
TP 3.0 - SIMULACION DE UN MODELO DE COLA

Fadua Dora Sicardi

Franco Giangiordano

Gonzalo Turconi

Ignacio Curti

4 de junio de 2022

ABSTRACT

En el presente trabajo de investigación, se quiere simular y poder observar el comportamiento de un sistema de colas MM1 y Modelo de Inventario a través del ingreso de distintos parámetros para la obtención de las diferentes medidas de rendimiento.

Keywords

Cola, Servidor, Analítico, Artículo Científico, Modelo, Simulación.

1. Introducción

El objetivo de este trabajo es simular y estudiar el comportamiento de un sistema de colas MM1 y Modelo de Inventario a través del ingreso de parámetros tales como la media de arribo, media de servicios y número de clientes que salen del sistema para la obtención de las distintas medidas de rendimiento.

Pero primero, ¿qué es una cola? Llamamos **cola** a una colección ordenada de ítems donde la adición de nuevos ítems tiene lugar en uno de los extremos. La cola puede estar compuesta teóricamente de cualquier cosa según sea el modelo que necesitemos: personas, procesos, solicitudes, vehículos, etc.

Podemos decir que convivimos a diario con diferentes modelos de colas, y es por eso que es importante estudiar **qué origina una cola**, y además saber cómo modelar correctamente la situación para poder analizar y posteriormente optimizar dicho modelo, haciendo por ejemplo, que los recursos de la computadora sean utilizados de forma eficiente en un sistema operativo, como también hacer que la gente espere lo menos posible en una fila en un supermercado.

¿Por qué se forma una Cola? Las causas fundamentales que hacen inevitables la aparición de una cola son dos: la primera es que la capacidad de los servidores que poseemos sea menor que la demanda requerida, la otra causa posible se relaciona con la variabilidad de los tiempos de servicio, o bien suceden ambas simultáneamente. Las colas producidas por la primera causa no representan una gran complejidad para enfrentarlas, mientras que si sucede la segunda representará un desafío de diseño y gestión.

A lo largo de este trabajo, explicaremos como modelar correctamente un sistema de colas.

2. Marco teórico

Se requiere una breve introducción a los conceptos, sistemas y soluciones analíticas de simulación para analizar y desarrollar los casos de estudio.

El sistema representa un proceso de interés, y para estudiarlo se realizan una serie de suposiciones y supuestos, para posteriormente formar un modelo analítico o, en su defecto, una simulación. Estas suposiciones a menudo toman la forma de relaciones lógicas o matemáticas, que forman el modelo utilizado para tratar de comprender el comportamiento de un sistema. Es importante en simulación poder conocer el estado de un sistema en cualquier momento, definiéndose este último como el conjunto de variables de estado, necesarias para describir un sistema en un momento dado.

Al usar un modelo, necesitamos poder simular el comportamiento del mundo real y asegurarnos de que el modelo funcione correctamente (comparando los resultados del modelo con los resultados experimentales en un sistema real). Una vez que se valida el modelo, puede responder preguntas de interés que a menudo son relevantes. Cambiar el modelo es mucho más conveniente que hacer modificaciones en un sistema real. Esto se debe a factores como el tiempo y

el costo. Si el modelo contiene relaciones simples, se puede resolver analíticamente. Sin embargo, la mayoría de los sistemas del mundo real son demasiado complejos para evaluarlos de esta manera y se investigan mediante simulación.

2.1. Modelo

Un **modelo** es una representación simplificada de la realidad, que facilita su comprensión y el estudio de su comportamiento. Debe mantener un equilibrio entre sencillez y capacidad de representación.

- **Modelo físico:** normalmente son construcciones en escala reducida o simplificada de obras, máquinas o sistemas de ingeniería para estudiar en ellos su comportamiento y permitir así perfeccionar los diseños, antes de iniciar la construcción de las obras u objetos reales. Por ese motivo, a este tipo de modelo se le suele llamar también modelo reducido o modelo simplificado.
- **Modelo analítico:** cualquier esquema simplificado e idealizado de un objeto, fenómeno real, constituido por símbolos y operaciones, relaciones matemáticas; es decir, una forma de representar cada uno de los tipos de entidades que intervienen en un cierto proceso de la realidad mediante objetos matemáticos. Las relaciones matemáticas formales entre los objetos del modelo, deben representar de alguna manera las relaciones reales existentes entre las diferentes entidades o aspectos del sistema u objeto real. Este tipo de modelo requerirá que se pueda seguir el camino inverso al modelado, permitiendo reinterpretar en la realidad las predicciones del modelo. Planteado un modelo matemático, tendremos que optar por realizar una Solución Analítica, y/o una Simulación por computadora.

2.2. Soluciones analíticas

Una vez representado cierto problema en forma de modelo matemático, si este no es extremadamente complejo, se puede proceder a aplicar el cálculo, el álgebra y otras herramientas matemáticas para deducir el comportamiento del sistema bajo estudio. De esta forma, se consiguen predicciones exactas que se reflejan en el sistema real, llamamos a esto solución analítica.

