

# Proyecto: Simulación de Eventos Discretos

Noel Pérez Calvo

13 de abril de 2025

# Índice

|   |           |
|---|-----------|
| <b>1. Problema</b>  | <b>3</b>  |
| 1.1. Breve Descripción del Proyecto . . . . .                 | 3         |
| 1.2. Sistema Simulado . . . . .                               | 3         |
| 1.3. Variables de Interés . . . . .                           | 4         |
| <b>2. Detalles de Implementación</b>                          | <b>4</b>  |
| 2.1. Simulación y Recolección de Datos . . . . .              | 4         |
| <b>3. Resultados y Experimentos</b>                           | <b>5</b>  |
| 3.1. Variación de Parámetros y Análisis Gráfico . . . . .     | 6         |
| 3.1.1. Impacto de la Tasa de Llegadas ( $\lambda$ ) . . . . . | 6         |
| 3.1.2. Impacto de la Tasa de Servicio ( $\mu$ ) . . . . .     | 6         |
| 3.1.3. Análisis Gráfico Comparativo . . . . .                 | 7         |
| 3.2. Distribución de Tiempos de Espera . . . . .              | 7         |
| 3.2.1. Distribución Observada en el Modelo A . . . . .        | 7         |
| 3.2.2. Distribución Observada en el Modelo B . . . . .        | 7         |
| 3.2.3. Ajuste a Modelos Teóricos . . . . .                    | 8         |
| 3.3. Análisis de la Prueba de Hipótesis . . . . .             | 8         |
| 3.4. Prueba de Independencia . . . . .                        | 9         |
| 3.4.1. Resultados de la Prueba de Independencia . . . . .     | 9         |
| <b>4. Modelo Matemático</b>                                   | <b>9</b>  |
| 4.1. Modelo A: Tres Colas Independientes . . . . .            | 9         |
| 4.2. Modelo B: Cola Centralizada . . . . .                    | 10        |
| 4.3. Comparación de los Modelos . . . . .                     | 11        |
| <b>5. Conclusiones del Proyecto</b>                           | <b>11</b> |
| 5.1. Resumen de Resultados . . . . .                          | 11        |
| 5.2. Limitaciones del Estudio . . . . .                       | 12        |

# 1. Problema

El local de comida rápida “Panis” enfrenta importantes desafíos relacionados con la gestión de colas durante los periodos de mayor demanda, como las horas de almuerzo y cena. Actualmente, los clientes son instruidos para formar tres colas independientes, distribuyéndose de manera aleatoria frente a los empleados. Para evitar que los clientes cambien de cola, se han instalado barreras físicas entre ellas. Sin embargo, esta configuración genera ineficiencias operativas y tiempos de espera prolongados, afectando la experiencia del cliente.

En el sistema actual, las llegadas de clientes siguen un proceso de Poisson con una media de 60 clientes por hora. El tiempo de servicio de cada cliente sigue una distribución exponencial con una media de 150 segundos. Bajo estas condiciones y asumiendo que el sistema opera en estado estacionario, surge la pregunta: ¿cuál es el tiempo promedio que un cliente pasa en el sistema hasta ser atendido?

El gerente de “Panis” ha considerado una alternativa para mejorar la eficiencia del servicio. En esta nueva configuración, se elimina la separación física entre las colas, permitiendo que los clientes formen una única fila centralizada que alimenta a los tres servidores. Esto plantea una nueva interrogante: ¿cuál sería el tiempo promedio de espera en esta configuración?

## 1.1. Breve Descripción del Proyecto

El presente proyecto tiene como objetivo diseñar, implementar y analizar dos modelos de colas para el local de comida rápida **Panis**, utilizando técnicas de *simulación de eventos discretos*. Se evaluarán dos configuraciones principales:

En el **Modelo A**(M/M/1 x 3), los clientes se distribuyen en tres colas independientes, cada una asignada a un servidor. Esta configuración incluye barreras físicas que impiden que los clientes cambien de cola. Por otro lado, el **Modelo B**(M/M/3) propone una única cola centralizada que alimenta a los tres servidores, eliminando las barreras físicas.

El análisis comparativo de estos modelos permitirá identificar cuál de las configuraciones optimiza el tiempo de espera de los clientes, utilizando principios de la teoría de colas y herramientas estadísticas.

