SIMULACIÓN DEL MODELO DE COLA SIMPLE

RAFAEL MARTÍN BIGIO
SANTIAGO ANDRÉS COFFEY
ANDRÉS SANTIAGO GREGOIRE

Resumen. En este artículo se simula un sistema de cola simple M/M/1 y se estiman algunos de sus parámetros, tales como el tiempo medio de clientes en cola y la longitud máxima, entre otros. Asimismo mediante una simulación de Montecarlo, se compara un modelo de cola simple de dos canales con otro de un único canal doblemente rápido.

1. Introducción

En un sistema de cola simple de tipo M/M/1/∞/FIFO, tanto el intervalo de tiempo entre arribos de clientes como el tiempo de servicio son variables aleatorias con distribución exponencial. Asimismo, se cuenta con un único servidor y se considera que la cola de clientes tiene capacidad infinita y sigue una disciplina *first in-first out*.

Si $Q(t)$ es la longitud de la cola en el instante t y $B(t)$ es el estado de ocupación del servidor en dicho instante, entonces el estado del sistema es $x = (Q, B)$. En consecuencia, el espacio de estados es $S = \{(n, m)/n = 0, 1, 2, \dots,\ m = 0, 1\}$. Por último, el conjunto de eventos es $\mathcal{E} = \{A, D\}$, donde el mnemónico A denota el arribo de un cliente, y D la salida del sistema de otro.

En este artículo se estiman algunos parámetros del sistema de cola simple, considerando que el tiempo medio entre llegadas sucesivas de clientes es de $1/\lambda = 1$ minuto y que el tiempo de atención de la facilidad es de $1/\mu = 1$ minuto.

En la sección 2 se simula el sistema para computar algunos estimadores característicos. En la sección 3 se compara el mismo sistema con el doble de tiempo de servicio con otro de dos canales con el mismo tiempo de servicio que el sistema original. Por último, en la sección 4 se exponen las conclusiones.

2. Simulación del sistema

Las simulaciones son realizadas en *GNU Octave 2.1.71* utilizando una modificación del algoritmo estándar de cola simple. En todos los casos, se fija como semilla para la función `rand` el valor 3000.

En la figura 1 se grafica la longitud de la cola entre $t = 0$ y el tiempo en el que llega el décimo cliente. Se puede observar que si bien la simulación tiene una corta duración, la máxima cantidad de clientes en cola resulta 4.

Se define $L(t)$ como la cantidad total de clientes en el sistema en el instante t. Considerando que no hay demoras para trasladar un cliente de la cola al servidor, los clientes que están en el sistema se encuentran encolados o siendo atendidos. En consecuencia, resulta que $L(t) = Q(t) + 1$ siempre que haya al menos un cliente en el sistema.

Se define la cantidad media de clientes en el sistema según el número de arribos como

$$\overline{L}(n) = \frac{\int_0^{T(n)} L(t)dt}{T(n)} \tag{1}$$

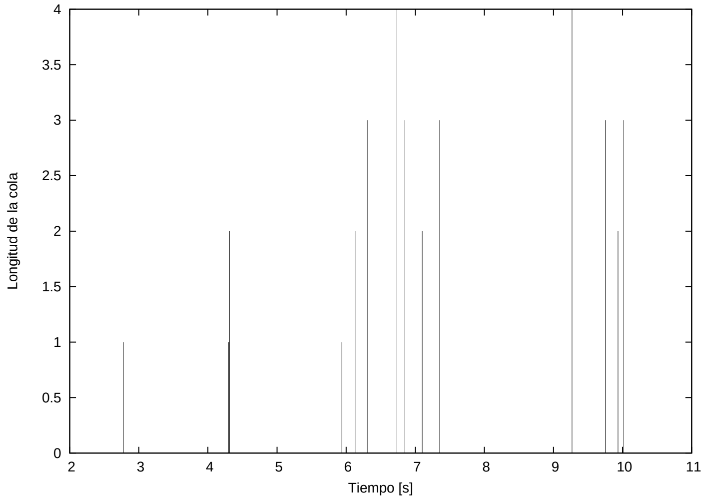

Figura 1. Longitud de la cola en función del tiempo.

