TP 3 - SIMULACIÓN DE UN SISTEMA M/M/1

Mateo, Lara
Cátedra Simulación
Universidad Tecnológica Nacional
Zeballos 1341, S2000
1laramateo@gmail.com

Retrivi, Vittorio
Cátedra Simulación
Universidad Tecnológica Nacional
Zeballos 1341, S2000
retrovitto@gmail.com

8 de Julio de 2020

ABSTRACT

Presentamos un análisis sobre un sistema de colas M/M/1 en el cual compararemos los diferentes resultados obtenidos analíticamente, simulando en python y en anylogic.

***K*eywords** Simulación · M/M/1 · Teoría de Colas · Anylogic

1 Introducción

Junto a los árboles de decisiones, con frecuencia los modelos de líneas de espera son útiles para la planificación de la capacidad. Frente a ciertos centros de trabajo, como el mostrador de pasajes de un aeropuerto, un centro de máquinas o un centro de cómputos central, tienden a formarse líneas de espera.

Es así porque los tiempos de llegada entre dos trabajos o clientes sucesivos varían y el tiempo de procesamiento también varía de un consumidor al siguiente. Los modelos de líneas de espera usan distribuciones de probabilidad para ofrecer estimaciones del tiempo de retraso promedio de los clientes, la longitud promedio de las filas de espera y la utilización del centro de trabajo. Los gerentes suelen usar esta información para elegir la capacidad más efectiva en términos de costos, hallando un equilibrio entre el servicio al cliente y el costo de la capacidad agregada.

Se conoce como línea de espera a una hilera formada por uno o varios clientes que aguardan para recibir un servicio. Los clientes pueden ser personas, objetos, máquinas que requieren mantenimiento, contenedores con mercancías en espera de ser embarcados o elementos de inventario a punto de ser utilizados. Las líneas de espera se forman a causa de un desequilibrio temporal ente la demanda de un servicio y la capacidad del sistema para suministrarlo.

En la mayoría de los problemas de líneas de espera que se presentan en la vida real, la tasa de demanda varía; es decir, los clientes llegan a intervalos imprevisibles. Lo más común es que también haya variaciones en el ritmo de producción del servicio, dependiendo de las necesidades del cliente.

Los pacientes que aguardan al médico en su consultorio y los taladros descompuestos que esperan en una instalación de reparación tienen mucho en común desde una perspectiva de Producción y Operaciones. Ambas utilizan recursos humanos y recursos de equipos para mantener los valiosos activos de producción (gente y máquinas) en buenas condiciones.

Los administradores de operaciones reconocen el trueque que se lleva a cabo entre el costo de ofrecer un buen servicio y el costo del tiempo de espera del cliente o la máquina. Los administradores desean que las filas de espera sean lo suficientemente cortas, de tal forma que los clientes no se sientan descontentos y se vayan sin comprar, o que compren pero nunca regresen. Sin embargo, los administradores están dispuestos a permitir alguna espera, si ésta es proporcional a un ahorro significativo en los costos del servicio. Cuando la empresa intenta elevar su nivel de servicio, se observa un incremento en los costos.

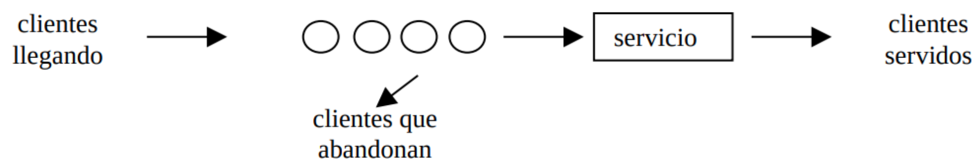
2 Teoría de Colas

Definimos a la teoría de colas como el estudio matemático de las líneas de espera o colas. Se construye un modelo de colas para que se puedan predecir las longitudes de las colas y el tiempo de espera.

La teoría de colas generalmente se considera una rama de la investigación de operaciones ya que los resultados a menudo se usan al tomar decisiones comerciales sobre los recursos necesarios para proporcionar un servicio.

2.1 Descripción de un sistema de colas

Un sistema de colas se puede describir de la siguiente forma. Un conjunto de "clientes" llega a un sistema en busca de un servicio, esperan si este no es inmediato, y abandonan el sistema una vez hayan sido atendidos. El término "cliente" es usado con un sentido genérico. el cliente puede ser ya sea un humano esperando la cola de un supermercado, piezas esperando su turno para ser procesadas o una lista de trabajo esperando para ser impresos en una impresora en red.



2.2 Características de los sistemas de colas

Para describir de forma adecuada a un sistema de cola se utilizan múltiples parámetros. Los mismos se describen a continuación.

Patrón de llegada de los clientes

Se corresponde con la forma en que los clientes llegan al sistema. Comúnmente, y en situaciones de cola habituales, la llegada es estocástica, es decir la llegada depende de una cierta variable aleatoria.

Patrones de servicio de los servidores

En general, los servidores poseen un tiempo de servicio variable, el cual sigue una determinada función de probabilidad.

Disciplina de cola

La disciplina de cola es la manera en que los clientes se ordenan al momento de ser servidos de entre los de la cola. Normalmente una cola normal sigue una disciplina FIFO (se atiende primero a quien llega primero). Sin embargo, es también habitual encontrar colas con el uso de la disciplina LIFO (atender primero al último), o que sigan reglas de secuencia basada en prioridades.

Capacidad del sistema

Es la cantidad de clientes admitidos por el sistema que pueden esperar en la cola. Un sistema de colas puede tener una capacidad ya sea finita o infinita.

Número de canales del servicio

Es la cantidad de servidores que se encuentran dando servicio. Cuando se habla de canales de servicio múltiples, se habla generalmente de una única cola alimentando a varios servidores.

2.2.1 Notación

- λ = Número de llegadas por unidad de tiempo
- μ = Número de servicios por unidad de tiempo si el servidor está ocupado
- c = Número de servidores en paralelo
- $\rho = \frac{\lambda}{c*\mu}$: Congestión de un sistema con parámetros: (λ, μ, c)
- $P_n(t)$: Probabilidad que haya n clientes en el sistema en el instante t=PrN(t)=n
- N: Número de clientes en el sistema en el estado estable
- P_n : Probabilidad de que haya n clientes en estado estable Pn=PrN=n
- L_q : Número medio de clientes en la cola
- T_q : Representa el tiempo que un cliente invierte en la cola
- S : Representa el tiempo de servicio
- $T = T_{q+S}$: Representa el tiempo total que un cliente invierte en el sistema
- $W_q = E[T_q]$: Tiempo medio de espera de los clientes en la cola
- P_b: probabilidad de que cualquier servidor esté ocupado

2.2.2 Rendimiento del sistema de cola

La tarea de un analista de colas se puede dividir en dos tipos: establecer mecanismos para medir la efectividad del sistema o diseñar un sistema "óptimo" (de acuerdo a algún criterio).

Diseñar eficientemente consiste, básicamente, en definir un sistema cuyo coste (de diseño y de operación) se justifique por el servicio que da. Dicho servicio se puede evaluar mediante el coste de "no darlo". De este modo al diseñar se pretende minimizar unos supuestos costes totales.

A partir de los datos que nos suministra la teoría de colas se puede obtener la información necesaria para definir el número de asientos necesarios en una sala de espera, o la estructura de etapas de un proceso de atención al cliente.

En cualquier caso, para poder tomar decisiones hacen falta datos que la teoría de colas puede dar en alguno de los siguientes tres aspectos:

1. tiempo de espera (en el total del sistema o en la cola)
2. cantidad de clientes esperando (en el sistema o en las colas)
3. tiempo ocioso de los servidores (total o particular de cada servicio)

El número medio de clientes en el sistema y en la cola se puede calcular de diferentes maneras:

$$L = E[n] = \sum_{n=0}^{\infty} np_n \tag{1}$$

$$L_q = E[n_q] = \sum_{n=c+1}^{\infty} (n-c)p_n \tag{2}$$

Sabemos por el teorema de Little, que la relación entre la longitud de la cola y el tiempo de espera es:

$$L = \lambda W \tag{3}$$

$$L_q = \lambda W_q \tag{4}$$

El tiempo de estancia de un cliente en el sistema se relaciona con el tiempo de espera de un cliente en la cola,

$$W = W_q + \frac{1}{\mu} \tag{5}$$

El número de clientes que por término medio se están atendiendo en cualquier momento es:

$$r = L - L_q = \lambda(W - W_q) = \frac{\lambda}{\mu} \tag{6}$$

En un sistema de un único servidor:

$$L - L_q = \sum_{n=1}^{\infty} np_n - \sum_{n=1}^{\infty} (n-1)p_n = \sum_{n=1}^{\infty} p_n = 1 - p_0 \tag{7}$$

La probabilidad de que un sistema de un único servidor esté vacío es $p_0 = 1 - p$.

