mu t)^j}{j!}
\end{aligned}$$

(5.33)

Donde:

$$\sum_{n=1}^{\infty}\rho^{n}\sum_{j=0}^{n-1}\frac{(\mu t)^{j}}{j!}=\sum_{k=0}^{\infty}\frac{(\mu t)^{k}}{k!}\sum_{i=k+1}^{\infty}\rho^{i}$$

Así pues:

$$\begin{aligned}
\sum_{k=0}^{\infty}\frac{(\mu t)^{k}}{k!}\sum_{i=k+1}^{\infty}\rho^{i} &= \sum_{k=0}^{\infty}\frac{(\mu t)^{k}}{k!}\left(\rho^{k+1}+\rho^{k+2}+\cdots\right)\\
&= \sum_{k=0}^{\infty}\frac{(\mu t)^{k}}{k!}\rho^{k}\left(\rho+\rho^{2}+\rho^{3}+\cdots\right)\\
&= \sum_{k=0}^{\infty}\frac{(\mu t)^{k}}{k!}\rho^{k}\left[\sum_{i=1}^{\infty}\rho^{i}\right]\\
&= \sum_{k=0}^{\infty}\frac{(\mu t)^{k}}{k!}\rho^{k}\left[\frac{\rho}{1-\rho}\right]\\
&= \left[\frac{\rho}{1-\rho}\right]\sum_{k=0}^{\infty}\frac{(\mu t\rho)^{k}}{k!}\\
\sum_{k=0}^{\infty}\frac{(\mu t)^{k}}{k!}\sum_{i=k+1}^{\infty}\rho^{i} &= \left[\frac{\rho}{1-\rho}\right]e^{\mu t\rho}
\end{aligned}$$

(5.34)

Volviendo a (5.33) utilizando el resultado de (5.34) obtenemos:

$$\begin{aligned}
\mathbf{W}_{q}(t) &= 1-(1-\rho)e^{-\mu t}\left[\frac{\rho}{1-\rho}\right]e^{\mu t\rho}\\
&= 1-\rho e^{-\mu t(1-\rho)}\\
&= 1-\frac{\lambda}{\mu}e^{-t(\mu-\lambda)}
\end{aligned}$$

(5.35)

De esta forma se obtiene:

$$\mathbf{W}_{q}(t)=\begin{cases}1-\dfrac{\lambda}{\mu}e^{-(\mu-\lambda)t} & ,t\geq 0\\ 0 & ,t\leq 0\end{cases}$$

(5.36)

Esta función de distribución es discontinua en t = 0, valiendo su salto.

$$\mathbf{W}_q(0) - \mathbf{W}_q(0^-) = 1 - \frac{\lambda}{\mu} = 1 - \rho = p_0$$

Se trata por tanto de una variable que es mezcla de continua y discreta; toma el valor de 0 con probabilidad p_0 y para todos los t > 0 tiene una componente continua con función de subdensidad dada por: $\frac{\lambda(\mu-\lambda)}{\mu}$ $e^{-(\mu-\lambda)t}$

3.3.5. El modelo M/M/1/k.

Se trata de un modelo como el M/M/1, ya estudiado, pero con limitación k para el tamaño de la cola. Donde, la distribución del tiempo entre dos intentos de llegadas al sistema de clientes consecutivos es una exp (λ), la distribución del tiempo de servicio es exp (μ) y sólo hay un servidor.Ademáselnúmerodeclientesquepuedenestarenlacolaescomomucho k, la población potencial es infinita y la disciplina es FIFO. Obviamente, en este modelo se puede dar el caso de que un cliente que intente entrar en el sistema no lo consiga, por estar la cola llena. A partir de las especificaciones anteriores se deducen fácilmente las tasas de llegada:

$$\lambda_n = \begin{cases} \lambda & , n = 0, 1, 2, 3, 4, \ldots, k \\ 0 & , n = k+1, k+2, \ldots \end{cases} \tag{5.38}$$

Mientras que las tasas de servicio son idénticas a las de un M/M/1,

$$\mu_n = \mu, \forall n = 1, 2, \ldots$$

Haciendo uso de λ_n y μ_n se obtienen inmediatamente los c_n:

$$c_n = \begin{cases} \rho^n & ; si\ n = 0, 1, 2, 3, 4, \ldots k+1 \\ 0 & ; si\ n = k+2, k+3, \ldots \end{cases} \tag{5.39}$$

Dado que la serie $\sum_{n-1}^{\infty} c_n$ tiene tan sólo un número finito de términos distintos de cero, es trivialmente convergente sin ninguna condición acerca de ρ. Esto puede interpretarse que por muy frecuente que sea la llegada de clientes al sistema en relación con la capacidad del servidor para dar servicio, la propia limitación en el tamaño de la cola fuerza a la estacionariedad, pues lo peor que podríamos imaginar es que prácticamente todo el tiempo estuviera el sistema saturado (es decir; P(N = k + 1) = 1). La suma de la serie de los c_n es realmente una suma finita.

$$\sum_{n=1}^{\infty} c_n = \sum_{n=1}^{k+1} \rho^n = \begin{cases} \dfrac{\rho^{k+2} - \rho}{\rho - 1} & , \rho \neq 1 \\ k+1 & , \rho = 1 \end{cases} \tag{5.40}$$

En la ecuación (5.40), observamos que si el valor de ρ = 1 la expresión es inmediata, ya que ρ en la sumatoria sería constante y k +1 veces ρ = 1, ahora bien para el caso en que ρ 6= 1 vemos lo siguiente:

$$\begin{aligned} \sum_{n=1}^{k+1} \rho^n &= \rho + \rho^2 + \cdots + \rho^{k+1} \\ \rho \sum_{n=1}^{k+1} \rho^n &= \rho(\rho + \rho^2 + \cdots + \rho^{k+1}) \\ \rho \sum_{n=1}^{k+1} \rho^n + \rho &= \rho + \rho^2 + \rho^3 + \cdots + \rho^{k+1} + \rho^{k+2} \\ &= \sum_{n=1}^{k+1} \rho^n + \rho^{k+2} \\ \sum_{n=1}^{k+1} \rho^n (\rho - 1) &= \rho^{k+2} - \rho \\ \sum_{n=1}^{k+1} \rho^n &= \frac{\rho^{k+2} - \rho}{(\rho - 1)} \end{aligned} \tag{5.41}$$

Esta distinción, ρ ≠ 1 o ρ = 1 habrá que hacerla a lo largo de todos los cálculos sucesivos.

Caso ρ ≠ 1: En primer lugar calculamos p_0:

$$p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} c_n} = \frac{1}{1 + \frac{\rho^{k+2} - \rho}{\rho - 1}} = \frac{\rho - 1}{\rho^{k+2} - 1} \tag{5.42}$$

$$p_n = \begin{cases} \dfrac{\rho - 1}{\rho^{k+2} - 1}, & n = 0, 1, 2 \ldots, k+1 \\ 0, & n = k+2, k+3, \ldots \end{cases} \tag{5.43}$$

El número medio de clientes en el sistema puede calcularse a partir de su definición:

$$
\begin{aligned}
L &= \sum_{n=0}^{\infty} np_n \\
&= \sum_{n=0}^{k+1} n\frac{\rho-1}{\rho^{k+2}-1}\rho^n \\
&= \frac{(\rho-1)\rho}{\rho^{k+2}-1}\sum_{n=0}^{k+1} n\rho^{n-1}
\end{aligned}
\tag{5.44}
$$