Una forma alternativa podría ser buscar una solución heurística, es decir buscar una solución aproximada a la exacta basado en una estrategia, método, o criterio usado para hacer más sencilla la predicción en modelos complejos. No siempre existe un método heurístico o es conveniente utilizar uno para hallar una solución a un problema complejo, o bien este es muy caótico para modelarlo solo con elementos matemáticos tradicionales y que tenga una solución realista, o quizás el tiempo requerido o el costo de recursos hace que no sea factible utilizar uno.

2.3. Simulación por computadora

Es un intento de modelar situaciones de la vida real por medio de un programa de computadora, lo que requiere ser estudiado para ver cómo es que trabaja el sistema y hacer predicciones sobre su comportamiento y que si bien no son exactas, se apunta a llegar a soluciones realistas y aproximadas a lo que en el sistema sucede. Su comportamiento puede cambiar en cada simulación según el conjunto de parámetros iniciales supuestos por el entorno. En una simulación también se puede buscar generar una muestra de escenarios representativos para un modelo en que una enumeración analítica completa de todos los estados posibles serían imposibles.

2.4. Teoría de Colas

2.4.1. Definición

La teoría de colas es utilizada para estudiar matemáticamente las colas y las líneas de espera en un sistema. Es utilizada para resolver problemas de la vida real como es en el caso del tráfico, las colas en los supermercados, entre otros. Se forman debido a un desequilibrio temporal entre la demanda del servicio y la capacidad del sistema para suministrarlo. 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. El matemático danés Agner Krarup Erlang, publicó el primer artículo sobre la teoría de colas en 1909.

2.4.2. Objetivos

- Identificar el nivel óptimo de capacidad del sistema que minimiza el coste global del mismo.

- Evaluar el impacto que las posibles alternativas de modificación de la capacidad del sistema tendrían en el coste total del mismo.
- Establecer un balance equilibrado óptimo entre las consideraciones cuantitativas de costes y las cualitativas de servicio.
- Prestar atención al tiempo de permanencia en el sistema o en la cola: la paciencia de los clientes depende del tipo de servicio específico considerado y eso puede hacer que un cliente abandone el sistema.

2.4.3. Componentes y Características de un Sistema de Colas

Para modelizar un sistema de colas, hay que tener en cuenta al menos estas 8 características:

- **Proceso básico de colas:** Los clientes que requieren un servicio se generan en una fase de entrada. Estos clientes entran al sistema y se unen a una cola. En determinado momento se selecciona un miembro de la cola, para proporcionarle el servicio, mediante alguna regla conocida como disciplina de servicio. Luego, se lleva a cabo el servicio requerido por el cliente en un mecanismo de servicio, después de lo cual el cliente sale del sistema de colas.
- **Fuente de entrada o población potencial:** Una característica de la fuente de entrada es su tamaño. El tamaño es el número total de clientes que pueden requerir servicio en determinado momento. Puede suponerse que el tamaño es infinito o finito.
- **Cliente:** Es todo individuo de la población potencial que solicita servicio como por ejemplo una lista de trabajo esperando para imprimirse.
- **Capacidad de la cola:** Es el máximo número de clientes que pueden estar haciendo cola (antes de comenzar a ser servidos). De nuevo, puede suponerse finita o infinita.
- **Disciplina de la cola:** La disciplina de la cola se refiere al orden en el que se seleccionan sus miembros para recibir el servicio. Por ejemplo, puede ser:
 - FIFO (first in first out) primero en entrar, primero en salir, según la cual se atiende primero al cliente que antes haya llegado.
 - LIFO (last in first out) también conocida como pila que consiste en atender primero al cliente que ha llegado el último.
 - RSS (random selection of service) que selecciona los clientes de manera aleatoria, de acuerdo a algún procedimiento de prioridad o a algún otro orden.
 - PS (Processor Sharing) sirve a los clientes igualmente. La capacidad de la red se comparte entre los clientes y todos experimentan con eficacia el mismo retraso.
- **Mecanismo de servicio:** El mecanismo de servicio consiste en una o más instalaciones de servicio, cada una de ellas con uno o más canales paralelos de servicio, llamados servidores.
- **Redes de colas:** Sistema donde existen varias colas y los trabajos fluyen de una a otra. Por ejemplo: las redes de comunicaciones o los sistemas operativos multitarea.
- **El proceso de servicio:** Define cómo son atendidos los clientes.

2.5. Enfoque de Modelo MM1

En este modelo se dispone sólo de un canal para dar servicio, las llegadas siguen un proceso de Poisson y la distribución del tiempo de servicio es exponencial. En este modelo la tasa de promedio de arribos al sistema se denomina con λ y la tasa promedio de servicio con μ . La capacidad del sistema es ilimitada, es decir, el tamaño de la cola se toma teóricamente como infinita y la disciplina de la cola es FIFO. El factor de utilización es conocido y es $\rho = \lambda/\mu$. La condición necesaria y suficiente para que un modelo MM1 tenga solución de equilibrio, es que $\rho < 1$, también denominada **condición de estabilidad**.

2.5.1. Medidas de Desempeño en un Modelo de Cola MM1

Teniendo en cuenta que la cantidad de clientes atendidos en el servidor por unidad de tiempo la denominamos con μ y la cantidad de llegadas por unidad de tiempo con λ , podemos definir las siguientes medidas de rendimiento:

- Utilización del Servidor: $\rho = \lambda/\mu$
- Tiempo promedio en la cola: se denomina con W_q .