La probabilidad de que un servidor (de un sistema de c servidores en paralelo) esté ocupado en el estado estable es:

$$p_b = \rho = \frac{\lambda}{c\mu} \tag{8}$$

2.3 Clasificación de los sistemas de cola -Notación Kendall

Al reconocer la diversidad de los sistemas de colas, Kendall [1953] propuso un sistema de notación para los sistemas de servidores paralelos el cual ha sido ampliamente adoptado.

Una versión resumida de esta convención está basada en el formato $A/B/c/N/K$.

Estas letras representan las siguientes características del sistema:

- A representa la distribución del tiempo de arribo.
- B representa la distribución del tiempo de servicio. [Los símbolos comunes para A y B incluyen M (exponencial o Markov), D (constante o determinista), Ek (Erlang de orden k), PH (tipo de fase), H (hiperexponencial), G (arbitrario o general), y GI (general independiente).]
- c representa el número de servidores paralelos.
- N representa la capacidad del sistema.
- K representa el tamaño de la población

2.4 Modelo de colas M/M/1

Para el ensayo, analizaremos un modelo de cola simple o también llamado M/M/1 De acuerdo a la notación previamente descripta, En este modelo se dispone de un sólo canal para dar servicio, las llegadas siguen un proceso de Poisson y la distribución del tiempo de servicio es exponencial. La capacidad del sistema es ilimitada y la disciplina de la cola es FIFO (First In - First Out)

Usaremos la siguiente notación:

- t_i: tiempo de arribo (o entre llegadas)del i-ésimo cliente (t_0=0)
- $A_i = t_i - t_{i-1}$: tiempo de arribos entre el $(i-1)$-ésimo y los arribos de los i-ésimos clientes
- S_i: tiempo que el servidor gasta atendiendo al i-ésimo cliente (Exclusivo del retraso del cliente en la cola)
- D_i: retraso de la cola del i-ésimo cliente
- $c_i = t_i + D_i + S_i$: tiempo en el que el i-ésimo cliente completa el servicio y se retira
- e_i: tiempo de ocurrencia del i-ésimo evento de cualquier tipo (i-ésimo valor que toma el reloj de la simulación, excluyendo el valor $e_0 = 0$)

Considere un sistema de colas de servidor único para el cual los tiempos de arribo $A_1, A_2, ...$ son variables aleatorias independientes, distribuidas idénticamente (ID) (" Distribuido idénticamente "significa que los tiempos entre llegadas tienen la misma distribución de probabilidad).

Un cliente quien llega y encuentra que el servidor inactivo ingresa al servicio inmediatamente, y los tiempos de servicio $S_1, S_2, ...$ de los clientes sucesivos son variables aleatorias de ID que son independientes de los tiempos entre llegadas.

Un cliente que llega y encuentra que el servidor está ocupado se une al final de una sola cola. Al completar el servicio para un cliente, el servidor elige a un cliente de la cola (si corresponde) en una primera entrada. primera salida (FIFO).

La simulación comenzará en el estado "vacío e inactivo"; es decir, no hay clientes presentes y el servidor está inactivo. En el momento 0, comenzaremos a esperar la llegada del primer cliente, que ocurrirá después del primer tiempo de arribo, A_1 "en lugar de en el momento 0 (lo que sería posiblemente válido, pero diferente, supuesto de modelado).

Deseamos simular este sistema hasta que un número fijo (n) de clientes haya completado sus retrasos en la cola, es decir, la simulación se detendrá cuando el enésimo cliente entre en servicio. Tenga en cuenta que el tiempo de finalización de la simulación es, por lo tanto, una variable aleatoria, dependiendo de los valores observados para las variables aleatorias entre llegadas y tiempo de servicio.

2.4.1 Utilización del servidor

$$\rho = \frac{\lambda}{\mu} \tag{9}$$

2.4.2 Probabilidad de n clientes en cola

A partir de los datos anteriores y mediante el análisis de procesos de nacimiento y muerte se puede concluir en que la probabilidad de que haya n clientes en el sistema es:

$$P_n = (1-\rho)\rho^n \text{ con } \rho = \frac{\lambda}{\mu} \tag{10}$$

2.4.3 Promedio de clientes en el sistema

$$L = E[n] = \sum_{n=0}^{\infty} np_n = (1-\rho)\sum_{n=0}^{\infty} n\rho^n = (1-\rho)\rho\sum_{n=0}^{\infty} n\rho^{n-1}$$

dado que

$$\sum_{n=0}^{\infty} n\rho^{n-1} = \frac{\partial(\sum n\rho^n)}{\partial\rho} = \frac{\partial(\frac{1}{1-\rho})}{\partial\rho} = \frac{1}{(1-\rho)^2}$$

se concluye que:

$$L = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda} \tag{11}$$

2.4.4 Promedio de clientes en la cola

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} \tag{12}$$

2.4.5 Tiempo promedio en el sistema

$$W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda} \tag{13}$$

2.4.6 Tiempo promedio en la cola

$$W_q = \frac{\rho}{\mu-\lambda} \tag{14}$$

2.4.7 Probabilidad de denegación del servicio

$$\text{Probabilidad de denegación del servicio} = 1 - \sum_{i=0}^{n}(1-\rho)\rho^i \tag{15}$$

2.4.8 Estimación del retraso promedio del cliente

Calcularemos el retraso promedio esperado en la cola de los clientes que completan sus retrasos durante la simulación; denotamos esta cantidad por $d(n)$.

Una interpretación es que $d(n)$ es el promedio de un gran número (en realidad, infinito) de retrasos promedio de n clientes. De una sola ejecución de la simulación que resulta en retrasos del cliente $D_1, D_2, ..., D_n$, un estimador obvio de $d(n)$ es

$$\hat{d}(n) = \frac{\sum_{i=1}^{n} D_i}{n} \tag{16}$$

La estimación de $d(n)$ proporciona información sobre el rendimiento del sistema desde el punto de vista de los clientes.

2.4.9 Estimación de promedio de clientes en la cola

Una de esas medidas para nuestro modelo simple aquí es el número promedio esperado de clientes en la cola, (pero no atendidos), denotado por $q(n)$, donde el n es necesario en la notación para indicar que este promedio se toma sobre el período de tiempo necesario para observar los n retrasos que definen nuestra regla de detención

Este es un tipo diferente de "promedio" que el retraso promedio en la cola, porque se toma en tiempo (continuo), en lugar de clientes (siendo discreto). Por lo tanto, debemos definir qué se entiende por este número de tiempo promedio de clientes en cola. Para hacer esto, deje que $Q(t)$ denote el número de clientes en cola en el tiempo t, para cualquier número real $t \geq 0$, y deje que $T(n)$ sea el tiempo requerido para observar nuestros n retrasos en la cola. Luego, para cada tiempo t entre 0 y $T(n)$, $Q(t)$ es un entero no negativo.

Además, si dejamos que p_i sea la proporción esperada (que será entre 0 y 1) del tiempo en que $Q(t)$ es igual a i, entonces una definición razonable de $q(n)$ sería

$$q(n) = \sum_{i=0}^{\infty} i p_i$$

Por lo tanto, $q(n)$ es un promedio ponderado de los valores posibles i para la longitud de la cola $Q(t)$, siendo los pesos la proporción esperada de tiempo que la cola pasa en cada una de sus posibles longitudes. Para estimar $q(n)$ a partir de una simulación, simplemente reemplazamos los p_i's con estimaciones de ellos, y obtenemos

$$\hat{q}(n) = \sum_{i=0}^{\infty} i \hat{p}_i \tag{17}$$

donde $\hat{p}_i$ es la proporción observada (en lugar de la esperada) del tiempo durante la simulación de que había i clientes en la cola.