Ahora bien, siguiendo las mismas pautas que las usadas en el M/M/1 en 5.24 para calcular la suma de una serie convertible en geométrica, podemos ahora calcular la suma de un número finito de términos de la misma. Así, definiendo:

$$
\begin{aligned}
\sum_{n=0}^{k+1}\rho^n &= \sum_{n=0}^{\infty}\rho^n - \sum_{n=k+2}^{\infty}\rho^n \\
&= 1+\frac{\rho}{1-\rho} - \sum_{n=k+2}^{\infty}\rho^n \\
&= \frac{1}{1-\rho} - (\rho^{k+2}+\rho^{k+3}+\cdots)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{1-\rho} - \rho^{k+2}(1+\rho+\rho^2+\cdots) \\
&= \frac{1}{1-\rho} - \rho^{k+2}\sum_{n=0}^{\infty}\rho^n \\
&= \frac{1}{1-\rho}(1-\rho^{k+2})
\end{aligned}
$$

Además sabemos que:

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

$$\begin{aligned}
\sum_{n=0}^{k+1} \frac{d}{d\rho}(\rho^n) &= \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)(1-\rho^{k+2}) \\
&= \frac{d}{d\rho}\left(\frac{1-\rho^{k+2}}{1-\rho}\right) \\
&= \frac{-(k+2)\rho^{k+1}+(k+2)\rho^{k+2}+1-\rho^{k+2}}{(1-\rho)^2} \\
&= \frac{(k+2-1)\rho^{k+2}-(k+2)\rho^{k+1}+1}{(1-\rho)^2} \\
&= \frac{(k+1)\rho^{k+2}-(k+2)\rho^{k+1}+1}{(1-\rho)^2}
\end{aligned} \tag{5.45}$$

Por lo tanto, de la expresión 5.45 podemos calcular L, de la siguiente forma:

$$\begin{aligned}
L &= \frac{(\rho-1)}{\rho^{k+2}-1}\cdot\left[\frac{k\rho^{k+3}+2\rho^{k+3}-(k+2)\rho^{k+2}+\rho-\rho^{k+3}}{(1-\rho)^2}\right] \\
&= \frac{(\rho-1)}{\rho^{k+2}-1}\cdot\left[\frac{(k+2)(\rho^{k+3}-\rho^{k+2})+\rho(1-\rho^{k+2})}{(1-\rho)^2}\right] \\
&= \frac{(\rho-1)}{\rho^{k+2}-1}\cdot\left[\frac{\rho(1-\rho^{k+2})-(k+2)(\rho^{k+2}-\rho^{k+3})}{(1-\rho)^2}\right]
\end{aligned}$$

$$\begin{aligned}
L &= \frac{(\rho-1)\rho}{\rho^{k+2}-1}\cdot\left[\frac{(k+1)\rho^{k+2}-(k+2)\rho^{k+1}+1}{(1-\rho)^2}\right] \\
&= \frac{(\rho-1)}{\rho^{k+2}-1}\cdot\left[\frac{(k+1)\rho^{k+3}-(k+2)\rho^{k+2}+\rho}{(1-\rho)^2}\right] \\
&= \frac{(\rho-1)}{\rho^{k+2}-1}\cdot\left[\frac{k\rho^{k+3}+\rho^{k+3}-(k+2)\rho^{k+2}+\rho+\rho^{k+3}-\rho^{k+3}}{(1-\rho)^2}\right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{(1-\rho)}{1-\rho^{k+2}}\cdot\left[\frac{\rho(1-\rho^{k+2})}{(1-\rho)^2}-\frac{(k+2)(\rho^{k+2}-\rho^{k+3})}{(1-\rho)^2}\right] \\
&= \frac{\rho}{1-\rho}-\frac{(k+2)\rho^{k+2}(1-\rho)^2}{(1-\rho)^2(1-\rho^{k+2})} \\
L &= \frac{\rho}{1-\rho}-\frac{(k+2)\rho^{k+2}}{(1-\rho^{k+2})}
\end{aligned} \tag{5.46}$$

El primer sumando de L es precisamente L del modelo M/M/1. Las fórmulas de Little y la relación entre tiempos medios pueden usarse para calcular las otras tres cantidades medias de interés. Para ello será necesario calcular λ, ya que ahora las λ_n no son constantes:

$$\begin{aligned}
\overline{\lambda} &= \sum_{n=0}^{\infty} \lambda \cdot p_n \\
&= \sum_{n=0}^{k} \lambda \cdot p_n \\
&= \lambda(1 - p_{k+1}); \quad p_n = p_0\rho^n \\
&= \lambda(1 - \frac{(\rho - 1)}{\rho^{k+2} - 1}\rho^{k+1}) \\
&= \lambda\frac{\rho^{k+2} - 1 - (\rho - 1)\rho^{k+1}}{\rho^{k+2} - 1} \\
&= \frac{\lambda(\rho^{k+1} - 1)}{\rho^{k+2} - 1}
\end{aligned} \tag{5.47}$$

A partir de esta expresión se tiene:

$$\begin{aligned}
W &= \frac{L}{\overline{\overline{\lambda}}} \\
&= \frac{\frac{(1-\rho)}{1-\rho^{k+2}} \cdot \frac{(k+1)\rho^{k+3}-(k+2)\rho^{k+2}+\rho}{(1-\rho)^2}}{\frac{\lambda(\rho^{k+1}-1)}{\rho^{k+2}-1}}
\end{aligned}$$

$$\begin{aligned}
W &= \frac{1-\rho}{1-\rho^{k+1}} \cdot \frac{(k+1)\rho^{k+3} - (k+2)\rho^{k+2} + \rho}{\lambda(1-\rho)^2} \\
&= \frac{1-\rho}{1-\rho^{k+1}} \cdot \frac{\rho - \rho^{k+2} - (k+1)(1-\rho)\rho^{k+2}}{\lambda(1-\rho)^2} \\
&= \frac{1-\rho}{1-\rho^{k+1}} \cdot \left[\frac{\rho(1-\rho^{k+1})}{\lambda(1-\rho)^2} - \frac{(k+1)(1-\rho)\rho^{k+2}}{\lambda(1-\rho)^2}\right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{\rho}{\lambda(1-\rho)} - \frac{(k+1)\rho^{k+2}}{\lambda(1-\rho^{k+1})} \\
W &= \frac{1}{\mu - \lambda} - \frac{(k+1)\rho^{k+2}}{\lambda(1-\rho^{k+1})}
\end{aligned} \tag{5.48}$$

A partir de esta última expersión podemos obtener W_q de la misma forma que en el modelo M/M/1

$$
\begin{aligned}
W_q &= W - \frac{1}{\mu} \\
&= \frac{1}{\mu-\lambda} - \frac{(k+1)\rho^{k+2}}{\lambda(1-\rho^{k+1})} - \frac{1}{\mu} \\
&= \frac{\lambda}{\mu(\mu-\lambda)} - \frac{(k+1)\rho^{k+2}}{\lambda(1-\rho^{k+1})}
\end{aligned}
\tag{5.49}
$$

De igual manera utilizando las fórmulas de Little se obtiene:

$$
\begin{aligned}
L_q &= \overline{\lambda} W_q \\
&= \overline{\lambda}\left(W - \frac{1}{\mu}\right) \\
&= L - \frac{\overline{\lambda}}{\mu}; \quad \text{De las expresiones } 5.46 \text{ y } 5.47 \\
&= \frac{\rho}{1-\rho} - \frac{(k+2)\rho^{k+2}}{1-\rho^{k+2}} - \frac{\frac{\lambda(\rho^{k+1}-1)}{\rho^{k+2}-1}}{\mu} \\
&= \frac{\rho}{1-\rho} - \frac{(k+2)\rho^{k+2}}{1-\rho^{k+2}} - \frac{\rho(1-\rho^{k+1})}{1-\rho^{k+2}} \\
&= \frac{\rho}{1-\rho} - \frac{(k+2)\rho^{k+2}}{1-\rho^{k+2}} - \frac{\rho-\rho^{k+2}+\rho^{k+3}-\rho^{k+3}}{1-\rho^{k+2}} \\
&= \frac{\rho}{1-\rho} - \frac{(k+2)\rho^{k+2}}{1-\rho^{k+2}} - \frac{\rho^{k+2}(\rho-1)+\rho(1-\rho^{k+2})}{1-\rho^{k+2}} \\
&= \frac{\rho}{1-\rho} - \rho - \frac{(k+1+\rho)\rho^{k+2}}{1-\rho^{k+2}} \\
&= \frac{\rho^2}{1-\rho} - \frac{(k+1+\rho)\rho^{k+2}}{1-\rho^{k+2}}
\end{aligned}
\tag{5.50}
$$

Esta última expresión, cuyo primer sumando es precisamente la fórmula para L_q en un modelo M/M/1. Obviamente, si el objetivo es calcular las cuatro cantidades (L, L_q, W y W_q) es más eficiente obtener el valor de $\overline{\lambda}$ y, una vez calculada una de las cuatro, obtener las tres restantes directamente de las fórmulas de Little y la relación entre tiempos medios.