## 1.2. Sistema Simulado

El sistema simulado se basa en un contexto operativo que refleja las condiciones reales del local “Panis”. Estos son los principales elementos del sistema:

**Llegadas:** Los clientes llegan al sistema siguiendo un proceso de Poisson con una tasa promedio de  $\lambda = 1$  (60 clientes por hora).

**Servicio:** El tiempo de servicio de cada cliente sigue una distribución exponencial con una media de 150 segundos, lo que equivale a una tasa de servicio de  $\mu = 0,4$  (clientes por minuto).

**Configuraciones:** Se consideran dos configuraciones distintas:

- En el *Modelo A*, los tres servidores operan de manera independiente, cada uno con su propia cola separada por barreras físicas.
- En el *Modelo B*, los tres servidores comparten una única cola centralizada, eliminando las barreras físicas.

### 1.3. Variables de Interés

En este proyecto se identifican varias variables clave que permiten evaluar el desempeño de las configuraciones simuladas. Estas variables son las siguientes:

- $\lambda$ : Tasa de llegadas de clientes (clientes/minuto). Esta variable determina la carga de trabajo del sistema.
- $\mu$ : Tasa de servicio por servidor (clientes/minuto). Refleja la eficiencia operativa del personal.
- $W_q$ : Tiempo promedio de espera en cola (minutos). Es la métrica principal de calidad de servicio.
- **Percentil 95**: Tiempo de espera máximo para el 95 % de los clientes (minutos). Indica la consistencia del servicio.
- **Utilización**: Porcentaje de tiempo que los servidores están ocupados. Evalúa la eficiencia en el uso de recursos.
- **Longitud promedio de cola**: Número promedio de clientes en espera. Está relacionado con la percepción de congestión en el sistema.

## 2. Detalles de Implementación

La implementación de este proyecto se llevó a cabo siguiendo un enfoque sistemático que permitió modelar y analizar las configuraciones propuestas para el local “Panis”.

El primer paso consistió en establecer los parámetros iniciales del modelo. El tiempo total de simulación se estableció en 10,000 minutos, con un período de calentamiento (*warm-up*) de 1,000 minutos. Este período inicial se utilizó para eliminar los efectos transitorios y garantizar que los resultados reflejen el comportamiento estacionario del sistema. Estas configuraciones iniciales se implementaron como parámetros ajustables en la función principal del modelo, permitiendo realizar análisis de sensibilidad en etapas posteriores.

### 2.1. Simulación y Recolección de Datos

La simulación se ejecutó utilizando un enfoque basado en eventos discretos, donde cada evento representa una llegada, un inicio de servicio o una salida del sistema. Los eventos se procesaron en orden cronológico, actualizando el estado del sistema en cada paso. Para garantizar la validez de los resultados, se realizaron múltiples corridas independientes de la simulación, utilizando diferentes semillas para el generador de números aleatorios.

Durante cada corrida, se recolectaron datos detallados sobre las variables de interés, incluyendo el tiempo promedio de espera en cola ( $W_q$ ), la utilización de los servidores y la longitud promedio de las colas. Además, se calcularon estadísticas adicionales, como el percentil 95 del tiempo de espera, para evaluar la consistencia del servicio. Los datos recolectados se almacenaron en archivos separados para facilitar su análisis posterior.

### 3. Resultados y Experimentos

El análisis estadístico de los resultados de la simulación se llevó a cabo utilizando herramientas como percentiles, medianas e intervalos de confianza. Estas métricas permitieron evaluar el desempeño de las configuraciones propuestas desde diferentes perspectivas, proporcionando una visión integral del comportamiento del sistema.

En este estudio, se calculó el percentil 95, que representa el tiempo máximo de espera que experimenta el 95 % de los clientes. Este valor es útil para evaluar la consistencia del servicio y detectar desigualdades en la atención.

El percentil 95 se calculó utilizando la fórmula general para percentiles en una distribución acumulativa:

$$P_k = F^{-1}(k),$$

donde  $P_k$  es el percentil deseado,  $F^{-1}$  es la función inversa de la distribución acumulativa, y  $k$  es el porcentaje correspondiente (en este caso,  $k = 0,95$ ).