Sin embargo, computacional-mente, es más fácil reescribir $\hat{q}(n)$ usando algunas consideraciones geométricas. Si dejamos que T_i sea el tiempo total durante la simulación de que la cola tiene una longitud i, entonces $T(n) = T_o + T_1 + T_2 + ...$ y $\hat{p}_i = \frac{T_i}{T(n)}$), para que podamos reescribir (18) como

$$\hat{q}(n) = \frac{\sum_{i=0}^{\infty} i T_i}{T(n)} \tag{18}$$

La suma en el numerador de la ecuación (18) es solo el área bajo la curva $Q(t)$ entre el comienzo y el final de la simulación: Recordando que "área bajo una curva" es una integral, podemos escribir

$$\sum_{i=0}^{\infty} i T_i = \int_0^{T(n)} Q(t) dt$$

y el estimador de $q(n)$ puede expresarse entonces como

$$\hat{q}(n) = \int_0^{T(n)} \frac{Q(t) dt}{T(n)} \tag{19}$$

2.4.10 Estimación de utilización del servidor

La utilización esperada del servidor es la proporción esperada de tiempo durante la simulación [del tiempo 0 al tiempo $T(n)$] de que el servidor está ocupado (es decir, no inactivo), y por lo tanto es un número entre 0 y 1; denotarlo por $u(n)$.

A partir de una sola simulación, entonces, nuestra estimación de $u(n)$ es $\hat{u}(n)$, que es la proporción de tiempo observada durante la simulación que el servidor está ocupado.

Ahora $\hat{u}(n)$ podría calcularse directamente a partir de la simulación observando las horas en que el servidor cambia de estado (inactivo a ocupado o viceversa) y luego haciendo las sustracciones y división apropiadas. Sin embargo, es más fácil considerar esta cantidad como un promedio de tiempo continuo, similar a la longitud promedio de la cola, definiendo la "función ocupada"

$$B(t) = \begin{cases} 0 & \text{si el servidor esta ocupado en el tiempo t} \\ 1 & \text{si el servidor esta inactivo en el tiempo t} \end{cases}$$

y $\hat{u}(n)$ podría expresarse como la proporción de tiempo en que $B(t) = 1$. Entonces

$$\hat{u}(n) = \frac{\int_0^{T(n)} B(t)dt}{T(n)} \tag{20}$$

y vemos nuevamente que $\hat{u}(n)$ es el promedio continuo de la función $B(t)$ correspondiente a nuestra noción de utilización.

3 Resultados

3.1 Tasa de arribo del 25% respecto a la tasa de servicio

Utilizando los parámetros:

- $\lambda = 1$
- $\mu = 4$

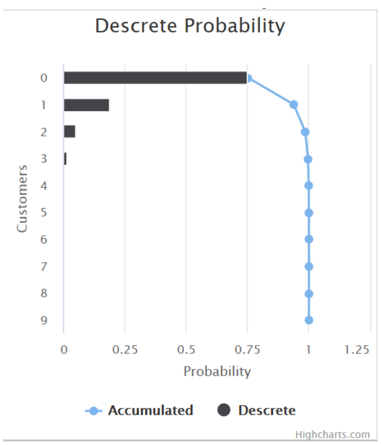
3.1.1 Forma analítica

De forma analítica obtenemos:

Utilización del servidor: $\rho = \frac{\lambda}{\mu} = \frac{1}{4} = 0.25$

Probabilidad de n clientes en la cola: $P_n = (1-\rho)\rho^n$
Ejemplificando, la probabilidad de:

- 1 cliente en la cola: 0.1875
- 2 clientes en la cola: 0.0469
- 3 clientes en la cola: 0.0117



Promedio de clientes en el sistema: $L = \frac{\lambda}{\mu-\lambda} = \frac{1}{4-1} = \frac{1}{3} = 0.3333$

Promedio de clientes en la cola: $L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{1^2}{4(4-1)} = \frac{1}{12} = 0.0833$

Tiempo promedio en el sistema: $W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda} = \frac{1}{4-1} = \frac{1}{3} = 0.3333$

Tiempo promedio en la cola: $W_q = \frac{\rho}{\mu-\lambda} = \frac{\frac{1}{4}}{4-1} = \frac{1}{12} = 0.0833$

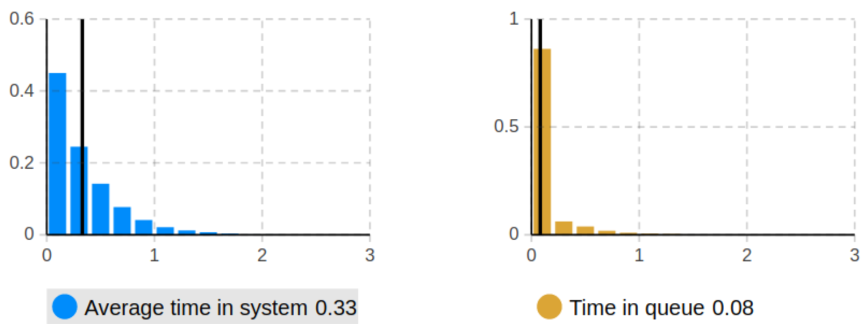
Probabilidad de denegación de servicio: $1 - \sum_{i=0}^{n}(1-\rho)\rho^i$
Ejemplificando, la probabilidad de denegar el servicio con una cola finita de:

- 0 clientes: 0.25
- 2 clientes: 0.0156
- 5 clientes: 0.0003
- 10 clientes: 0.0001
- 50 clientes: tiende a 0

3.1.2 Simulación en AnyLogic

Tras realizar una simulación utilizando el software AnyLogic y utilizando los mismos parámetros de λ y μ obtenemos los siguientes resultados:

- Promedio de clientes en el sistema: 0.329 (sobre un total de 19970 muestras)
- Promedio de clientes en la cola: 0.081 (sobre un total de 19970 muestras)
- Tiempo promedio en el sistema: 0.33
- Tiempo promedio en la cola: 0.08



3.1.3 Simulación en Python

Así mismo, realizamos la misma simulación en el lenguaje de programación Python utilizando los mismos parámetros λ y μ con el objetivo de comparar los datos obtenidos. El programa finaliza al superar los 2000 clientes atendidos. Los valores obtenidos son los siguientes:

- Utilización del servidor (promedio de 10 experimentos): 0.25071284469874616
- Promedio de clientes en el sistema (promedio de 10 experimentos): 0.33831062522505245
- Promedio de clientes en la cola (promedio de 10 experimentos): 0.08077609488042767
- Tiempo promedio en el sistema (promedio de 10 experimentos): 0.3369349508011162
- Tiempo promedio en la cola (promedio de 10 experimentos): 0.08269125300722595
- Probabilidad de n clientes en la cola
  - 1 cliente en la cola: 0.19
  - 2 clientes en la cola: 0.043164
  - 3 clientes en la cola: 0.0093576
- Probabilidad de denegar el servicio con una cola finita de
  - 0 clientes (promedio de 10 experimentos): 0.25025
  - 2 clientes (promedio de 10 experimentos): 0.01405
  - 5 clientes (promedio de 10 experimentos): 0.00015
  - 10 clientes (promedio de 10 experimentos): 0
  - 50 clientes (promedio de 10 experimentos): 0