- Número promedio de clientes en la cola: se calcula como $L_q = \lambda W_q$.
- Tiempo promedio en el sistema: se denomina con W_s y se calcula $W_s = W_q + 1/\mu$.
- Número promedio de clientes en el sistema: se calcula como $L_s = \lambda W_s = L_q + \lambda/\mu$.
- Probabilidad de n clientes en el sistema: $P_n = (1 - p)p^n$
- Probabilidad de denegación de servicio: es la probabilidad de que haya más clientes de lo que la cola puede soportar, por lo tanto se le denega el servicio al cliente: $1 - \sum_{i=0}^n P_i$
- **Promedio esperado de retraso en cola $d(n)$** : Deviene de el parámetro L_q . El estimador para esta medida es:

$$\hat{d} = \frac{\sum_{i=1}^n D_i}{n} \quad (1)$$

- **Promedio esperado de clientes en cola $q(n)$** :Proviene del parámetro L_q . Pero primero es necesario definir, el número de clientes en la cola en el momento t como $Q(t) = L$. El tiempo en el que ocurren n retrasos, es decir el tiempo total de la simulación, como $T(n)$. Considerando,

$$Q(t) \geq 0, \forall 0 \leq t \leq T(n) \quad (2)$$

y p_i la proporción de $T(n)$ en el que $Q(t) = i$, definimos:

$$\hat{q}(n) = \sum_{i=0}^{k-1} i \hat{p}_i \quad (3)$$

Siendo k la cantidad de clientes que soporta el sistema. Es decir que $q(n)$ es el promedio ponderado de los posibles valores i de clientes en cola. El peso es la proporción de tiempo en el que hay dicha cantidad de clientes en espera.

\hat{p}_i es la proporción observada de cada cantidad de clientes en cola.

$$p_i = \frac{T_i}{T(n)} \quad (4)$$

Entonces, reemplazando en la ecuación anterior:

$$\hat{q}(n) = \frac{\sum_{i=0}^{k-1} i T_i}{T(n)} \quad (5)$$

- **Proporción esperada de tiempo del servidor en estado ocupado $u(n)$** : Deviene de la probabilidad de que el servidor no esté vacío, $p_{N>0} = 1 - p_0$. Para calcular su estimador $\hat{\mu}(n)$, primero se define la "función ocupada":

$$B(t) = \begin{cases} 1 & \text{si } N_t > 0 \\ 0 & \text{si } N_t = 0 \end{cases} \quad (6)$$

2.5.2. Subprocedimientos

- **Rutina de inicializacion**: Se ejecuta solamente al inicio de la simulación. Se inicializa el reloj de sistema y las variables de las medidas de rendimiento en 0, y se establece el servidor como disponible. Se configura el Próximo tiempo de partida en infinito (de manera conceptual), en la práctica se setea un valor arbitrario muy grande, y se genera un valor para el Próximo tiempo de arribo. De esta manera, inevitablemente el Tipo del próximo evento será Arribo.
- **Rutina de cronometraje**: Determina Tipo del siguiente evento eligiendo el menor valor entre el Próximo tiempo de arribo y Próximo tiempo de partida de la colección Lista de eventos y avanza el Reloj del sistema al tiempo elegido.

- **Rutina de evento:** Actualiza el Estado del sistema cuando ocurre un tipo determinado de evento. Si lo que ocurre es un arribo actualiza la variable Próximo tiempo de arribo sumando al tiempo actual del reloj un valor aleatorio en base a la distribución de arribo con parámetro λ , si lo que sucede es una partida y entra un nuevo cliente al servidor hace lo mismo para lo mismo para Próximo tiempo de partida en base a la distribución de servicio con parámetro μ . A su vez actualiza las variables que se usarán para calcular las medidas de rendimiento
- **Generador de reporte:** Calcula y devuelve las medidas de rendimiento necesarias cuando termina la simulación.
- **Programa principal:** Determina el flujo de la simulación invocando a cada procedimiento anteriormente mencionado.

2.6. Modelo de Inventarios

Los modelos de inventarios son metodos que ayudan a reducir o minimizar los niveles de inventario requeridos en la produccion.

2.6.1. Ventajas del modelo

Algunas de las ventajas que podemos mencionar son:

- Modelo muy sencillo al calcularse y ponerse en practica, especialmente si se compara con los demas modelos que lo sustituyen en el mercado.
- Demuestra ser eficiente, a pesar de que se basa en estimaciones y predicciones, lo que significa que proporciona excelentes resultados.