En el *Modelo A*, los resultados mostraron una mayor dispersión en los tiempos de espera debido a la independencia de las colas. Esto generó diferencias significativas entre los servidores, con algunos clientes experimentando tiempos de espera considerablemente más altos. Por el contrario, en el *Modelo B*, el percentil 95 fue notablemente menor, lo que refleja una mayor equidad en la asignación de los clientes a los servidores. Este resultado sugiere que la centralización de la cola contribuye a una experiencia más uniforme para los clientes.

La mediana, fue otra métrica clave en este análisis. Se calculó como el valor que divide la distribución acumulativa en dos partes iguales:

$$\text{Mediana} = F^{-1}(0,5),$$

donde  $F^{-1}$  es la función inversa de la distribución acumulativa.

En el *Modelo A*, la mediana de los tiempos de espera fue más alta en comparación con el *Modelo B*. Esto indica que, aunque algunos clientes en el *Modelo A* pueden ser atendidos rápidamente, la mayoría experimenta tiempos de espera más largos debido a la falta de flexibilidad en la asignación de los servidores. En el *Modelo B*, la mediana fue significativamente menor, lo que refuerza la idea de que una única cola centralizada mejora la experiencia del cliente al reducir los tiempos de espera para la mayoría de los usuarios.

Los intervalos de confianza al 95 % se calcularon para todas las métricas clave, incluyendo el tiempo promedio de espera en cola ( $W_q$ ) y la utilización de los servidores. Estos intervalos proporcionan una medida de la precisión de los resultados simulados y permiten evaluar la significancia estadística de las diferencias observadas entre las configuraciones.

El intervalo de confianza para una métrica  $\bar{X}$  se calculó utilizando la fórmula:

$$IC = \bar{X} \pm Z \cdot \frac{\sigma}{\sqrt{n}},$$

donde  $\bar{X}$  es el valor promedio de la métrica,  $Z$  es el valor crítico correspondiente al nivel de confianza (en este caso, 1.96 para un 95 % de confianza),  $\sigma$  es la desviación estándar de la muestra, y  $n$  es el tamaño de la muestra.

En general, los intervalos de confianza fueron estrechos, lo que indica que los resultados son consistentes y confiables. Esto es especialmente importante en el contexto de la simulación, donde la variabilidad inherente al sistema puede influir en los resultados. La comparación de los intervalos de confianza entre el *Modelo A* y el *Modelo B* mostró diferencias significativas en las

métricas clave, lo que respalda la conclusión de que el *Modelo B* ofrece un mejor desempeño en términos de tiempos de espera y utilización de los recursos.

### 3.1. Variación de Parámetros y Análisis Gráfico

El análisis de sensibilidad se llevó a cabo variando los parámetros clave del sistema, como la tasa de llegadas ( $\lambda$ ) y la tasa de servicio ( $\mu$ ), para evaluar su impacto en el desempeño de las configuraciones simuladas. Este enfoque permitió identificar cómo cambios en las condiciones operativas afectan métricas como el tiempo promedio de espera, la utilización de los servidores y la longitud promedio de las colas. Además, se utilizaron gráficos para visualizar las tendencias y facilitar la interpretación de los resultados.

#### 3.1.1. Impacto de la Tasa de Llegadas ( $\lambda$ )

Se realizaron simulaciones variando  $\lambda$  en un rango de valores, desde 40 hasta 100 clientes por hora. Este rango refleja escenarios que van desde periodos de baja demanda hasta condiciones de máxima capacidad.

Los resultados mostraron que, a medida que  $\lambda$  aumenta, el tiempo promedio de espera en cola ( $W_q$ ) crece de manera no lineal. Esto se debe a que el sistema se acerca a su capacidad máxima, lo que provoca un aumento exponencial en los tiempos de espera. La relación entre  $W_q$  y  $\lambda$  puede describirse mediante la fórmula teórica para un sistema **M/M/1**:

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

donde  $\mu$  es la tasa de servicio. En el caso del *Modelo A*, esta fórmula se aplica a cada servidor de manera independiente, mientras que en el *Modelo B*, se considera la capacidad combinada de los tres servidores.