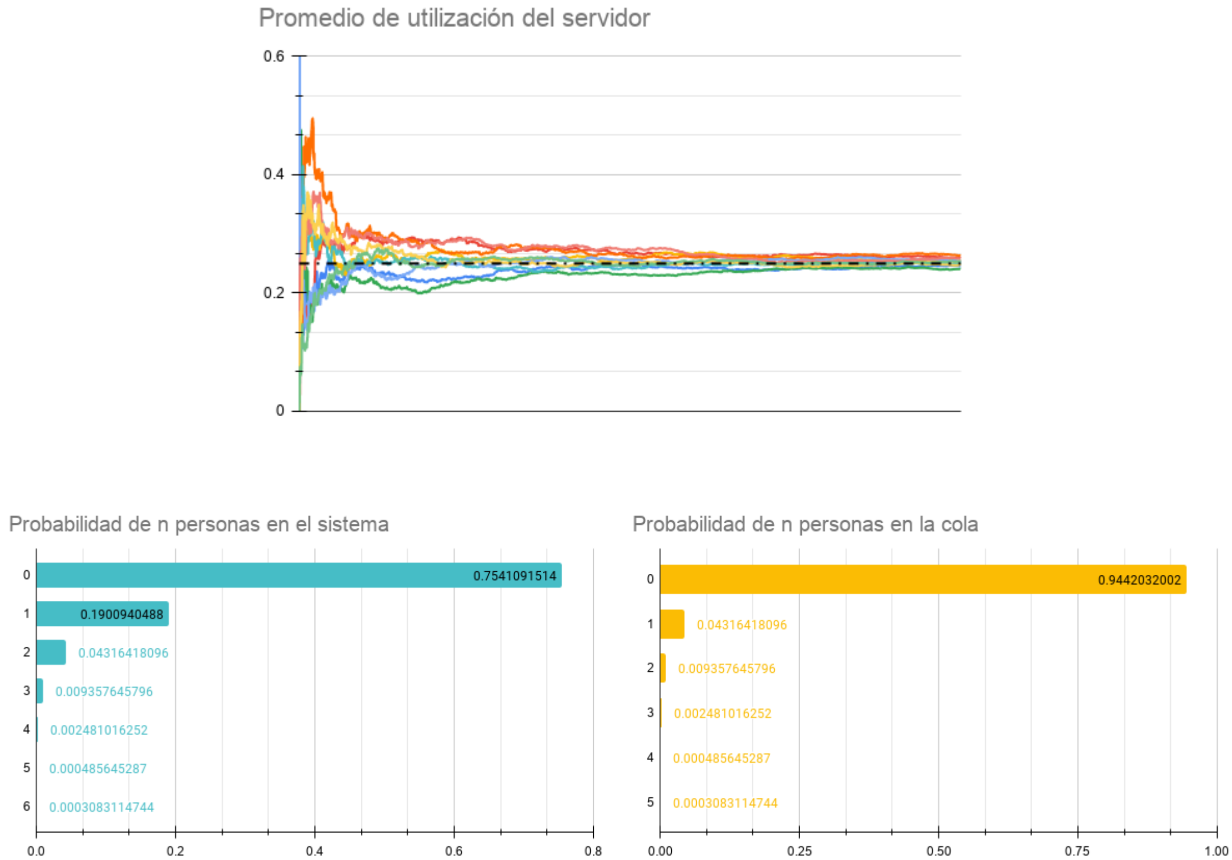


Figure 1: Resultados correspondientes a los datos del 1er experimento

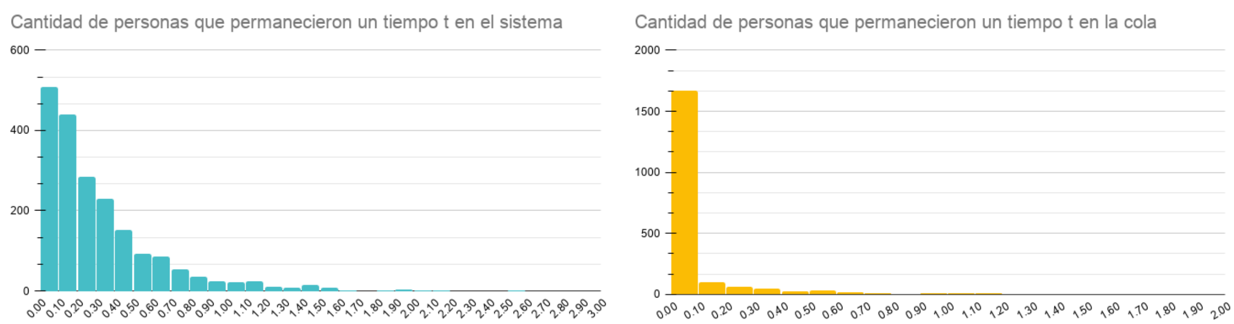


Figure 2: Resultados correspondientes a los datos del 1er experimento

3.2 Tasa de arribo del 50% respecto a la tasa de servicio

Utilizando los parámetros:

- $\lambda = 4$
- $\mu = 8$

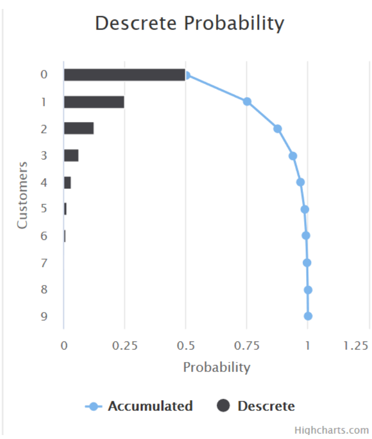
3.2.1 Forma analítica

De forma analítica obtenemos:

Utilización del servidor: $\rho = \frac{\lambda}{\mu} = \frac{4}{8} = 0.5$

Probabilidad de n clientes en la cola: $P_n = (1-\rho)\rho^n$
Ejemplificando, la probabilidad de:

- 1 cliente en la cola: 0.25
- 2 clientes en la cola: 0.125
- 3 clientes en la cola: 0.0625



Promedio de clientes en el sistema: $L = \frac{\lambda}{\mu-\lambda} = \frac{4}{8-4} = \frac{1}{1} = 1$

Promedio de clientes en la cola: $L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{4^2}{8(8-4)} = \frac{16}{32} = 0.5$

Tiempo promedio en el sistema: $W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda} = \frac{1}{8-4} = \frac{1}{4} = 0.25$

Tiempo promedio en la cola: $W_q = \frac{\rho}{\mu-\lambda} = \frac{\frac{4}{8}}{8-4} = \frac{1}{8} = 0.125$

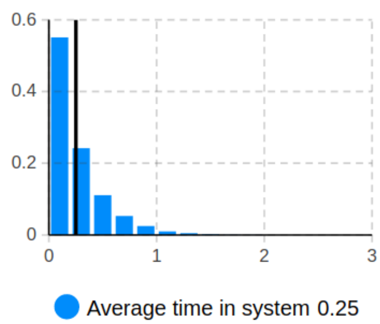
Probabilidad de denegación de servicio: $1 - \sum_{i=0}^{n}(1-\rho)\rho^i$
Ejemplificando, la probabilidad de denegar el servicio con una cola finita de:

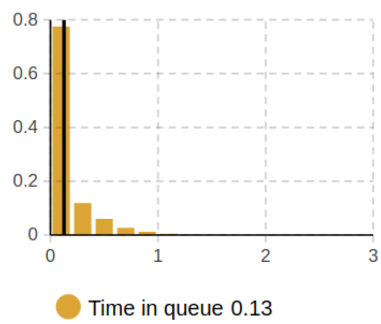
- 0 clientes: 0.5
- 2 clientes: 0.125
- 5 clientes: 0.015625
- 10 clientes: 0.00042
- 50 clientes: tiende a 0

3.2.2 Simulación en AnyLogic

Tras realizar una simulación utilizando el software AnyLogic y utilizando los mismos parámetros de λ y μ obtenemos los siguientes resultados:

- Promedio de clientes en el sistema: 1 (sobre un total de 79572 muestras)
- Promedio de clientes en la cola: 0.5 (sobre un total de 79572 muestras)
- Tiempo promedio en el sistema: 0.25
- Tiempo promedio en la cola: 0.13





3.2.3 Simulación en Python

Así mismo, realizamos la misma simulación en el lenguaje de programación Python utilizando los mismos parámetros λ y μ con el objetivo de comparar los datos obtenidos. El programa finaliza al superar los 2000 clientes atendidos. Los valores obtenidos son los siguientes:

- Utilización del servidor (promedio de 10 experimentos): 0.5020193140134399
- Promedio de clientes en el sistema (promedio de 10 experimentos): 1.0717223644587448
- Promedio de clientes en la cola (promedio de 10 experimentos): 0.4687186301365463
- Tiempo promedio en el sistema (promedio de 10 experimentos): 0.24587229382377881
- Tiempo promedio en la cola (promedio de 10 experimentos): 0.12237133193562974
- Probabilidad de n clientes en la cola
  - 1 cliente en la cola: 0.23333
  - 2 clientes en la cola: 0.12212
  - 3 clientes en la cola: 0.061308
- Probabilidad de denegar el servicio con una cola finita de
  - 0 clientes (promedio de 10 experimentos): 0.528
  - 2 clientes (promedio de 10 experimentos): 0.12458
  - 5 clientes (promedio de 10 experimentos): 0.1451
  - 10 clientes (promedio de 10 experimentos): 0
  - 50 clientes (promedio de 10 experimentos): 0