Aunque el modelo M/M/1/k no contiene al M/M/1 como caso particular, en el caso $\rho < 1$ (para el cual el M/M/1 es estacionario) el modelo M/M/1/k debe tender al M/M/1 cuando $k \to \lambda$ (que es tanto como decir que el tamaño máximo permitido para la cola es más y más grande).

En efecto, los resultados que ofrecen las fórmulas anteriores para los p_n, L, L_q, W y W_q coinciden con los que aparecen en el M/M/1. Así por ejemplo:

$$
\begin{aligned}
\lim_{k\longrightarrow\infty} L_{M/M/1/k} &= \frac{\rho}{1-\rho} - \lim_{k\longrightarrow\infty}\frac{(k+2)\rho^{k+2}}{1-\rho^{k+2}} \\
&= \frac{\rho}{1-\rho} \\
&= L_{M/M/1}
\end{aligned}
\tag{5.51}
$$

Haciendo uso de la regla de L Hopital queda que el límite del numerador es cero y por lo tanto el límite del modelo M/M/1/k coincide con el de un modelo M/M/1.

CASO ρ =1: En este caso los resultados son fáciles de obtener y vienen dados de la siguiente manera:

$$
p_0 = \frac{1}{1+\sum_{n=1}^{\infty} c_n} = \frac{1}{1+k+1} = \frac{1}{k+2} \tag{5.52}
$$

tomando encuenta que $c_n = \rho^n = 1$, para $n = 1, 2, 3, \ldots, k+1$, entonces

$$
p_n = \frac{1}{k+2}, \forall n \ = 0, 1, \ldots, k+1 \tag{5.53}
$$

Se puede decir que N tiene una distribución uniforme discreta sobre su conjunto de valores posibles, y en este caso,

- Número medio de clientes en el sistema:

$$
\begin{aligned}
L &= \sum_{n=0}^{k+1} np_n = \sum_{n=0}^{k+1} n\frac{1}{k+2} \\
&= \frac{1}{(k+2)}\frac{(k+2)(k+1)}{2} \\
&= \frac{k+1}{2}
\end{aligned}
\tag{5.54}
$$

- Tiempo medio de llegada de los clientes:

$$
\begin{aligned}
\overline{\lambda} &= \sum_{n=0}^{\infty} \lambda p_n \\
&= \lambda\sum_{n=0}^{k} p_n = \lambda\left[p_0 + p_1 + p_2 + \cdots + p_k\right] \\
&= \lambda\left[\frac{1}{k+2} + \frac{1}{k+2} + \cdots + \frac{1}{k+2}\right] \\
&= \lambda\left[\frac{k+1}{k+2}\right]
\end{aligned}
\tag{5.55}
$$

- Tiempo medio de clientes en el sistema:

$$W = \frac{L}{\overline{\lambda}} = \frac{\frac{k+1}{2}}{\frac{\lambda(k+1)}{k+2}} = \frac{k+2}{2\lambda} \tag{5.56}$$

- Tiempo medio de clientes en cola:

$$W_q = W - \frac{1}{\mu} = \frac{k+2}{2\lambda} - \frac{1}{\lambda} = \frac{k}{2\lambda} \tag{5.57}$$

- Número medio de clientes en cola:

$$L_q = \overline{\lambda} W_q = \frac{\lambda(k+1)}{(k+2)} \frac{k}{2\lambda} = \frac{k(k+1)}{2(k+2)} \tag{5.58}$$

Para cualquier valor de λ (igual o distinto de 1), si se desea obtener las funciones de distribución del tiempo que un cliente está en el sistema y del tiempo que un cliente está en la cola, es necesario previamente definir q_n, como la probabilidad de que haya n clientes en el sistema justo cuando una nueva llegada se esta produciendo. Denotando, como siempre, por N el número de clientes en el sistema y por T el tiempo que falta para que se produzca la llegada del siguiente cliente, las q_n representan P(N = n|T = 0). Usando la notación f(t|N = n) para la función de densidad de la variable T condicionada a que N = n, se sigue que f(t|N = n) corresponde a la densidad de una exp(λ_n).

Así, aplicando la regla de Bayes, se tiene:

$$\begin{aligned} q_n &= P(N = n|_{T=0}) \\ &= \frac{f(0|_{N=n})p_n}{\sum_{m=0}^{k} f(0|_{N=m})p_m} \\ &= \frac{\lambda_n e^{-\lambda_n 0} p_n}{\sum_{m=0}^{k} \lambda_m e^{-\lambda_m 0} p_m} \\ &= \frac{\lambda \cdot p_n}{\lambda \sum_{m=0}^{k} p_m} \\ &= \frac{p_n}{1 - p_{k+1}} \end{aligned}$$

(5.59)

Para n = 0,1,...,k; mientras que q_n = 0paran = k+1,k+2,... para el caso del modelo M/M/c (incluyendo el caso de c = 1) debe de verificarse que las variables T y N son independientes y, por lo tanto:

$$q_n = P(N = n|_{T=0}) = P(N = n) = p_n$$

Las funciones de distribución de las variables W y W_q vienen dadas por:

$$W = 1 - e^{-\mu t} \sum_{n=0}^{k} q_n \sum_{r=0}^{n} \frac{(\mu t)^r}{r!}, \; si \; t \geq 0 \; (y \; W(t) = 0 \; en \; otro \; caso) \tag{5.60}$$

$$W_q = 1 - e^{-\mu t} \sum_{n=1}^{k} q_n \sum_{r=0}^{n-1} \frac{(\mu t)^r}{r!}, \; si \; t \geq 0 \; (y \; W_q(t) = 0 \; si \; t < 0) \tag{5.61}$$

Este modelo (y en otros posteriores) el significado de ρ como intensidad de tráfico se desvirtúa. Aquí ρ no puede interpretarse como el cociente entre el número medio de llegadas de clientes al sistema por unidad de tiempo entre el número medio de clientes a los que el servidor tendría capacidad de dar servicio por unidad de tiempo, sino más bien como un cociente semejante, pero donde el numerador representa el número medio de intentos de llegada, más que de llegadas efectivas al sistema. De hecho, por este motivo ρ puede ser mayor o igual que 1, aun siendo el sistema estacionario. El valor de $\overline{\lambda}$ sí representa el número medio de entradas efectivas de clientes en el sistema por unidad de tiempo y, así, la verdadera intensidad de tráfico podría medirse a través de que efectivamente sí es siempre menor que 1.

$$\overline{\rho} = \frac{\overline{\lambda}}{\mu} = \begin{cases} \dfrac{\frac{\lambda(k+1)}{k+2}}{\mu} = \dfrac{k+1}{k+2} & ; si \; \rho = 1 \\ \dfrac{\frac{\lambda(\rho^{k+1}-1)}{\rho^{k+2}-1}}{\mu} = \dfrac{\rho^{k+2} - \rho}{\rho^{k+2} - 1} & ; si \; \rho \neq 1 \end{cases}$$