2.6.2. Componentes de un modelo de inventario

- **Nivel real del inventario:** El nivel real del inventario (I o $I(t)$) se actualiza tras cada demanda y tras cada llegada de un pedido. Está dado por el antiguo valor del nivel del inventario menos la cantidad de artículos demandados o más el tamaño del pedido, según corresponda. Si la demanda es mayor al nivel de inventario, I será negativo. Si I es negativo al momento del arribo del pedido, se compensará primero, si es posible, el faltante de artículos y luego, si hay sobrante, se añadirá al inventario.
Existen otros dos componentes derivados del nivel real del inventario. I^+ es el máximo entre $I(t)$ y 0. Representa la cantidad de artículos en el inventario que tienen costo de mantenimiento.
 I^- es el máximo entre $-I(t)$ y 0. Representa la cantidad de artículos que deben ser compensados una vez llegue el pedido,
- **Tiempo entre demandas:** Los tiempos entre demandas (T_0) son Variables Aleatorias Independientes distribuidas Idénticamente (VAII). Junto a un generador de numeros pseudoaleatorios generan los tiempos entre una demanda y la siguiente.
- **Tamaño de la demanda:** El tamaño de la muestra (D) es una VAII. Junto a un generador de numeros pseudoaleatorios determina la cantidad de artículos que serán demandados.
- **Número mínimo de artículos en stock:** A final de cada período se verifica si el nivel del inventario es menor al número mínimo de artículos en stock (s). En caso de serlo se realiza el pedido.
- **Número máximo de artículos en stock:** El número máximo de artículos en stock (S) determina el límite de artículos que pueden existir en el inventario al mismo tiempo.
- **Política de pedidos:** La política de pedidos (Z) determina la cantidad de artículos que se piden a final de un período. Está dada por:

$$B(t) = \begin{cases} S - I & \text{si } I < s \\ 0 & \text{si } I \geq s \end{cases} \quad (7)$$

- **Momento en que el proveedor entrega el pedido:** El momento en que el proveedor entrega el pedido (M_p) es una VAII y junto a un generador de numeros pseudoaleatorios determina el tiempo que tarda el pedido en arribar tras ser solicitado al principio del pedido.

- **Costo por pedido:** El costo por pedido (c_p) de una cantidad está dado por la cantidad de artículos solicitados (z) multiplicado por el costo unitario por artículo (i). Se le suele agregar un valor que refiere al costo de administración y/o envío llamado costo fijo (K). Está dada por:

$$c_p = k + iz \quad (8)$$

- **Costo de mantenimiento:** El costo de mantenimiento (C_M) son todos los costos asociados al almacenamiento del inventario hasta que es vendido. Está dada por:

$$C_M = h\bar{I}^+ \quad (9)$$

donde h es el costo de mantenimiento de un artículo por período y \bar{I}^+ es el nivel actual de inventario.

- **Costo por faltante:** Este se da cuando se realiza una demanda de artículos mayor a la cantidad de artículos en el inventario. Está dado por:

$$C_F = \pi\bar{I}^- \quad (10)$$

donde π refiere al costo por faltante por artículo por período y \bar{I}^- refiere a la cantidad de artículos faltantes para cumplir la demanda.

- **Costo total:** El costo total (C_T) es el componente a minimizar. Para hacerlo solo podemos variar los parámetros s y S . Está dado por:

$$C_T = C_p + C_M + C_F \quad (11)$$

3. Exposición de las gráficas de los modelos realizados en la simulación

3.1. Gráficas correspondientes al modelo de inventario

3.1.1. Python

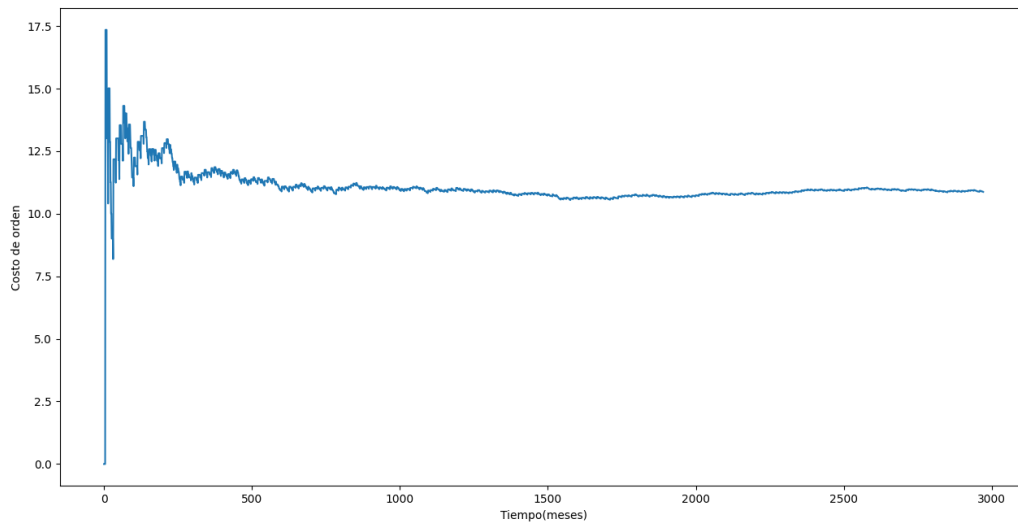


Figura 1: Gráfica costo de orden en meses

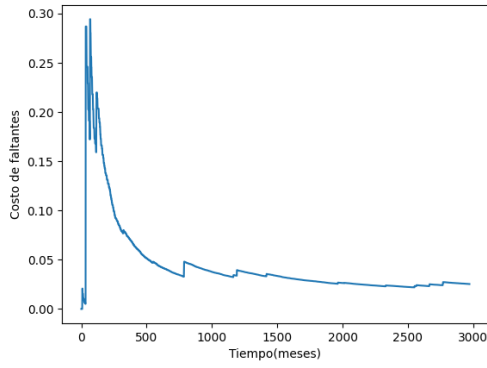


Figura 2: Gráfica de costo de faltantes en meses

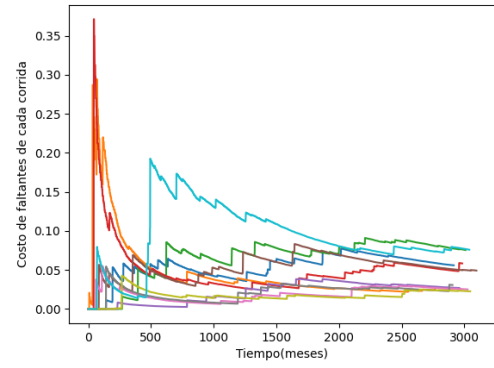


Figura 5: Gráfica de costo de faltante en cada corrida en meses

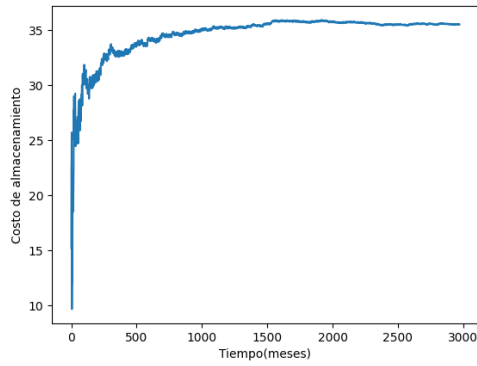


Figura 3: Gráfica de costo de almacenamiento en meses

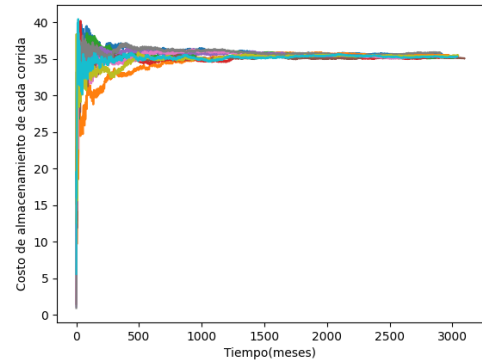


Figura 6: Gráfica de costo de almacenamiento en cada corrida en meses

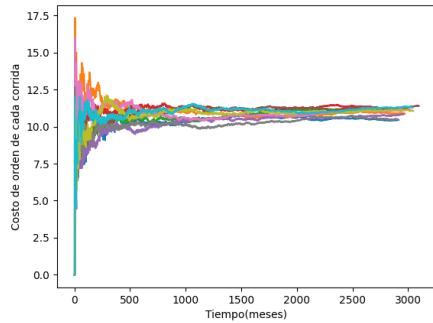


Figura 4: Gráfica de costo de orden en cada corrida en meses

3.2. Gráficas correspondientes al modelo de cola MM1

3.2.1. Python

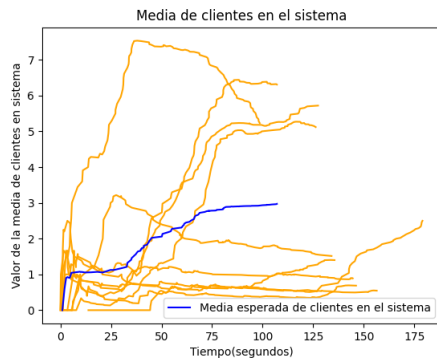


Figura 1: Gráfica promedio de clientes en el sistema.

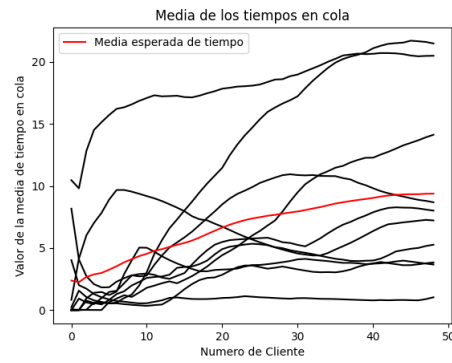


Figura 4: Gráfica tiempo promedio en cola

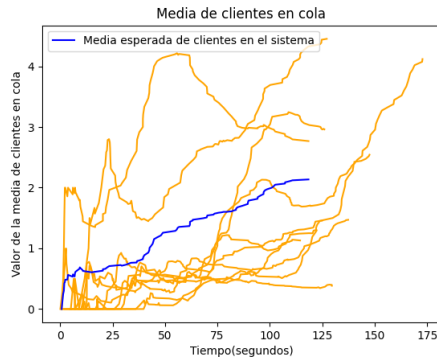


Figura 2: Gráfica promedio de clientes en cola

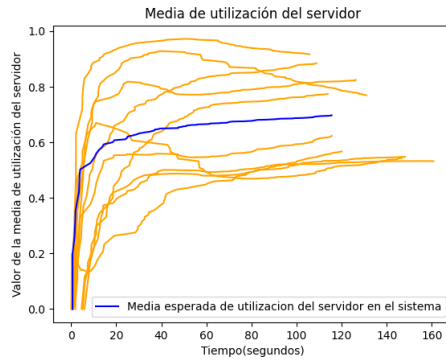


Figura 5: Gráfica utilización del servidor.