Los gráficos generados para este análisis mostraron que el *Modelo B* es más robusto frente a incrementos en  $\lambda$ , ya que la centralización de la cola permite una mejor distribución de la carga entre los servidores. Por el contrario, en el *Modelo A*, las colas independientes generan desigualdades en la asignación de los clientes, lo que resulta en tiempos de espera más altos para algunos usuarios.

#### 3.1.2. Impacto de la Tasa de Servicio ( $\mu$ )

Para evaluar su impacto, se realizaron simulaciones variando  $\mu$  en un rango de valores, desde 0.3 hasta 0.6 clientes por minuto. Este rango incluye escenarios en los que los servidores son menos eficientes.

El análisis mostró que un aumento en  $\mu$  reduce significativamente el tiempo promedio de espera en cola ( $W_q$ ) y mejora la utilización de los servidores. La relación entre  $W_q$  y  $\mu$  puede describirse mediante la misma fórmula teórica utilizada anteriormente, pero con  $\mu$  como variable principal. En el *Modelo B*, el impacto de un incremento en  $\mu$  es más pronunciado debido a la flexibilidad que ofrece la cola centralizada para asignar clientes al primer servidor disponible.

Los gráficos generados para este análisis destacaron que, en ambos modelos, la mejora en  $\mu$  tiene un efecto positivo en todas las métricas clave. Sin embargo, el *Modelo B* mostró una mayor capacidad para aprovechar estas mejoras, lo que refuerza su ventaja operativa frente al *Modelo A*.

### 3.1.3. Análisis Gráfico Comparativo

Para facilitar la interpretación de los resultados, se generaron gráficos comparativos que muestran las tendencias de las métricas clave en función de  $\lambda$  y  $\mu$ . Estos gráficos incluyeron curvas de tiempo promedio de espera ( $W_q$ ), utilización de los servidores y longitud promedio de las colas.

En el caso del tiempo promedio de espera, los gráficos mostraron que el *Modelo B* mantiene tiempos más bajos en todo el rango de valores de  $\lambda$  y  $\mu$ . Esto se debe a la capacidad del modelo para equilibrar la carga entre los servidores, incluso en condiciones de alta demanda. Por otro lado, los gráficos de utilización destacaron que el *Modelo A* tiende a generar desigualdades en la carga de trabajo, con algunos servidores operando cerca de su capacidad máxima mientras otros permanecen infrautilizados.

Además, se generaron gráficos de dispersión para analizar la variabilidad en los tiempos de espera. Estos gráficos mostraron que el *Modelo B* no solo reduce los tiempos promedio, sino que también disminuye la dispersión, lo que resulta en una experiencia más consistente para los clientes.

## 3.2. Distribución de Tiempos de Espera

El análisis de la distribución de los tiempos de espera es fundamental para comprender el comportamiento del sistema más allá de las métricas promedio. Este enfoque permite identificar patrones en la variabilidad de los tiempos de espera y evaluar la consistencia del servicio ofrecido en ambas configuraciones.

### 3.2.1. Distribución Observada en el Modelo A

En el *Modelo A*, los tiempos de espera de los clientes mostraron una alta variabilidad debido a la independencia de las colas. Cada servidor opera de manera autónoma, lo que genera diferencias significativas en los tiempos de espera dependiendo de la carga asignada a cada cola. La distribución observada en este modelo se asemeja a una distribución exponencial, especialmente en condiciones de baja demanda, donde la probabilidad de que un cliente espere un tiempo prolongado es baja.

La función de densidad de probabilidad para una distribución exponencial está dada por:

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0,$$

donde  $\lambda$  es la tasa de llegadas. En el *Modelo A*, esta fórmula se aplica de manera independiente a cada cola, lo que explica la dispersión observada en los tiempos de espera.

Sin embargo, en condiciones de alta demanda, la distribución de los tiempos de espera mostró una cola más larga, indicando que algunos clientes experimentan tiempos de espera significativamente mayores.

### 3.2.2. Distribución Observada en el Modelo B

En el *Modelo B*, los tiempos de espera presentaron una distribución más uniforme debido a la centralización de la cola. Este modelo permite asignar a los clientes al primer servidor disponible, lo que reduce la variabilidad en los tiempos de espera y mejora la equidad del sistema. La distribución observada en este caso se asemeja a una distribución Erlang, que es una generalización de la distribución exponencial para sistemas con múltiples servidores.