3.3.6. El modelo M/M/c.

Este modelo estudia los tipos de colas en donde la distribución de los tiempos entre llegadas son exponenciales con parámetro λ, y además los tiempos de servicio se distribuyen exponencialmente con parámetro µ y en este caso existen c servidores, bajo la suposición de que cada servidor realiza la misma función o actividad con el mismo nivel de eficacia que los demás servidores. Con respecto a los demás parámetros del modelo tenemos que la población potencial y la capacidad de la cola es infinita, la forma de elegir a los clientes en cola para que sean servidos es FIFO. Es por ello que este modelo es considerado como la generalización del modelo M/M/1. Por tanto, las tasas de llegada vienen dadas por:

$$\lambda_n = \lambda; \quad \forall n = 0, 1, \ldots$$

Y las tasas de servicio son:

$$\mu_n = \begin{cases} n\mu; & \forall 1 \leq n \leq c \\ c\mu; & \forall n > c \end{cases} \tag{5.62}$$

De forma similar al caso M/M/1 para obtener c_n:

$$c_n = \frac{\lambda_{n-1}\lambda_{n-2}\ldots\lambda_0}{\mu_n\mu_{n-1}\ldots\mu_1} = \prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i} = \begin{cases} \dfrac{\lambda^n}{n!\mu^n} & ;\text{si } 1 \leq n \leq c \\ \dfrac{\lambda^n}{c!c^{n-c}\mu^n} & ;\text{si } \forall n > c \end{cases} \tag{5.63}$$

Es necesario también establecer la condición estacionaria para el sistema ($\rho < 1$), o lo que es lo mismo, siempre que se cumpla $\lambda < c\mu$. Para ello basta con determinar la convergencia de la serie siguiente:

$$\begin{aligned} \sum_{n=1}^{\infty} c_n &= \sum_{n=1}^{c-1}\frac{\lambda^n}{n!\mu^n} + \sum_{n=c}^{\infty}\frac{\lambda^n}{c!c^{n-c}\mu^n} \\ &= \sum_{n=1}^{c-1}\frac{\lambda^n}{n!\mu^n} + \sum_{n=c}^{\infty}\frac{\lambda^c\lambda^{n-c}}{c!c^{n-c}\mu^c\mu^{n-c}} \end{aligned}$$

$$\begin{aligned} &= \sum_{n=1}^{c-1}\frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{c!\mu^c}\sum_{n=c}^{\infty}\frac{\lambda^{n-c}}{c^{n-c}\mu^{n-c}}; \quad \rho = \frac{\lambda}{c\mu} \\ &= \sum_{n=1}^{c-1}\frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{c!\mu^c}\sum_{n=c}^{\infty}\rho^{n-c} \end{aligned} \tag{5.64}$$

Entonces a partir del término c-ésimo, la serie es geométrica de razón ρ y por tanto convergente siempre que $\lambda < c\mu$ que puede ser interpretado como el número medio de clientes que entran al sistema por unidad de tiempo, debe ser menor que el número medio de clientes a los que se le completa su servicio, por unidad de tiempo; multiplicada por el número de servidores que hay en el sistema. Probar la convergencia en la serie $\sum_{n=c}^{\infty} \rho^{n-c}$ resulta sencilla.

$$\begin{aligned}
\sum_{n=c}^{\infty} \rho^{n-c} &= 1 + \rho + \rho^2 + \rho^3 + \ldots \\
&= 1 + \rho(1 + \rho + \rho^2 + \rho^3 + \ldots) \\
&= 1 + \rho\left(\sum_{n=c}^{\infty} \rho^{n-c}\right) \\
(1-\rho)\sum_{n=c}^{\infty} \rho^{n-c} &= 1 \\
\sum_{n=c}^{\infty} \rho^{n-c} &= \frac{1}{1-\rho}
\end{aligned} \tag{5.65}$$

Entonces de 5.64, la suma de las c_n es:

$$\begin{aligned}
\sum_{n=1}^{\infty} c_n &= \sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{c!\mu^c} \cdot \frac{1}{1-\rho} \\
&= \sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{c(c-1)!\mu\mu^{c-1}(1-\rho)} \\
&= \sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{(c-1)!\mu^{c-1}(c\mu-\lambda)}
\end{aligned} \tag{5.66}$$

Análogamente al modelo M/M/1 se puede calcular p_0 de la siguiente manera:

$$\begin{aligned}
p_0 &= \frac{1}{1+\sum_{n=1}^{\infty} c_n} \\
&= \frac{1}{1+\sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{(c-1)!\mu^{c-1}(c\mu-\lambda)}}
\end{aligned} \tag{5.67}$$

A la hora de interpretar el cálculo para p_0 debemos tener en cuenta, que si el número de servidores del sistema es alto, los términos con n! en el denominador de p_0, puede ocurrir errores del tipo "overflow" o pérdidas de tiempo/datos por exceso de cálculos. Es importante hacer notar que en estos casos es ineficiente aplicar las fórmulas cada vez que se desee calcular los valores, es más útil reutilizar dichos datos. Por ejemplo: si c = 100, en el denominador de c_{98} aparecería 98!, que, en lugar de calcularlo directamente es más eficiente multiplicar por 98 al término 97! que aparece en c_{97}. En resumen, es más eficiente definir $c_0 = 1$ y al utilizar cálculos recursivos, tenemos:

$$c_n = c_{n-1}\frac{\lambda_{n-1}}{\mu_n} = c_{n-1}\frac{\lambda}{n\mu}, \quad \forall n = 1, 2, \ldots, c-1$$

Esto se deduce de la definición de los c_n. Además para el caso del término $\dfrac{\lambda^c}{(c-1)!\mu^{c-1}(c\mu-\lambda)}$ que presenta $\sum_{n=c}^{\infty} c_n$ en (5.64), vemos que:

$$\frac{\lambda^c}{(c-1)!\mu^{c-1}(c\mu-\lambda)} = \frac{\lambda^{c-1}}{(c-1)!\mu^{c-1}}\left(\frac{\lambda}{c\mu-\lambda}\right) = d_{c-1}\frac{\lambda}{c\mu-\lambda}$$

y que denotaremos por $D_{\geq c}$, puede calcularse fácilmente a partir de d_{c-1} mediante

$$D_{\geq c} = d_{c-1}\frac{\lambda}{c\mu-\lambda}$$

donde $c_0 = 1$, $c_n = c_{n-1}\frac{\lambda}{n\mu}$, $\forall n = 1, 2, \ldots, c-1$ y $D_{\geq c} = d_{c-1}\frac{\lambda}{c\mu-\lambda}$.

Ahora para determinar una expresión explícita para las p_n:

$$p_n = c_n p_0 = \begin{cases} \dfrac{\lambda^n}{n!\mu^n} p_0; & \forall 1 \leq n < c \\ \\ \dfrac{\lambda^c}{c!\mu^c}\rho^{n-c} p_0; & \forall n \geq c \end{cases} \tag{5.68}$$

Nuevamente, la aplicación directa de la fórmula anterior, resulta ser poco eficiente en términos computacionales. De manera similar al cálculo de p_0, también aquí podemos proceder más eficientemente. Una vez calculado p_0, y si se mantiene en memoria los valores de c_n el cálculo de los p_n sería muy sencillo. Así, si $n = 1,2,...,c-1$, se tiene $p_n = c_n p_0$. Cuando $n \geq c$ los p_n pueden ser calculados recursivamente ya que p_{c-1} se ha obtenido antes.

$$
\begin{aligned}
p_n &= c_n p_0 \\
&= c_{n-1}\frac{\lambda_{n-1}}{\mu_n}p_0 \\
&= \frac{\lambda_{n-1}}{\mu_n}p_{n-1} \\
&= \frac{\lambda}{c\mu}p_{n-1} \\
&= \rho . p_{n-1}, \forall n \geq c
\end{aligned} \tag{5.69}
$$

Entonces, la manera de proceder es la siguiente:

Calcular directamente $p_n = c_n p_0$; $1 \leq n < c$

Recursivamente $p_n = \rho . p_{n-1}$; $n \geq c$ Utilizando las expresiones obtenidas para p_n, puede encontrarse fácilmente el valor de L_q.