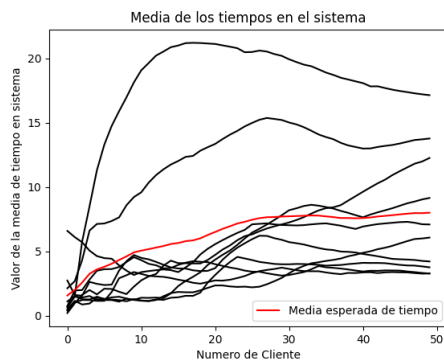


Figura 3: Gráfica tiempo promedio en sistema

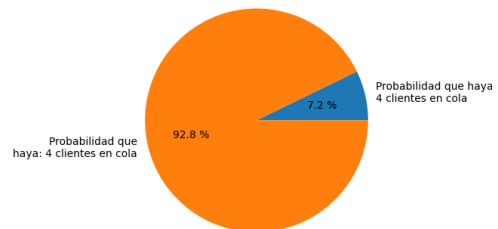


Figura 6: Gráfica probabilidad de encontrar 4 clientes en cola

3.2.2. Anylogic

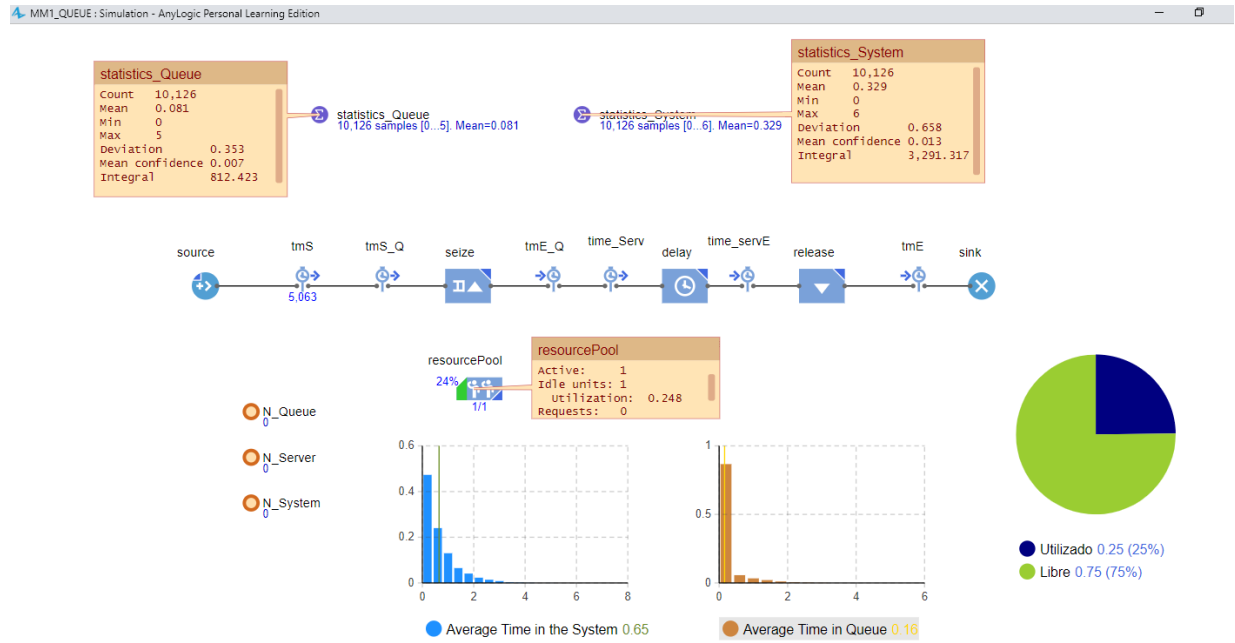


Figura 1: Resultados del primer caso, con tasa de arribos=0.5, tasa de servicio=2