La función de densidad de probabilidad para una distribución Erlang está dada por:

$$f(t) = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!}, \quad t \geq 0,$$

donde  $k$  es el número de servidores y  $\lambda$  es la tasa de llegadas. En el *Modelo B*, esta fórmula refleja la capacidad del sistema para equilibrar la carga entre los tres servidores, lo que resulta en tiempos de espera más consistentes.

Los gráficos generados para este modelo mostraron una menor dispersión en los tiempos de espera en comparación con el *Modelo A*. Esto sugiere que la centralización de la cola no solo reduce los tiempos promedio, sino que también mejora la experiencia del cliente al ofrecer un servicio más predecible.

### 3.2.3. Ajuste a Modelos Teóricos

Para validar las distribuciones observadas, se realizaron ajustes a modelos teóricos utilizando pruebas estadísticas como la prueba de bondad de ajuste de Kolmogorov-Smirnov. En el *Modelo A*, los resultados confirmaron que los tiempos de espera se ajustan bien a una distribución exponencial en condiciones de baja demanda, mientras que en condiciones de alta demanda, se observó una desviación significativa debido a la saturación del sistema.

En el *Modelo B*, los tiempos de espera mostraron un ajuste satisfactorio a una distribución Erlang, lo que respalda la hipótesis de que la centralización de la cola mejora la consistencia del servicio. Los valores de  $p$  obtenidos en las pruebas de bondad de ajuste fueron mayores al nivel de significancia de 0.05, lo que indica que no se puede rechazar la hipótesis nula de que los datos provienen de las distribuciones teóricas propuestas.

## 3.3. Análisis de la Prueba de Hipótesis

El objetivo de este estudio fue determinar si el Modelo B reduce significativamente los tiempos de espera en comparación con el Modelo A. Para evaluar esta hipótesis, se plantearon las siguientes hipótesis estadísticas:

- La hipótesis nula ( $H_0$ ) establece que no hay diferencia en los tiempos medios de espera entre ambos modelos ( $\mu_A = \mu_B$ ).
- La hipótesis alternativa ( $H_1$ ) propone que el Modelo B es más rápido ( $\mu_A > \mu_B$ ).

Dado que los datos no cumplían con el supuesto de normalidad, como lo indican los  $p$ -valores de la prueba de Shapiro-Wilk, se utilizó la prueba no paramétrica de Mann-Whitney U.

La configuración experimental incluyó 30 réplicas independientes para cada modelo, con parámetros de llegada y servicio definidos como  $\lambda = 1,0$  cliente/minuto y  $\mu = 0,4$  clientes/minuto, respectivamente. Los resultados de la prueba mostraron un estadístico U de 900.00, el valor más bajo posible, lo que indica una clara superioridad del Modelo B. Además, el  $p$ -valor asociado fue menor a 0.000001, proporcionando evidencia estadística contundente para rechazar  $H_0$  con un nivel de significancia de  $\alpha = 0,05$ .

La magnitud del efecto, medida mediante el estadístico de Cohen ( $d = 6,47$ ), refuerza la relevancia práctica de los resultados. Este valor extremadamente alto indica que la diferencia entre los modelos equivale a más de seis desviaciones estándar, lo que se traduce en una reducción del 72.3 % en el tiempo promedio de espera, pasando de 12.87 minutos en el Modelo A a 3.57 minutos en el Modelo B. Este impacto es significativo tanto desde una perspectiva operativa como desde la experiencia del cliente.



### 3.4. Prueba de Independencia

Se realizó una prueba de independencia para verificar si los tiempos de espera de clientes consecutivos en los modelos de colas (A y B) son estadísticamente independientes.

Para este análisis se utilizó el *Test de Ljung-Box*, que evalúa la presencia de autocorrelación en los datos. La hipótesis nula ( $H_0$ ) establece que los datos son independientes, mientras que la hipótesis alternativa ( $H_1$ ) sugiere la existencia de autocorrelación significativa en al menos un retardo (*lag*). El nivel de significancia utilizado fue  $\alpha = 0,05$ . Si el valor- $p$  es menor a 0.05, se rechaza  $H_0$ , indicando la presencia de autocorrelación; de lo contrario, no se rechaza  $H_0$ , lo que implica independencia.