$$
L_q = E(N_q) \;=\; 0(p_0 + p_1 + \ldots + p_{c-1}) + \sum_{n=c}^{\infty}(n-c)p_n; \;\; \text{donde } p_n = c_n p_0
$$

$$
\begin{aligned}
&= \sum_{n=c}^{\infty}(n-c)c_n p_0 \\
&= \sum_{n=c}^{\infty}(n-c)\frac{\lambda^n}{c!c^{n-c}\mu^n}p_0 \\
&= \sum_{n=c}^{\infty}(n-c)\frac{\lambda^{n-c}\lambda^c}{c!c^{n-c}\mu^{n-c}\mu^c}p_0 \\
&= \frac{\lambda^c}{c!\mu^c}p_0\sum_{n=c}^{\infty}(n-c)\rho^{(n-c)}; \;\; \text{sea } k = n-c
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\lambda^c}{c!\mu^c}p_0\sum_{k=0}^{\infty}k\rho^k \\
&= \frac{\lambda^c}{c!\mu^c}p_0\frac{\rho}{(1-\rho)^2}; \;\; \rho = \frac{\lambda}{c\mu} \\
&= \frac{\lambda^{c+1}p_0}{(c\mu)^2(c-1)!\mu^{c-1}\left(\frac{c\mu-\lambda}{c\mu}\right)^2} \\
&= \frac{\lambda^{(c+1)}p_0}{(c-1)!\mu^{(c-1)}(c\mu-\lambda)^2}
\end{aligned} \tag{5.70}
$$

En la práctica, dado que $D_{\geq c} = \frac{\lambda^c}{(c-1)!\mu^{(c-1)}(c\mu-\lambda)}$, puede calcularse de forma eficiente mediante la expresión:

$$L_q = D_{\geq c} \cdot \frac{\lambda p_0}{c\mu - \lambda}$$

Mediante la segunda fórmula de Little, se puede obtener W_q (Tiempo medio de espera en cola) a partir de L_q calculado anteriormente, entonces:

Según Little:

$$\begin{aligned} W_q &= \frac{1}{\lambda} L_q \\ &= \frac{1}{\lambda} \cdot \frac{\lambda^{c+1} p_0}{(c-1)!\mu^{c-1}(c\mu-\lambda)^2} \\ &= \frac{\lambda^c p_0}{(c-1)!\mu^{c-1}(c\mu-\lambda)^2} \end{aligned} \tag{5.71}$$

Ahora se puede obtener W (Tiempo medio de espera en el sistema) de la siguiente forma

$W = W_q + \frac{1}{\mu}$ y luego el número medio de clientes en el sistema $L = \lambda W$. Entonces ya conocidas λ, μ, c y p_0. Es fácil obtener:

$$\begin{aligned} W &= W_q + \frac{1}{\mu} \\ &= \frac{\lambda^c \cdot p_0}{(c-1)!\mu^{c-1}(c\mu-\lambda)^2} + \frac{1}{\mu} \end{aligned} \tag{5.72}$$

Luego $L = \lambda W$

$$\begin{aligned} L &= \lambda W \\ &= \lambda \left[\frac{\lambda^c \cdot p_0}{(c-1)!\mu^{c-1}(c\mu-\lambda)^2} + \frac{1}{\mu} \right] \\ &= \frac{\lambda^{c+1} \cdot p_0}{(c-1)!\mu^{c-1}(c\mu-\lambda)^2} + \frac{\lambda}{\mu} \end{aligned} \tag{5.73}$$

También, si en algún momento deseamos conocer la función de distribución del tiempo que un cliente tarda en la cola de un sistema, ésta puede calcularse de la siguiente forma:

Lema. Sea $F_{Tq}(t)$ la función de distribución del tiempo que un cliente pasa en cola, entonces:

$$F_{T_q}(t) = \begin{cases} 1 - \dfrac{c(\lambda/\mu)^c}{c!(c-\lambda/\mu)} p_0 & ; si\ t = 0 \\ \dfrac{\mu(\lambda/\mu)^c(1-e^{-(\mu c-\lambda)t})}{(c-1)!(\mu c-\lambda)} p_0 + F_{T_q}(0) & ; si\ t > 0 \end{cases} \quad (5.74)$$

Demostración:

$$F_{T_q}(0) = P(T_q = 0) = P(N \leq c-1) = \sum_{n=0}^{c-1} p_n = p_0 \sum_{n=0}^{c-1} \frac{\lambda^n}{n!\mu^n}$$

Pero sabemos que

$$\begin{aligned} p_0 &= \frac{1}{1 + \sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{(c-1)!\mu^{c-1}(c\mu-\lambda)}} \\ &= \frac{1}{\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{(c-1)!\mu^{c-1}(c\mu-\lambda)}} \end{aligned}$$

si y sólo si

$$\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\mu^n} = \frac{1}{p_0} - \frac{c(\lambda/\mu)^c}{c!(c-\lambda/\mu)}$$

Entonces

$$F_{T_q}(0) = 1 - \frac{c(\lambda/\mu)^c}{(c-\lambda/\mu)c!} \cdot p_0$$

Ahora, si $n \geq c$, el proceso Z(t) (el número de salidas hasta el instante t) se comporta como un proceso de Poisson (cμ). Esto significa que los tiempos entre llegadas consecutivas τ_1, τ_2,... son v.a.i.i.d. exp. (cμ). Consecuentemente la variable T_k, representa el tiempo transcurrido hasta que terminen k servicios y se distribuye según una $\Gamma(p = k, a = c\mu)$. Entonces:

$$
\begin{aligned}
F_{T_q}(t) &= P(T_q \le t) = P(T_q = 0) + P(0 < T_q < t) \\
&= F_{T_q}(0) + \sum_{n=c}^{\infty} p_n \underbrace{\int_0^t f(x)dx}_{\Gamma(p=n-c+1,a=c\mu)} \\
&= F_{T_q}(0) + \sum_{n=c}^{\infty} c_n p_0 \int_0^t \frac{(c\mu)^{n-c+1}}{(n-c)!} x^{n-c} e^{-c\mu x} dx; \quad Sea\ r = \frac{\lambda}{\mu} \\
&= F_{T_q}(0) + p_0 \sum_{n=c}^{\infty} \frac{r^n}{c^{n-c} c!} \int_0^t \frac{(c\mu)^{n-c+1}}{(n-c)!} x^{n-c} e^{-c\mu x} dx \\
&= F_{T_q}(0) + p_0 \frac{r^c}{(c-1)!} \int_0^t \mu e^{-\mu c x} \left[\sum_{n=c}^{\infty} \frac{(\mu r x)^{n-c}}{(n-c)!} \right] dx \\
&= F_{T_q}(0) + p_0 \frac{r^c}{(c-1)!} \int_0^t \mu e^{-\mu(c-r)x} dx \\
&= F_{T_q}(0) + p_0 \frac{r^c}{(c-1)!} \cdot \frac{\mu}{\mu(c-r)} \left[-e^{-\mu(c-r)x} \right]_0^t \\
&= F_{T_q}(0) + p_0 \frac{r^c}{(c-1)!} \cdot \frac{\left(1 - e^{-\mu(c-r)t}\right)}{(c-r)} \\
F_{T_q}(t) &= 1 - \frac{c r^c}{(c-r)c!} p_0 + \frac{r^c}{(c-1)!} \cdot \frac{\left(1 - e^{-\mu(c-r)t}\right)}{(c-r)} p_0 \qquad (5.75)
\end{aligned}
$$

3.3.7. El modelo M/M/c/k.

Este modelo es una extensión del modelo M/M/1/k, ya que en este caso existen c servidores y la capacidad de la cola es limitada a k clientes. Donde las tasas entre llegadas se comportan como las del modelo M/M/1/k, mientras que las tasas de servicio son iguales a las de un M/M/c:

$$
\lambda_n = \begin{cases} \lambda; & \text{si } n = 0, 1, 2, \ldots, k + c - 1 \\ 0; & \text{si } n = k + c, k + c + 1, k + c + 2, \ldots \end{cases}
$$

$$
\mu_n = \begin{cases} n\mu; & \text{si } n = 1, 2, \ldots, c \\ c\mu; & \text{si } n = c + 1, c + 2, \ldots \end{cases}
$$