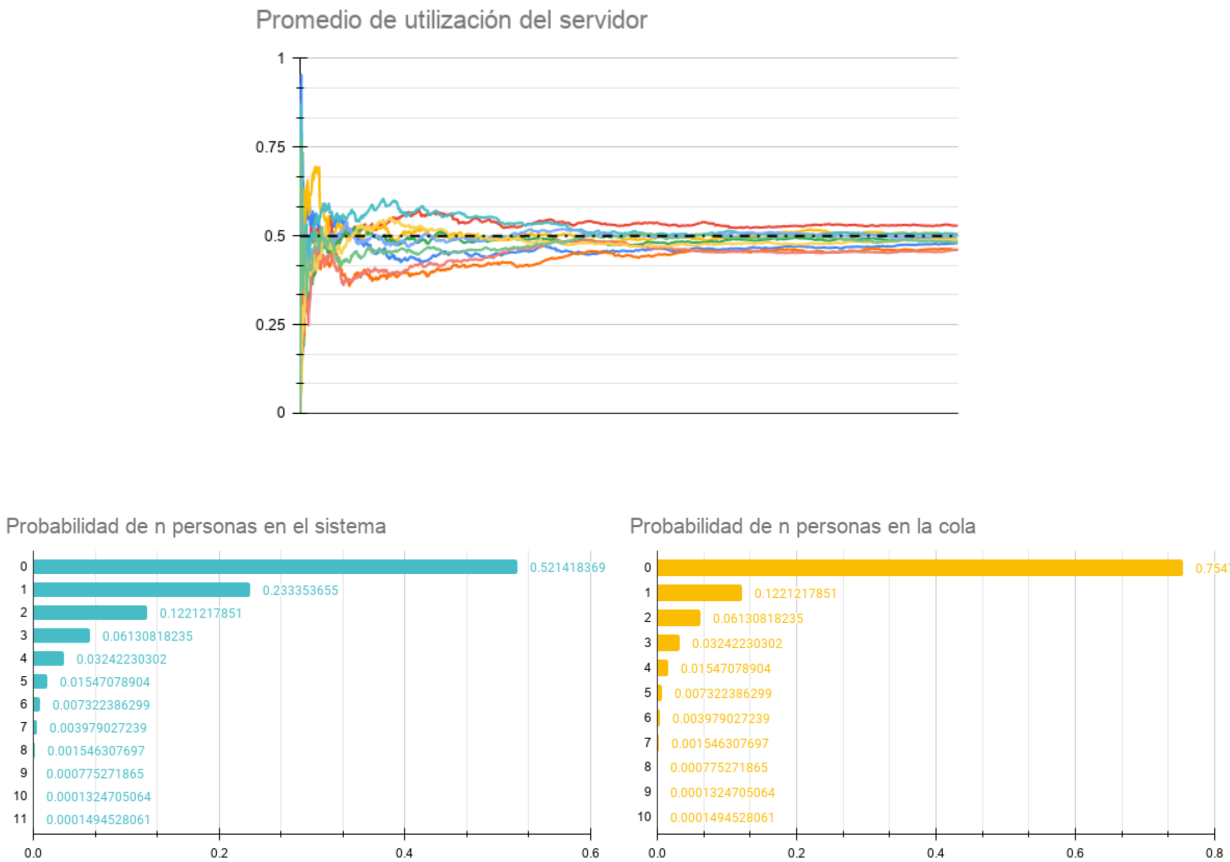


Figure 3: Resultados correspondientes a los datos del 1er experimento

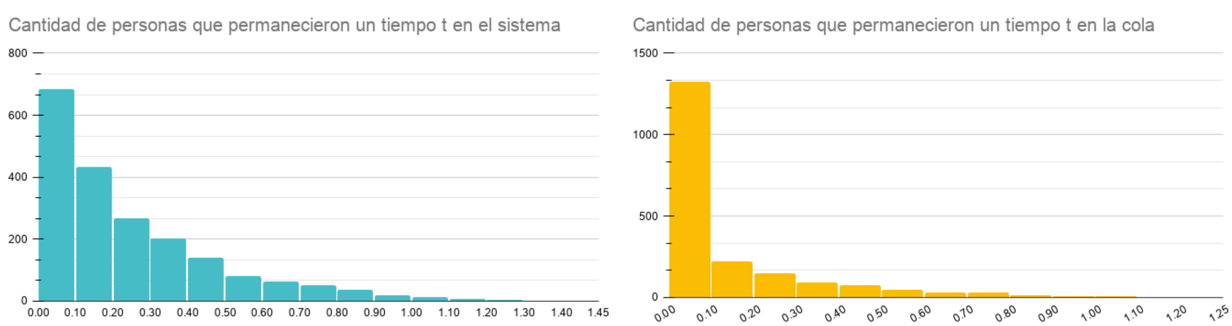


Figure 4: Resultados correspondientes a los datos del 1er experimento

3.3 Tasa de arribo del 75% respecto a la tasa de servicio

Utilizando los parámetros:

- $\lambda = 3$
- $\mu = 4$

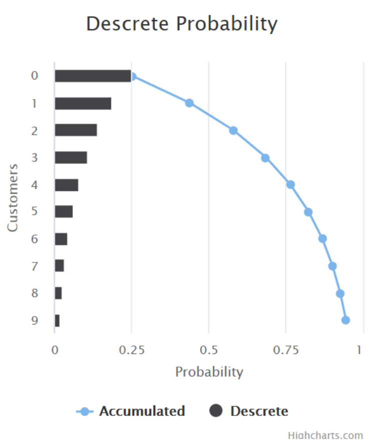
3.3.1 Forma analítica

De forma analítica obtenemos:

Utilización del servidor: $\rho = \frac{\lambda}{\mu} = \frac{3}{4} = 0.75$

Probabilidad de n clientes en la cola: $P_n = (1-\rho)\rho^n$
Ejemplificando, la probabilidad de:

- 1 cliente en la cola: 0.1875
- 2 clientes en la cola: 0.1406
- 3 clientes en la cola: 0.1055



Promedio de clientes en el sistema: $L = \frac{\lambda}{\mu-\lambda} = \frac{3}{4-3} = \frac{1}{1} = 3$

Promedio de clientes en la cola: $L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{3^2}{4(4-3)} = \frac{9}{4} = 2.25$

Tiempo promedio en el sistema: $W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda} = \frac{1}{4-3} = \frac{1}{1} = 1$

Tiempo promedio en la cola: $W_q = \frac{\rho}{\mu-\lambda} = \frac{\frac{3}{4}}{4-3} = \frac{3}{4} = 0.75$

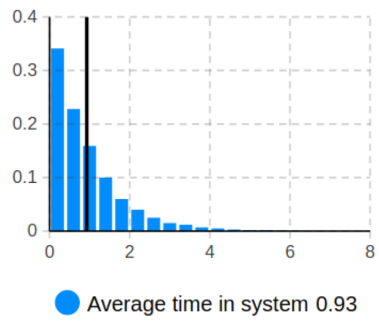
Probabilidad de denegación de servicio: $1 - \sum_{i=0}^{n}(1-\rho)\rho^i$
Ejemplificando, la probabilidad de denegar el servicio con una cola finita de:

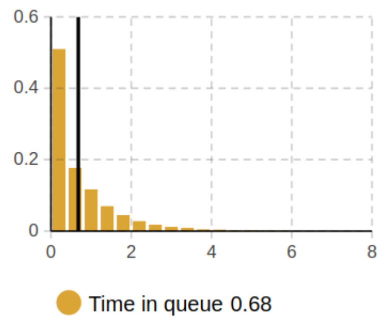
- 0 clientes: 0.75
- 2 clientes: 0.421875
- 5 clientes: 0.17797851562
- 10 clientes: 0.04223513603
- 50 clientes: tiende a 0

3.3.2 Simulación en AnyLogic

Tras realizar una simulación utilizando el software AnyLogic y utilizando los mismos parámetros de λ y μ obtenemos los siguientes resultados:

- Promedio de clientes en el sistema: 2.033 (sobre un total de 59653 muestras)
- Promedio de clientes en la cola: 2.77 (sobre un total de 59653 muestras)
- Tiempo promedio en el sistema: 0.93
- Tiempo promedio en la cola: 0.68





3.3.3 Simulación en Python

Así mismo, realizamos la misma simulación en el lenguaje de programación Python utilizando los mismos parámetros λ y μ con el objetivo de comparar los datos obtenidos. El programa finaliza al superar los 2000 clientes atendidos. Los valores obtenidos son los siguientes:

- Utilización del servidor (promedio de 10 experimentos): 0.7481570337540937
- Promedio de clientes en el sistema (promedio de 10 experimentos): 3.1479747721494924
- Promedio de clientes en la cola (promedio de 10 experimentos): 2.270693716831123
- Tiempo promedio en el sistema (promedio de 10 experimentos): 0.978905744694128
- Tiempo promedio en la cola (promedio de 10 experimentos): 0.7098027827106713
- Probabilidad de n clientes en la cola
  - 1 cliente en la cola: 0.17298
  - 2 clientes en la cola: 0.12984
  - 3 clientes en la cola: 0.100817
- Probabilidad de denegar el servicio con una cola finita de
  - 0 clientes (promedio de 10 experimentos): 0.74915
  - 2 clientes (promedio de 10 experimentos): 0.4423
  - 5 clientes (promedio de 10 experimentos): 0.151347
  - 10 clientes (promedio de 10 experimentos): 0.0023
  - 50 clientes (promedio de 10 experimentos): 0

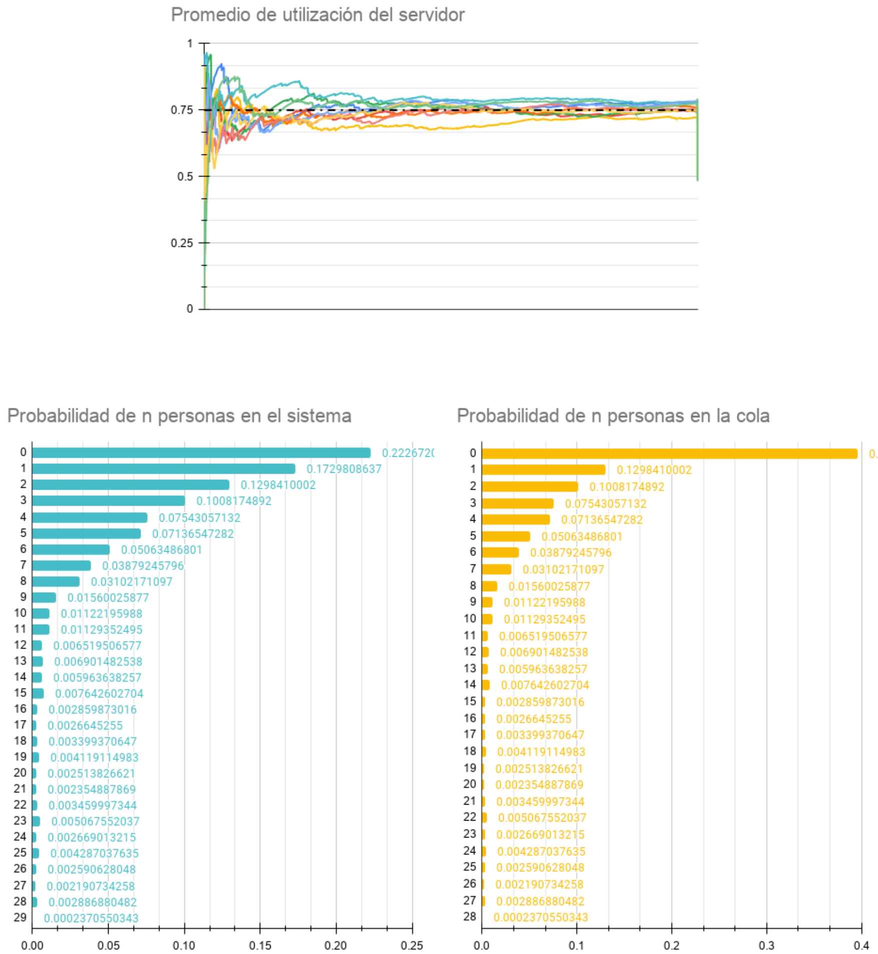


Figure 5: Resultados correspondientes a los datos del 1er experimento

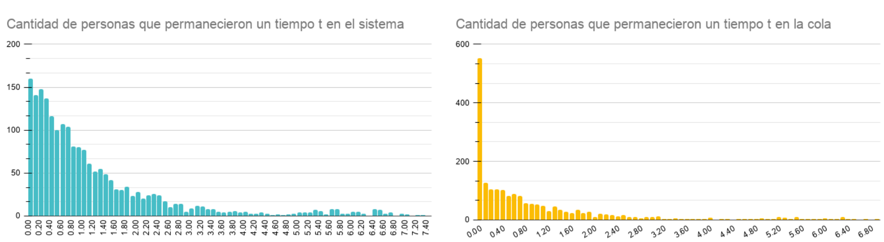


Figure 6: Resultados correspondientes a los datos del 1er experimento

3.4 Tasa de arribo del 100% respecto a la tasa de servicio

Utilizando los parámetros:

- $\lambda = 1$
- $\mu = 1$

Si bien μ debe ser mayor a λ para ser un sistema estable, decidimos investigar los posibles resultados.

3.4.1 Forma analítica

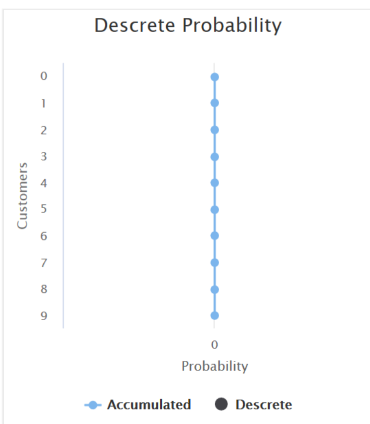
De forma analítica obtenemos:

Utilización del servidor: $\rho = \frac{\lambda}{\mu} = \frac{1}{1} = 1$

Probabilidad de n clientes en la cola: $P_n = (1 - \rho)\rho^n$

Ejemplificando, la probabilidad de:

- 1 cliente en la cola: 0
- 2 clientes en la cola: 0
- 3 clientes en la cola: 0



Promedio de clientes en el sistema: $L = \frac{\lambda}{\mu - \lambda} = \frac{1}{1-1} = \frac{1}{0} =$ tiende a infinito

Promedio de clientes en la cola: $L_q = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{1^2}{1(1-1)} = \frac{1}{0} =$ tiende a infinito

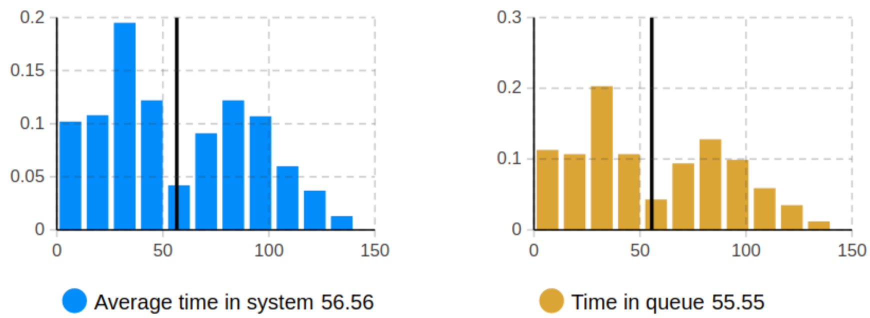
Tiempo promedio en el sistema: $W = \frac{L}{\lambda} = \frac{1}{\mu - \lambda} = \frac{1}{-1} = \frac{1}{0}$ tiende a infinito

Tiempo promedio en la cola: $W_q = \frac{\rho}{\mu - \lambda} = \frac{\frac{1}{1}}{1-1} = \frac{1}{0}$ tiende a infinito

3.4.2 Simulación en AnyLogic

Tras realizar una simulación utilizando el software AnyLogic y utilizando los mismos parámetros de λ y μ obtenemos los siguientes resultados:

- Promedio de clientes en el sistema: 56162 (sobre un total de 19862 muestras)
- Promedio de clientes en la cola: 55171 (sobre un total de 19862 muestras)
- Tiempo promedio en el sistema: 56.56
- Tiempo promedio en la cola: 55.55



3.4.3 Simulación en Python

Así mismo, realizamos la misma simulación en el lenguaje de programación Python utilizando los mismos parámetros λ y μ con el objetivo de comparar los datos obtenidos. El programa finaliza al superar los 2000 clientes atendidos. Los valores obtenidos son los siguientes:

- Utilización del servidor (promedio de 10 experimentos): 0.9779369586915887
- Promedio de clientes en el sistema (promedio de 10 experimentos): 19.0811531711
- Promedio de clientes en la cola (promedio de 10 experimentos): 18.1128774253
- Tiempo promedio en el sistema (promedio de 10 experimentos): 19.1598847979
- Tiempo promedio en la cola (promedio de 10 experimentos): 18.1668314192
- Probabilidad de n clientes en la cola
  - 1 cliente en la cola: 0.02295
  - 2 clientes en la cola: 0.02256
  - 3 clientes en la cola: 0.01985

Con respecto a la probabilidad de n clientes en la cola, en primer lugar, todas las n cantidades de clientes en cola poseen la misma probabilidad de ser halladas, es decir, son equiprobables. En segundo lugar la simulación en Python posee un límite de usuarios en el sistema, el cual, una vez superado el programa se detiene. Si la simulación se realizara de forma infinita, los valores de n clientes en cola comenzarían a tender todos a 0 como fue calculado anteriormente de forma analítica.

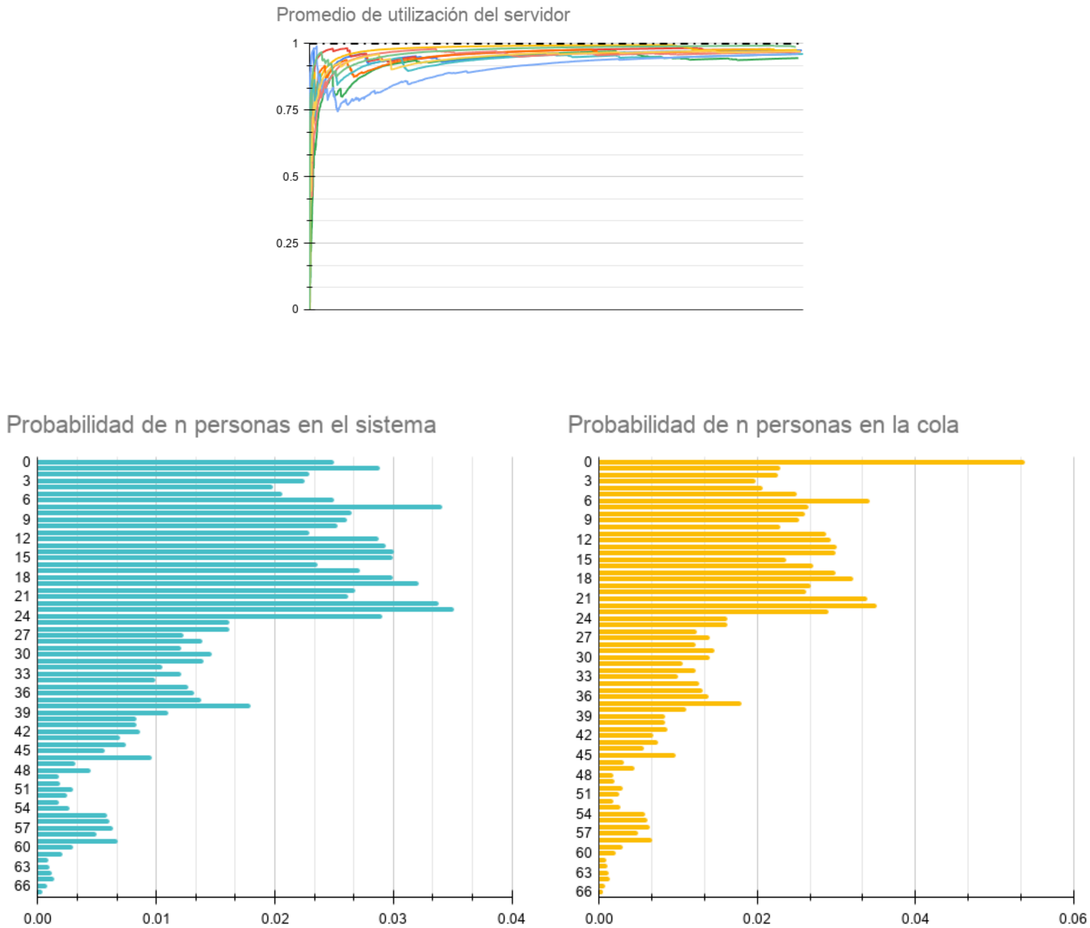


Figure 7: Resultados correspondientes a los datos del 1er experimento

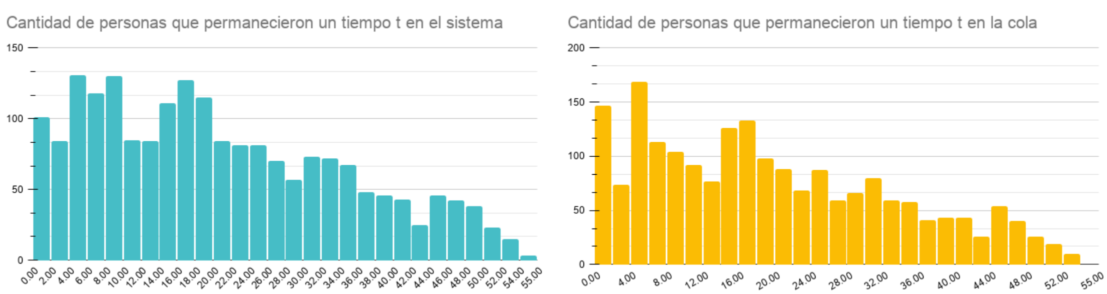


Figure 8: Resultados correspondientes a los datos del 1er experimento

3.5 Tasa de arribo del 125% respecto a la tasa de servicio

Utilizando los parámetros:

- $\lambda = 5$
- $\mu = 4$

Si bien μ debe ser mayor a λ para ser un sistema estable, decidimos investigar los posibles resultados.

3.5.1 Forma analítica

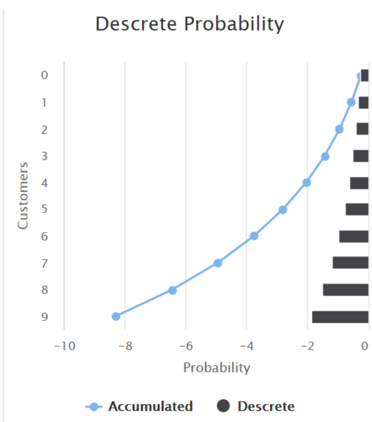
De forma analítica obtenemos:

Utilización del servidor: $\rho = \frac{\lambda}{\mu} = \frac{5}{4}$

Probabilidad de n clientes en la cola: $P_n = (1-\rho)\rho^n$

Ejemplificando, la probabilidad de:

- 1 cliente en la cola: $-\frac{5}{16}$
- 2 clientes en la cola: $-\frac{25}{64}$
- 3 clientes en la cola: $-\frac{125}{256}$



Promedio de clientes en el sistema: $L = \frac{\lambda}{\mu-\lambda} = \frac{5}{4-5} = -5$

Promedio de clientes en la cola: $L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{5^2}{4(4-5)} = -\frac{25}{4}$

Tiempo promedio en el sistema: $W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda} = \frac{1}{4-5} = -1$

Tiempo promedio en la cola: $W_q = \frac{\rho}{\mu-\lambda} = \frac{\frac{5}{4}}{4-5} = -\frac{5}{4}$

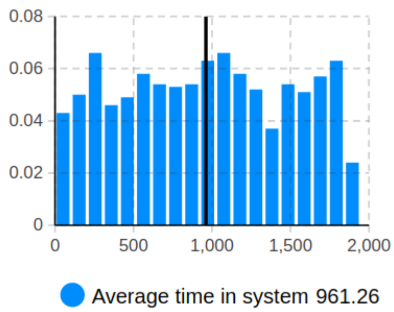
Probabilidad de denegación de servicio: $1 - \sum_{i=0}^{n}(1-\rho)\rho^i$

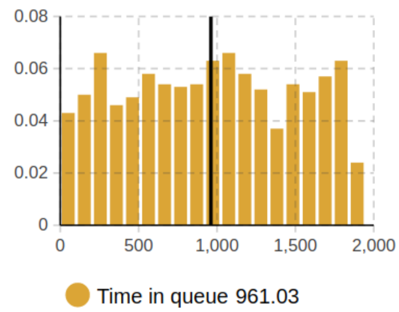
La probabilidad de denegar el servicio cuando $\lambda > \rho$ es 1. Esto se debe a que debido a que llegan mas usuarios al sistema del que el mismo puede atender, la longitud de la cola tiende a infinito mientras que la capacidad de la misma se mantiene constante.