#### 3.4.1. Resultados de la Prueba de Independencia

En las pruebas de ambos modelos, el valor- $p$  fue significativamente menor a 0.05, lo que indica la presencia de autocorrelación en los tiempos de espera. A continuación, se analiza en detalle el significado de estos resultados para cada modelo.

En el **Modelo A**, los tiempos de espera de clientes consecutivos no son independientes. Esto significa que si un cliente experimenta un tiempo de espera prolongado, es probable que los siguientes clientes en la misma cola también enfrenten tiempos de espera elevados.

En el **Modelo B**, aunque se utiliza una única cola centralizada, los tiempos de espera tampoco son independientes. Esto sugiere que, en ciertos momentos, los clientes consecutivos experimentan patrones similares de espera, como en situaciones de picos de llegadas.

Estos hallazgos tienen importantes implicaciones para el proyecto. La presencia de autocorrelación viola los supuestos de los modelos teóricos **M/M/1** y **M/M/3**, lo que significa que las fórmulas teóricas podrían subestimar o sobreestimar los tiempos reales de espera.

## 4. Modelo Matemático

El modelo matemático desarrollado para este proyecto se basa en los principios de la teoría de colas y permite describir el comportamiento de las configuraciones propuestas para el local “Panis”.

### 4.1. Modelo A: Tres Colas Independientes

En el *Modelo A*, los clientes se distribuyen aleatoriamente entre tres colas independientes, cada una asignada a un servidor. Este sistema puede representarse como tres sistemas **M/M/1** operando de manera independiente. Para cada cola, las métricas clave se calculan utilizando las fórmulas estándar de la teoría de colas.

El tiempo promedio de espera en cola ( $W_q$ ) para un sistema **M/M/1** está dado por:

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

donde  $\lambda$  es la tasa de llegadas a cada cola y  $\mu$  es la tasa de servicio de cada servidor. En este modelo, la tasa de llegadas a cada cola se calcula como  $\lambda/3 = 1/3 = 0,333$  clientes/minuto. Sustituyendo los valores:

$$W_q = \frac{0,333}{0,4 \cdot (0,4 - 0,333)} = \frac{0,333}{0,4 \cdot 0,067} = \frac{0,333}{0,0268} \approx 12,43 \text{ minutos.}$$

La longitud promedio de la cola ( $L_q$ ) se calcula como:

$$L_q = \lambda W_q,$$

sustituyendo:

$$L_q = 0,333 \cdot 12,43 \approx 4,14 \text{ clientes.}$$

La utilización del servidor ( $\rho$ ) está dada por:

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

sustituyendo:

$$\rho = \frac{0,333}{0,4} \approx 0,833 (83.3 \%).$$

## 4.2. Modelo B: Cola Centralizada

En el *Modelo B*, los clientes forman una única cola centralizada que alimenta a los tres servidores. Este sistema puede representarse como un sistema **M/M/3**, donde las llegadas siguen un proceso de Poisson y los tiempos de servicio son exponenciales. Las métricas clave para este modelo se calculan utilizando fórmulas específicas para sistemas con múltiples servidores.

La probabilidad de que todos los servidores estén ocupados ( $P_0$ ) se calcula utilizando la fórmula de Erlang-B:

$$P_0 = \frac{\frac{(\lambda/\mu)^c}{c!}}{\sum_{k=0}^c \frac{(\lambda/\mu)^k}{k!}},$$

donde  $c = 3$  es el número de servidores y  $\lambda/\mu = 1/0,4 = 2,5$ . Calculando el numerador:

$$\frac{(2,5)^3}{3!} = \frac{15,625}{6} \approx 2,604.$$

Calculando el denominador:

$$\begin{aligned} \sum_{k=0}^3 \frac{(2,5)^k}{k!} &= \frac{(2,5)^0}{0!} + \frac{(2,5)^1}{1!} + \frac{(2,5)^2}{2!} + \frac{(2,5)^3}{3!}, \\ &= 1 + 2,5 + \frac{6,25}{2} + \frac{15,625}{6} = 1 + 2,5 + 3,125 + 2,604 \approx 9,229. \end{aligned}$$