Como consecuencia de esto se obtiene:

$$c_n = \begin{cases} \dfrac{\lambda^n}{n!\mu^n} & ;\text{si } n = 1, 2, \ldots, c \\ \dfrac{\lambda^n}{c!c^{n-c}\mu^n} & ;\text{si } n = c+1, c+2, \ldots, k+c \\ 0 & ;\text{si } n = k+c+1, k+c+2, \ldots \end{cases} \tag{5.76}$$

Entonces la serie $\sum_{n-1}^{\infty} c_n$ tiene sólo un número finito de términos distintos de cero y, por tanto, siempre es convergente. Así pues el sistema es estacionario siempre (independientemente del valor de ρ). Para calcular p_0 necesitamos encontrar una expresión, lo más sencilla posible para la suma de la serie:

$$\begin{aligned} \sum_{n=1}^{\infty} c_n = \sum_{n=1}^{k+c} c_n &= \sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \sum_{n=c}^{k+c} \frac{\lambda^n}{c!c^{n-c}\mu^n} \\ &= \sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{c!\mu^c} \sum_{n=c}^{k+c} \rho^{n-c} \end{aligned}$$

Obteniendo como resultado:

$$\sum_{n=1}^{\infty} c_n = \begin{cases} \displaystyle\sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{c!\mu^c} \cdot \frac{1-\rho^{k+1}}{1-\rho} & ;\text{si } \rho \neq 1 \\ \displaystyle\sum_{n=1}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c(k+1)}{c!\mu^c} & ;\text{ si } \rho = 1 \end{cases} \tag{5.77}$$

Como consecuencia:

$$p_0 = \begin{cases} \left(\displaystyle\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c}{c!\mu^c} \cdot \frac{1-\rho^{k+1}}{1-\rho}\right)^{-1} & ;\text{si } \rho \neq 1 \\ \left(\displaystyle\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\mu^n} + \frac{\lambda^c(k+1)}{c!\mu^c}\right)^{-1} & ;\text{si } \rho = 1 \end{cases} \tag{5.78}$$

Al igual que en el caso de un sólo servidor usaremos la notación ρ = λ/cμ. Tal como ya se comentó en el modelo M/M/c, la implementación directa de las fórmulas anteriores no es precisamente la manera más eficiente de calcular p_0. Entonces, si empleamos las fórmulas recursivas puede calcularse p_0 mediante

$$p_0 = \frac{1}{\sum_{n=0}^{c-1} c_n + D_{\geq c}}$$

donde: $c_0 = 1, c_n = c_{n-1}\frac{\lambda}{n\mu}, \forall n = 1, 2, \ldots$

Además:

$$D_{\geq c} = \begin{cases} d_{c-1}\dfrac{\rho - \rho^{k+2}}{1-\rho} & ; \text{si } \rho \neq 1 \\ d_{c-1}(k+1) & ; \text{si } \rho = 1 \end{cases} \tag{5.79}$$

A partir de p_0, y los parámetros de entrada del modelo, pueden obtenerse expresiones explícitas para las p_n:

$$p_n = c_n \cdot p_0 = \begin{cases} \dfrac{\lambda^n}{n!\mu^n} p_0 & ; \text{si } n = 1, 2, \ldots, c \\ \dfrac{\lambda^c}{c!\mu^c} \rho^{n-c} p_0 & ; \text{si } n = c+1, c+2, \ldots, k+c \end{cases} \tag{5.80}$$

Nuevamente, las fórmulas implementadas directamente a partir de estas expresiones son muy poco eficientes y resulta preferible proceder de forma análoga al modelo M/M/c. Es decir; reutilizando los valores de c_n, necesarios para la implementación del cálculo eficiente de p_0, los p_n se obtienen de manera sencilla: $p_n = c_n \cdot p_0$, $\forall n = 1,2,...,c-1$. Por otra parte, si $n \geq c$, el término p_n debe ser calculado recursivamente, comenzando en p_{c-1} que ya ha sido calculado:

$$p_n = \frac{\lambda_{n-1}}{\mu_n} p_{n-1} = \rho \cdot p_{n-1}, \forall n = c, c+1, \ldots, k+c$$

Fórmula que, aun siendo válida para todo valor de ρ, se trivializa si ρ = 1, dando como resultado.

$$p_{k+c} = p_{k+c-1} = \cdots = p_c = p_{c-1} \quad \text{para } \rho = 1$$

Las cuatro medidas de interés (L, L_q, W, y W_q) pueden calcularse a partir de los valores de las p_n recién encontradas. Posiblemente el cálculo que resulte más sencillo sea el de L_q, por el cual comenzaremos. Si ρ ≠ 1, a partir de la igualdad.

$$\sum_{n=0}^{k+1} nx^{n-1} = \frac{(k+1)x^{k+2} - (k+2)x^{k+1} + 1}{(x-1)^2} \tag{5.81}$$

Demostrada y utilizada ya para el modelo M/M/1/k, se tiene.

$$\begin{aligned} L_q &= \sum_{n=c+1}^{c+k} (n-c)p_n = \sum_{n=c}^{c+k} (n-c)\frac{\lambda^n}{c!c^{n-c}\mu^n}p_0 \\ &= \frac{\lambda^c}{c!\mu^c} \cdot p_0 \cdot \rho \sum_{n=c}^{c+k} (n-c)\rho^{n-c-1} \\ &= \frac{\lambda^c}{c!\mu^c} \cdot p_0 \cdot \rho \sum_{j=0}^{k} j \cdot \rho^{j-1} \\ &= \frac{\lambda^c}{c!\mu^c} \cdot p_0 \cdot \rho \cdot \frac{k\rho^{k+1} - (k+1)\rho^k + 1}{(\rho-1)^2} \\ &= \frac{\lambda^c \cdot p_0 \cdot \rho \cdot [1 + k\rho^{k+1} - (k+1)\rho^k]}{c!\mu^c(\rho-1)^2} \end{aligned} \tag{5.82}$$

Sin embargo, si ρ = 1 entonces

$$\begin{aligned} L_q &= \sum_{n=c+1}^{c+k} (n-c) \cdot p_n \\ &= p_{c-1} \sum_{N=c+1}^{c+k} (n-c) \\ &= p_{c-1}\frac{k(k+1)}{2} \\ &= \frac{\lambda^{c-1}k(k+1)}{2(c-1)!\mu^{c-1}} \end{aligned}$$

En resumen, la expresión explícita para el número medio de clientes en la cola es

$$L_q = \begin{cases} \dfrac{\lambda^{c-1} k(k+1) \cdot p_0}{2(c-1)!\mu^{c-1}} & ;\text{si } \rho = 1 \\ \\ \dfrac{\lambda^c \cdot p_0 \cdot \rho[1 + k\rho^{k+1} - (k+1)\rho^k]}{c!\mu^c(1-\rho)^2} & ;\text{si } \rho \neq 1 \end{cases} \tag{5.84}$$

Sin embargo, para la obtención de un cálculo mucho más eficiente usaremos,

$$L_q = \begin{cases} \dfrac{k(k+1)}{2} p_{c-1} & ;\text{si } \rho = 1 \\ \\ \dfrac{[1 + k\rho^{k+1} - (k+1)\rho^k]\rho^2}{(1-\rho)^2} p_{c-1} & ;\text{si } \rho \neq 1 \end{cases} \tag{5.85}$$

Para la obtención de las demás medidas debemos de usar las fórmulas de Little y calcular, primeramente, $\bar{\lambda}$:

$$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n \cdot p_n = \lambda \sum_{n=0}^{k+c+1} p_n = \lambda(1 - p_{k+c}),$$

Que es la implementación a usar para un cálculo eficiente. Una fórmula más explícita que proporcionará este valor es:

$$\bar{\lambda} = \lambda \left(1 - \frac{\lambda^c}{c!\mu^c} \rho^k p_0\right)$$

Que se simplifica a $\bar{\lambda} = \lambda \left(1 - \frac{\lambda^{c-1}}{(c-1)!\mu^{c-1}} p_0\right)$ Si ρ = 1. De nuevo, en este modelo vuelve a ocurrir que ρ no representa la intensidad de tráfico efectiva.