3.5.2 Simulación en AnyLogic

Tras realizar una simulación utilizando el software AnyLogic y utilizando los mismos parámetros de λ y μ obtenemos los siguientes resultados:

- Promedio de clientes en el sistema: 4777.77 (sobre un total de 89975 muestras)
- Promedio de clientes en la cola: 4776.77 (sobre un total de 89975 muestras)
- Tiempo promedio en el sistema: 961.26
- Tiempo promedio en la cola: 961.03





3.5.3 Simulación en Python

Así mismo, realizamos la misma simulación en el lenguaje de programación Python utilizando los mismos parámetros λ y μ con el objetivo de comparar los datos obtenidos. El programa finaliza al superar los 2000 clientes atendidos. Los valores obtenidos son los siguientes:

- Utilización del servidor (promedio de 10 experimentos): 0.998002643694956
- Promedio de clientes en el sistema (promedio de 10 experimentos): 276.142469147
- Promedio de clientes en la cola (promedio de 10 experimentos): 275.144341968
- Tiempo promedio en el sistema (promedio de 10 experimentos): 55.5245586787
- Tiempo promedio en la cola (promedio de 10 experimentos): 55.2737135609
- Probabilidad de n clientes en la cola
  - 1 cliente en la cola: 0.001707666
  - 2 clientes en la cola: 0.001078986
  - 3 clientes en la cola: 0.000291259
- Probabilidad de denegar el servicio con una cola finita de
  - 0 clientes (promedio de 10 experimentos): 1
  - 2 clientes (promedio de 10 experimentos): 1
  - 5 clientes (promedio de 10 experimentos): 1
  - 10 clientes (promedio de 10 experimentos): 1
  - 50 clientes (promedio de 10 experimentos): 1

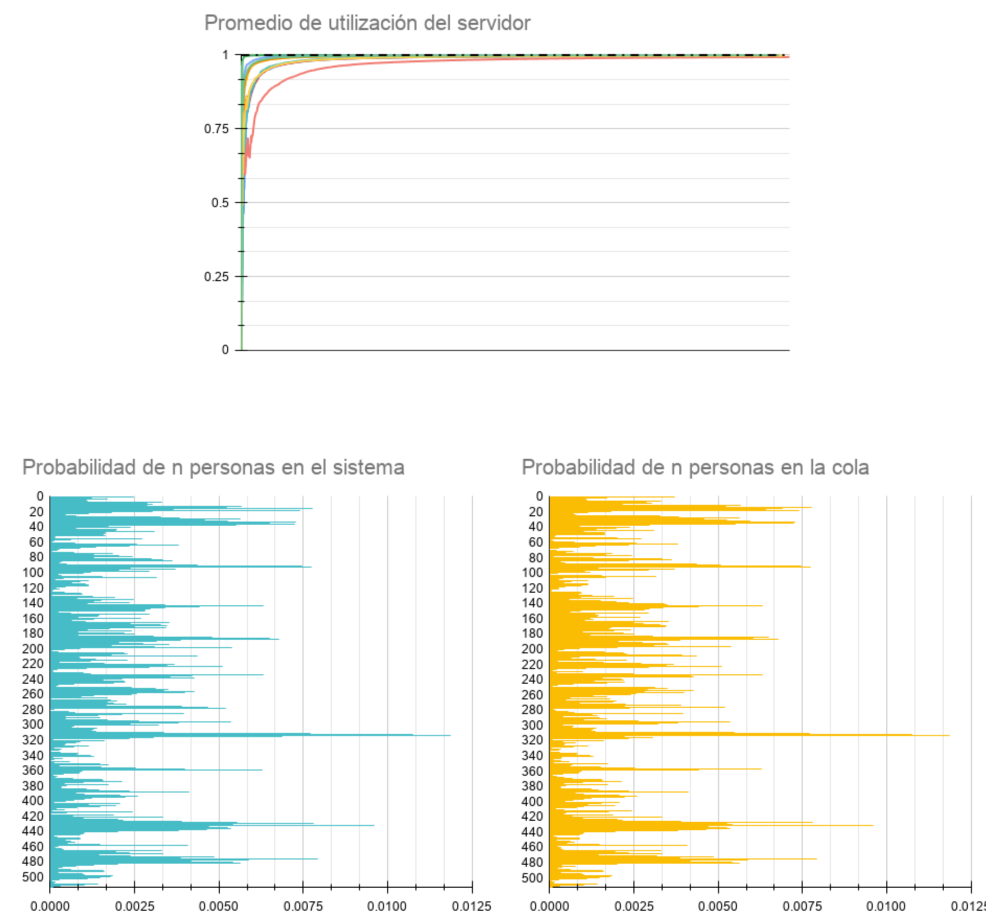


Figure 9: Resultados correspondientes a los datos del 1er experimento

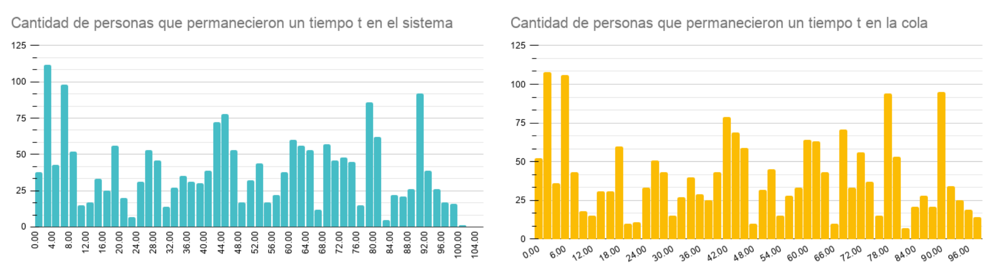


Figure 10: Resultados correspondientes a los datos del 1er experimento

4 Discusión

Al observar los resultados de los análisis realizados sobre las diferentes tasas de arribo respecto a la tasa de servicio notamos:

- Que las gráficas generadas analíticamente, con la simulación en Python y con la simulación en Anylogic tienen una composición visual casi idéntica.
- Que los valores que obtuvimos mediante las 3 metodologías fueron muy similares.
- Que al usar un μ menor igual a λ los valores de rendimiento tienden a infinito o son negativos.

5 Conclusiones

Al realizar los análisis que prometimos notamos que los valores que aproximamos son muy parecidos a los valores esperados que calculamos analíticamente. Lo cual son muy buenas noticias para nosotros que adoramos la idea de poder simular eventos de la vida real.

En este experimento los valores que poseemos fácilmente podían ser calculados, ya que, al ser un sistema tan simple las herramientas estadísticas que poseemos al alcance eran suficientes para solucionar las problemáticas que podría tener la cola y como optimizarla. Pero en muchas ocasiones la metodología analítica podría requerir mucho esfuerzo y recursos que no poseemos si es que es posible seguirla en ese entorno. Entonces concluimos finalmente que esta simulación fue una buena aproximación a la realidad y es una buena herramienta que merece ser aprendida.

References

[1] W David Kelton Averill M. Law. *Simulation Modeling Analysis*.

[2] *Discrete-Event System Simulation FOURTH EDITION*.

[3] Universidad Nacional de Mar del Plata. *Modelos de líneas de espera*.

[4] Universidad Politécnica de Valencia. *Aplicando teoría de Colas en Dirección de Operaciones*.

[5] Queueing theory.