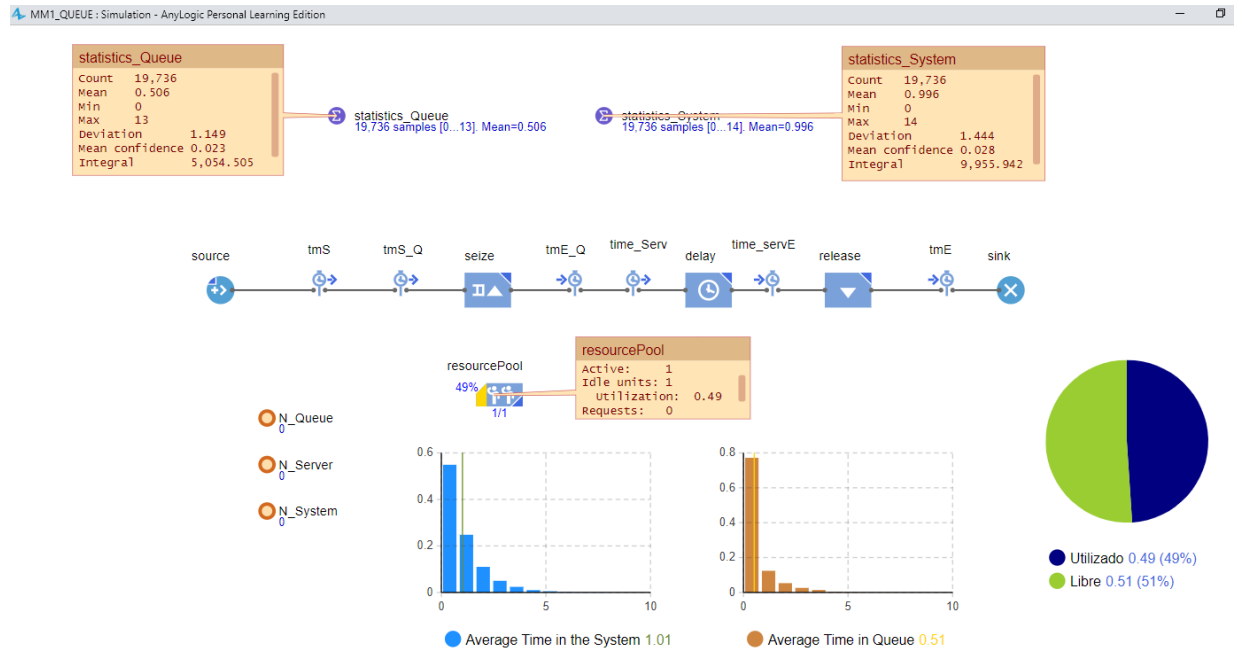


Figura 2: Resultados del primer caso, con tasa de arribos=1.0, tasa de servicio=2

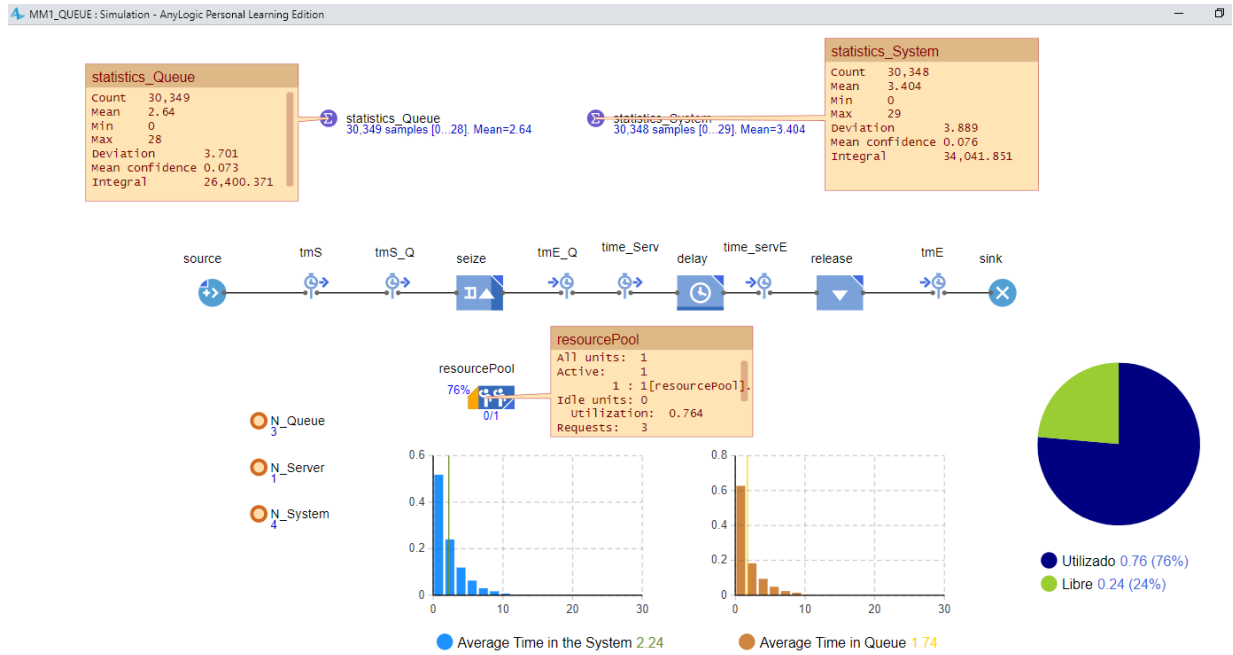


Figura 3: Resultados del primer caso, con tasa de arribos=1.5, tasa de servicio=2