Este valor puede calcularse mediante

$$\begin{aligned} \bar{\rho} = \frac{\bar{\lambda}}{c\mu} &= \rho \left(1 - \frac{\lambda^c}{c!\mu^c} \rho^k p_0\right) \\ \bar{\rho} &= \rho - \frac{\lambda^c}{c!\mu^c} \rho^{k+1} p_0 \end{aligned} \tag{5.86}$$

A partir de los valores de L_q y $\bar{\lambda}$, calculados de forma eficiente, pueden usarse las fórmulas de Little para obtener:

$$
\begin{aligned}
W_q &= \frac{L_q}{\lambda}, \\
W &= W_q + \frac{1}{\mu}, \\
L &= \bar{\lambda} W.
\end{aligned}
$$

Esimportantemencionarqueenestemodeloresultabastantecomplicadoobtenerladistribución del tiempo que un cliente está en el sistema. Sin embargo es posible determinar el tiempo que un cliente pasa en cola a través de la siguiente expresión:

$$\mathbf{W}_q(t) = 1 - e^{-c\mu t} \sum_{n=c}^{k+c-1} q_n \sum_{r=0}^{n-c} \frac{(c\mu t)^r}{r!}, \quad \text{si } t \geq 0 \quad (\text{y} \quad \mathbf{W}_q(t) = 0 \text{ si } t < 0)$$

donde las q_n tienen el mismo significado que en el modelo M/M/1/k, es decir, la probabilidad de que haya n clientes en el sistema justo cuando una llegada se está produciendo, viene dada por

$$q_n = \frac{p_n}{1 - p_{k+c}}, n = 0, 1, 2, \ldots, k + c - 1$$

3.3.8. El modelo M/M/c/c (un caso especial del modelo M/M/c/k).

Un caso especial de este modelo es cuando la capacidad de la cola es igual al número de servidores que hay en el sistema, es decir; k = c. Entonces bajo estas condiciones nunca se producirán colas, es por ello que muchas veces este modelo no es considerado como un sistema de colas propiamente dicho. Sin embargo; resulta de especial importancia este modelo ya que es considerado como un modelo clásico de las centrales telefónicas, en la cual el número de servidores es el número de líneas en el sistema. Además, en este tipo de modelos existe un valor que es especialmente relevante, nos referimos a la probabilidad de que el sistema esté saturado, es decir, haya c clientes en el sistema, frecuentemente identificado en las centrales de telefonía como “la probabilidad de no tener línea”, y cuyo valor como se ha visto con anterioridad es:

$$p_c = \frac{(c\rho)^c / c!}{\sum_{i=0}^{c} (c\rho)^i / i!} \tag{5.87}$$

También se puede determinar la probabilidad de que haya n elementos en el sistema por:

$$p_n = \frac{\frac{(\lambda/\mu)^n}{n!}}{\sum_{i=0}^{c} \frac{(\lambda/\mu)^i}{i!}} \tag{5.88}$$

Demostrar que (5.88) es cierta resulta sencillo, para ello partimos de la expresión pn = cnp0 utilizada anteriormente para denotar la probabilidad de que hayan n clientes en el sistema, entonces:

$$\begin{aligned} p_n &= c_n p_0 \\ &= p_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} \\ &= p_0 \prod_{i=1}^{n} \frac{\lambda}{i\mu}; \quad \forall 1 \leq n \leq c \\ &= p_0 \left(\frac{\lambda^n}{n!\mu^n} \right) \\ p_n &= \frac{(\frac{\lambda}{\mu})^n}{n!} p_0 \end{aligned}$$

Como p_n representa probabilidades, entonces se cumple que:

$$1 = \sum_{i=0}^{\infty} p_i = p_0 \sum_{i=0}^{c} \frac{(\frac{\lambda}{\mu})^i}{i!}$$

Entonces

$$p_0{}^{-1} = \sum_{i=0}^{c} \frac{(\frac{\lambda}{\mu})^i}{i!}$$

Por lo tanto

$$p_n = \frac{(\lambda/\mu)^n}{n!} p_0 = \frac{(\lambda/\mu)^n}{n! \sum_{i=0}^{c} \frac{(\lambda/\mu)^i}{i!}} \quad , \forall 1 \leq n \leq c$$

3.3.9. El modelo M/G/1.

Como su nombre indica, se trata del sistema de una cola con un único servidor, con tiempo entre llegadas de clientes consecutivos con distribución exponencial y con tiempo de servicio de distribución arbitraria, G. En este contexto, λ seguirá siendo el parámetro de la exponencial que rige los tiempos entre dos llegadas de clientes consecutivos. Esto significa que el número de clientes que llegan al sistema por unidad de tiempo será también λ. Como la distribución G no tiene por qué ser exponencial, el parámetro μ pasará ahora a significar el inverso del tiempo medio de servicio, es decir:

$$\mu = \frac{1}{E[G]}$$

Esta cola viene caracterizada por las siguientes hipótesis:

1. Tiempos entre llegadas exponenciales independientes; es decir, $\tau_1,\tau_2,...$ son variables aleatorias independientes e igualmente distribuidas (v.a.i.i.d) exp(λ). Por tanto, N(t): número de llegadas hasta el instante t sigue un proceso estocástico (P.E) de Poisson de parámetro λ.

2. Tiempos de servicio independientes; es decir, $s_1,s_2,...$ son v.a.i.i.d según una distribución de probabilidad cualquiera F(·). 3. $\tau_1,\tau_2,...$ y $s_1,s_2,...$ son independientes.

4. hay un único canal de servicio.

Como el tipo de distribución para el tiempo de llegadas es desconocido, los cálculos de las medidas de efectividad para este tipo de modelos se pueden obtener mediante la fórmula que se definirá a continuación.

3.3.10. La fórmula de Pollaczek-Khintchine.

Sea:

$$L = \rho + \frac{\rho^2 + \lambda^2\sigma_S^2}{2(1-\rho)} \tag{5.89}$$

Donde:

$\rho = \frac{\lambda}{\mu}$, $E[S] = \frac{1}{\mu}$, $V[S] = \sigma_S^2$ y L

que es el número medio de clientes en el sistema en estado estacionario.

Para demostrar (5.89) veamos la siguiente relación:

$$X_{n+1} = X_n - U(X_n) + A_{n+1}, \tag{5.90}$$

Donde:

$$U(X_n) = \begin{cases} 1, & X_n > 0 \\ 0, & X_n = 0 \end{cases}$$

Si tomamos esperanzas en (5.90) tenemos:

$$E[X_{n+1}] = E[X_n] - E[U(X_n)] + E[A_{n+1}]$$

Y tomando en cuenta que en estado estacionario se verifican las igualdades.

$$L = E[X_n] = E[X_{n+1}]$$

Entonces la ecuación (5.90) nos queda de la siguiente forma:

$$E[U(X_n)] = E[A_{n+1}]$$

Así pues:

$$\begin{aligned} E[A_{n+1}] &= E[E[A_{n+1}/S_{n+1}]] \\ &= \int_0^\infty E[A_{n+1}/S_{n+1}]\, g(t)dt \\ &= \int_0^\infty \lambda t g(t) dt \\ &= \lambda \int_0^\infty t g(t) dt \\ &= \lambda E[S_{n+1}] \\ &= \frac{\lambda}{\mu} \\ &= \rho \end{aligned} \tag{5.91}$$