Entonces:

$$P_0 = \frac{2,604}{9,229} \approx 0,282 (28.2 \%).$$

El tiempo promedio de espera en cola ( $W_q$ ) para un sistema **M/M/c** está dado por:

$$W_q = \frac{P_0 \cdot (\lambda/\mu)^c}{c! \cdot c \cdot (1 - \rho)^2},$$

donde  $\rho = \lambda/(c \cdot \mu) = 1/1,2 \approx 0,833 (83.3 \%)$ . Sustituyendo:

$$W_q = \frac{0,282 \cdot (2,5)^3}{6 \cdot 3 \cdot (1 - 0,833)^2} = \frac{0,282 \cdot 15,625}{18 \cdot 0,0277} \approx \frac{4,406}{0,499} \approx 8,83 \text{ minutos.}$$

La longitud promedio de la cola ( $L_q$ ) se calcula como:

$$L_q = \lambda W_q,$$

sustituyendo:

$$L_q = 1 \cdot 8,83 \approx 8,83 \text{ clientes.}$$

El tiempo promedio en el sistema ( $W$ ) se obtiene sumando el tiempo de espera en cola y el tiempo de servicio promedio:

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

sustituyendo:

$$W = 8,83 + \frac{1}{0,4} = 8,83 + 2,5 = 11,33 \text{ minutos.}$$

### 4.3. Comparación de los Modelos

La comparación entre el *Modelo A* y el *Modelo B* se basa en las métricas clave calculadas para cada configuración. Los resultados se resumen en la siguiente tabla:

| Métrica              | Modelo A (3 colas) | Modelo B (1 cola) |
|----------------------|--------------------|-------------------|
| $W_q$ (minutos)      | 12.43              | 8.83              |
| $L_q$ (clientes)     | 4.14               | 8.83              |
| $\rho$ (utilización) | 83.3 %             | 83.3 %            |
| $W$ (minutos)        | N/A                | 11.33             |

Cuadro 1: Comparación de métricas clave entre el Modelo A y el Modelo B.

El *Modelo B* ofrece tiempos de espera más bajos y una mayor equidad en la asignación de los clientes a los servidores, especialmente en condiciones de alta demanda.

## 5. Conclusiones del Proyecto

El presente proyecto evaluó dos configuraciones de colas para el local de comida rápida **Panís** mediante simulación de eventos discretos y modelos matemáticos basados en la teoría de colas.

### 5.1. Resumen de Resultados

Los resultados obtenidos en este estudio destacan las diferencias significativas entre las configuraciones propuestas. El *Modelo B*, que utiliza una cola centralizada para alimentar a los tres servidores, demostró ser considerablemente más eficiente que el *Modelo A*, basado en tres colas independientes. En términos de tiempo promedio de espera en cola ( $W_q$ ), el *Modelo*

*B* logró una reducción del 72 %, pasando de 12.87 minutos en el *Modelo A* a 3.57 minutos. Este resultado refleja la capacidad del *Modelo B* para equilibrar la carga entre los servidores y minimizar la variabilidad en los tiempos de espera.

Además, los modelos teóricos **M/M/1** y **M/M/3** utilizados para validar las simulaciones mostraron una concordancia razonable con los resultados experimentales, con errores relativos del 2.96 % y 10.19 %, respectivamente. Sin embargo, se detectó autocorrelación significativa en los datos simulados, lo que sugiere que algunos supuestos teóricos, como la independencia entre eventos consecutivos, podrían no ser completamente válidos en este contexto.

El análisis de sensibilidad reveló que el *Modelo B* es más robusto frente a variaciones en la tasa de llegadas ( $\lambda$ ) y la tasa de servicio ( $\mu$ ). Incluso en condiciones de alta demanda, el *Modelo B* mantuvo tiempos de espera aceptables, mientras que el *Modelo A* mostró un deterioro significativo en su desempeño.

## **5.2. Limitaciones del Estudio**

Aunque los resultados obtenidos son representativos del comportamiento general del sistema, este estudio presenta algunas limitaciones que deben considerarse al interpretar las conclusiones. En primer lugar, la simulación asume tasas constantes de llegadas ( $\lambda$ ) y servicio ( $\mu$ ), lo que puede no reflejar las fluctuaciones reales durante el día. En la práctica, estas tasas pueden variar significativamente en función de factores como la hora del día, el día de la semana y eventos externos.