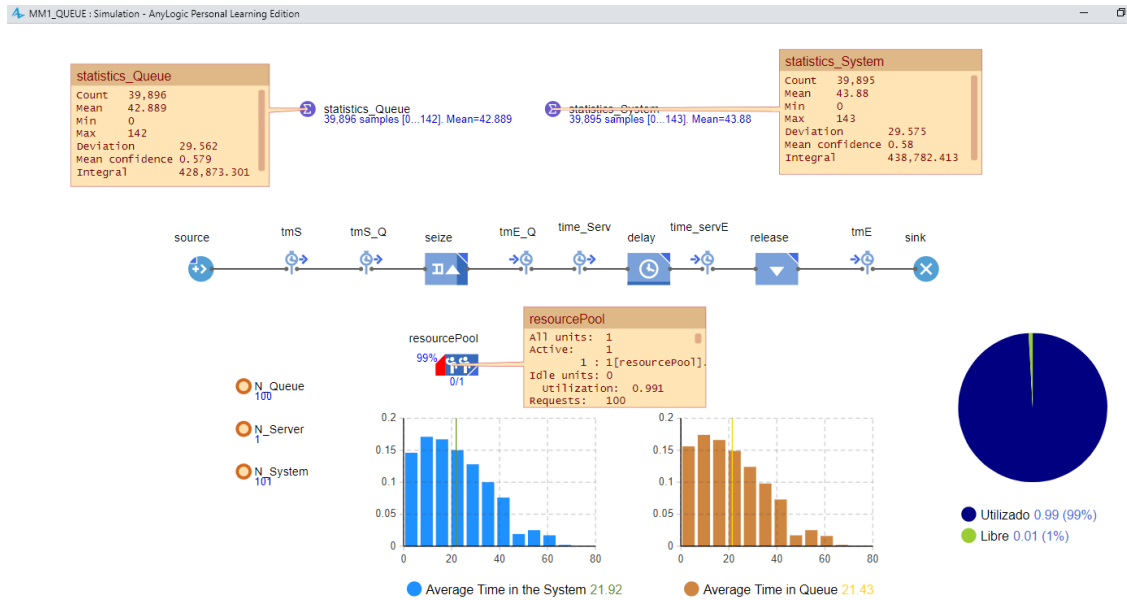


Figura 4: Resultados del primer caso, con tasa de arribos=2.0, tasa de servicio=2

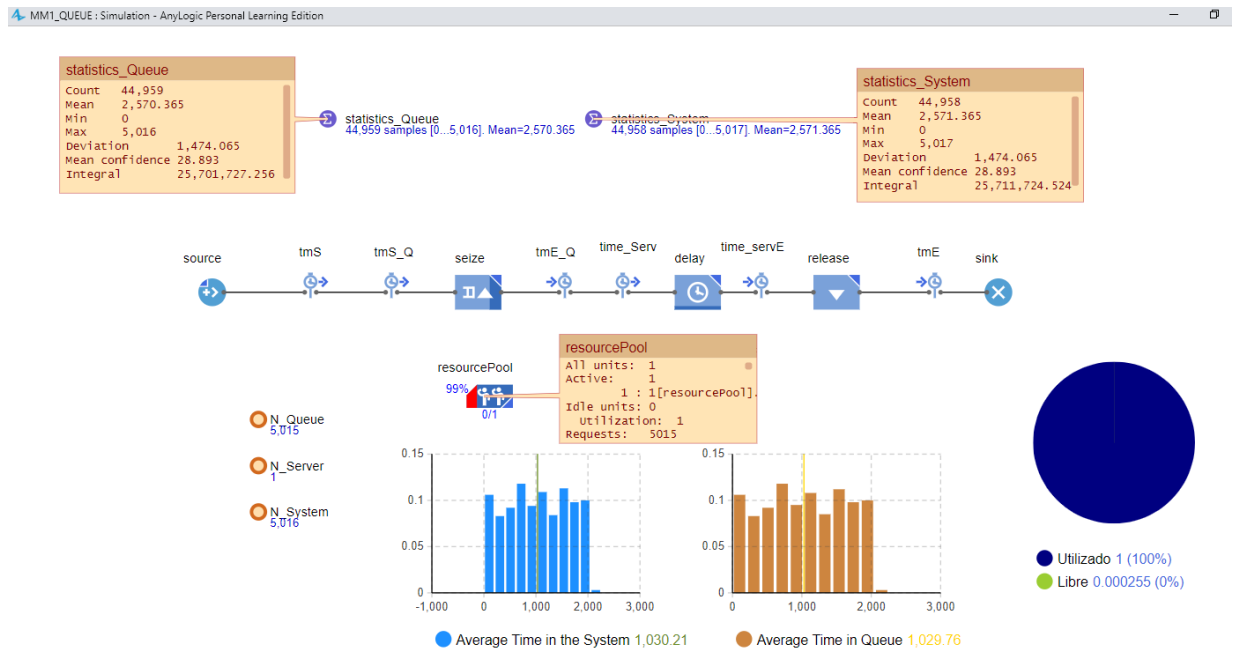


Figura 5: Resultados del primer caso, con tasa de arribos=2.5, tasa de servicio=2

4. Conclusión

Si bien los sistemas de colas tipo M/M/1 pueden ser resueltos mediante el método analítico este resulta ser un método tedioso y largo. Las simulaciones, por otro lado, son capaces de entregar resultados aproximados en cuestión de segundos.

A pesar de que no den un resultado exacto y que una corrida pueda variar moderadamente con respecto a otra, con suficientes tiradas se obtiene un promedio fiable cercano al teórico esperado.

Entre los dos tipos de simulación, programa propio en Python y modelo en Anylogic, encontramos que Anylogic ofrece una gran simplificación a la hora de modelar. No sólo facilita a la hora de construir el modelo sino también al momento de mostrar datos, resultados y gráficas.

Referencias

- [1] Python
<https://python-para-impacientes.blogspot.com/2014/08/graficos-en-ipython.html>
- [2] Anylogic
<https://www.anylogic.com/getting-started/>
- [3] Teoría de colas
https://es.wikipedia.org/wiki/Teor%C3%ADa_de_colas
- [4] Simulation modelling and Analysis - Averill Law
<http://www.averill-law.com/simulation-book/>