Elevando al cuadrado (5.90) y tomando esperanzas tenemos:

$$
\begin{aligned}
X_{n+1}^2 &= [X_n - U(X_n) + A_{n+1}]^2 \\
&= [X_n - U(X_n) + A_{n+1}]\,[X_n - U(X_n) + A_{n+1}] \\
&= X_n^2 + X_n A_{n+1} - X_n U(X_n) + A_{n+1} X_n + A_{n+1}^2 - A_{n+1} U(X_n) \\
&\quad - U(X_n) X_n - U(X_n) A_{n+1} + U(X_n)^2 \\
&= X_n^2 + U(X_n)^2 + A_{n+1}^2 + 2X_n A_{n+1} - 2U(X_n) X_n - 2A_{n+1} U(X_n) \\
E[X_{n+1}^2] &= E[X_n^2] + E[U(X_n)^2] + E[A_{n+1}^2] + 2E[X_n]E[A_{n+1}] \\
&\quad - 2E[U(X_n)]E[X_n] - 2E[U(X_n)]E[A_{n+1}] \\
0 &= E[U(X_n)^2] + E[A_{n+1}^2] - 2E[U(X_n)]E[X_n] \\
&\quad - 2E[U(X_n)]E[A_{n+1}] + 2E[X_n]E[A_{n+1}] \\
&= \rho + E[A_{n+1}^2] - 2L - 2\rho^2 + 2L\rho \\
2L - 2L\rho &= \rho + E[A_{n+1}^2] - 2\rho^2 \\
L &= \frac{\rho - 2\rho^2 + E[A_{n+1}^2]}{2(1-\rho)}
\end{aligned} \tag{5.92}
$$

Desarrollando

$$
\begin{aligned}
E[A_{n+1}^2] &= V[A_{n+1}] + E[A_{n+1}]^2 \\
&= V[A_{n+1}] + \rho^2
\end{aligned} \tag{5.93}
$$

$$
\begin{aligned}
V[A_{n+1}] &= E[V[A_{n+1}/S_{n+1}]] + V[E[A_{n+1}/S_{n+1}]] \\
&= E[\lambda S_{n+1}] + V[\lambda S_{n+1}] \\
&= \lambda E[S_{n+1}] + \lambda^2 V[S_{n+1}] \\
&= \lambda \frac{1}{\mu} + \lambda^2 \sigma_S^2 \\
&= \rho + \lambda^2 \sigma_S^2
\end{aligned} \tag{5.94}
$$

Asi que:

$$
\begin{aligned}
L &= \frac{\rho - 2\rho^2 + \rho + \lambda^2\sigma_S^2 + \rho^2}{2(1-\rho)} \\
&= \frac{2\rho(1-\rho) + \lambda^2\sigma_S^2 + \rho^2}{2(1-\rho)} \\
&= \rho + \frac{\lambda^2\sigma_S^2 + \rho^2}{2(1-\rho)}
\end{aligned} \tag{5.95}
$$

A partir de la ecuación anterior se pueden obtener W, Wq y Lq ya que siguen verificándose las fórmulas de Little y la relación entre tiempos medios en el sistema de la cola, vemos entonces los siguientes resultados:

$$W = \frac{L}{\lambda},\ W_q = W - \frac{1}{\mu},\ L_q = \lambda W_q.$$

4. CONCLUSIONES.

La teoría de las colas es el estudio matemático de las colas o líneas de espera. La formación de colas es, por supuesto, un fenómeno común que ocurre siempre que la demanda efectiva de un servicio excede a la oferta efectiva.

Con frecuencia, las empresas deben tomar decisiones respecto al caudal de servicios que debe estar preparada para ofrecer. Sin embargo, muchas veces es imposible predecir con exactitud cuándo llegarán los clientes que demandan el servicio y/o cuanto tiempo será necesario para dar ese servicio; es por eso que esas decisiones implican dilemas que hay que resolver con información escasa. Estar preparados para ofrecer todo servicio que se nos solicite en cualquier momento puede implicar mantener recursos ociosos y costos excesivos. Pero, por otro lado, carecer de la capacidad de servicio suficiente causa colas excesivamente largas en ciertos momentos. Cuando los clientes tienen que esperar en una cola para recibir nuestros servicios, están pagando un coste, en tiempo, más alto del que esperaban. Las líneas de espera largas también son costosas por tanto para la empresa ya que producen pérdida de prestigio y pérdida de clientes.

La teoría de las colas en si no resuelve directamente el problema, pero contribuye con la información vital que se requiere para tomar las decisiones concernientes prediciendo algunas características sobre la línea de espera: probabilidad de que se formen, el tiempo de espera promedio.

Pero si utilizamos el concepto de "clientes internos" en la organización de la empresa, asociándolo a la teoría de las colas, nos estaremos aproximando al modelo de organización empresarial "just in time" en el que se trata de minimizar el costo asociado a la ociosidad de recursos en la cadena productiva.

5. REFERENCIAS BIBLIOGRAFICAS.

- A. K. Erlang,(1996) “Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges”, Birkhauser Boston.
- Cao Abad, Ricardo et. al. (2007-2008): Teoría de colas. Departamento de matemáticas (Universidad de A Coruña).
- Domingo Morales González (2003), Teoría de Colas. ISBN 84-605-1045-X.
- Gross, Donaldy CarlHarris (2008).Resumen traducido del libro Fundamentals of Queueing Theory, Wiley.
- Hillier F.S. y Lieberman G.J. (2010). Introducción a la investigación de operaciones. McGrawHill
- https://es.wikipedia.org/wiki/Teor%C3%ADa_de_colas.
- https://www.monografias.com/trabajos18/teoria-colas/teoria-colas.shtml#conclu.
- https://es.wikipedia.org/wiki/Teor%C3%ADa_de_colas
- https://es.wikipedia.org/wiki/Teor%C3%ADa_de_colas.
- http://blogdocente.usfx.bo/ramiro-villegas/2017/06/05/el-sistema-financiero-boliviano/.
- http://ri.ues.edu.sv/id/eprint/12510/1/19200861.pdf.

ANEXOS

Figura 1 Sistema de colas básico

Figura 2 Sistema de colas en espera

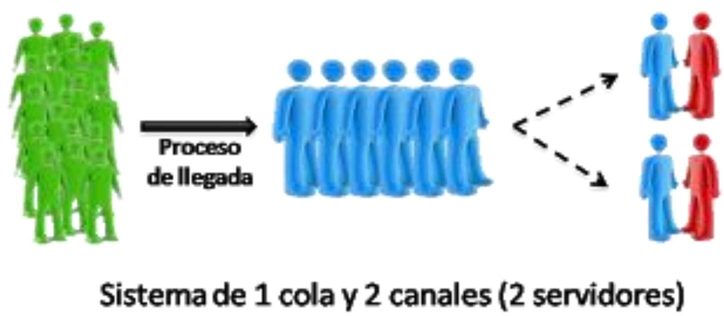


Figura 3 Sistema de colas una cola para cada canal

Figura 4 Sistema de colas multietapa

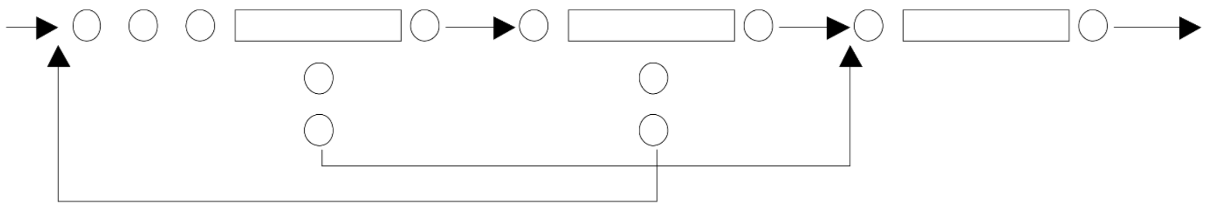


Figura 5 Proceso de nacimiento y muerte de un sistema de colas