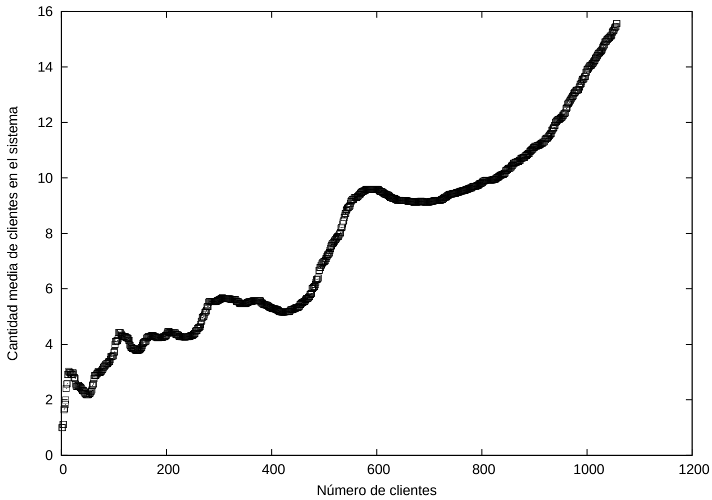

Figura 2. Cantidad media de clientes en el sistema en función de la cantidad de arribos.

y la cantidad media de clientes en cola según la cantidad de llegadas como

$$\overline{Q}(n) = \frac{\int_0^{T(n)} Q(t)dt}{T(n)} \tag{2}$$

En la figura 2 se grafica $\overline{L}(n)$ para una realización con 1000 clientes. El gráfico de $\overline{Q}(n)$ resulta lo mismo desplazado en una unidad según lo señalado anteriormente. Se observa que conforme aumenta el número de clientes de la simulación, la cantidad de los mismos en el sistema (y en cola) aumenta.

Asimismo, se define la ocupación media del servidor según la cantidad de llegadas como

$$\overline{B}(n) = \frac{\int_0^{T(n)} B(t)dt}{T(n)} \tag{3}$$

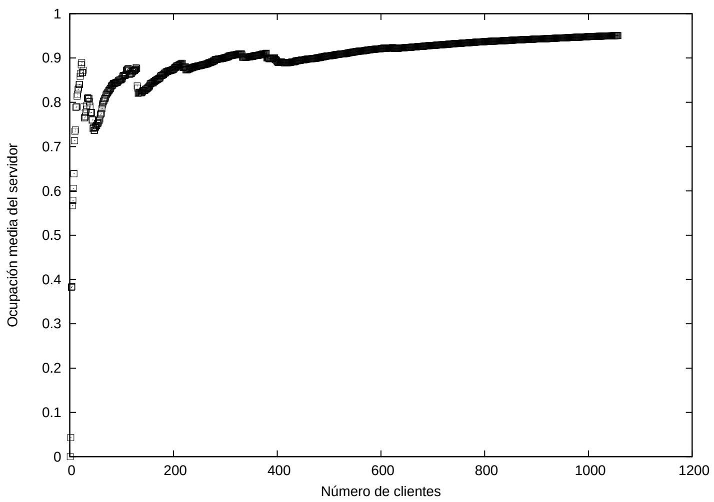

FIGURA 3. Ocupación media del servidor en función de la cantidad de arribos.

En la figura 3 se grafica $\overline{B}(n)$. Se puede observar que la ocupación media del servidor tiende a 1 conforme aumenta el número de clientes. Este resultado es razonable puesto que, si el número de clientes en cola crece indefinidamente, el servidor siempre tiene uno nuevo para atender. Esto también se puede deducir a partir de los parámetros del sistema, dado que la *intensidad de tráfico* es $\rho = \frac{\lambda}{\mu} = 1$. Es decir que el sistema es inestable, y en consecuencia, el número de clientes en cola aumenta indefinidamente.

Mediante una simulación de Montecarlo se computan varios parámetros poblacionales. Para ello se obtienen realizaciones de sus estimadores a partir de simulaciones del modelo con 100 clientes. Como criterio de detención se fija que la discrepancia de la norma de las varianzas de dichas variables aleatorias sea menor que una cota arbitrariamente prefijada en 0,005. La fórmula recursiva para computar cada varianza es

$$\sigma_t^2 = \frac{t-1}{t}\sigma_{t-1}^2 + \frac{1}{t-1}(x_t - \mu_t)^2 \tag{4}$$

donde σ_i^2, x_i y μ_i son la varianza, el dato y la media de la i-ésima iteración respectivamente. Con este criterio de detención se obtienen 92 iteraciones. En la tabla 1 se presentan los parámetros estimados, computados en un intervalos de confianza con un nivel de significación $\alpha = 0{,}05$.

En la figura 4 se grafica la varianza de la norma de los estimadores. Se puede observar que su evolución es inversamente proporcional al número de realizaciones.

3. Comparación de sistemas

3.1. Simulación de cola simple con distinto tiempo de servicio. Se desea estudiar la influencia del parámetro μ en el sistema. Para ello, se realizan nuevamente las mismas simulaciones que en la sección anterior estableciendo $\mu = 2$, es decir, fijando el tiempo medio de servicio a la mitad que en el caso anterior.

En las figura 5 se grafica la cantidad media de clientes en el sistema en función de los arribos. Se puede observar que con estos parámetros la curva queda acotada (no crece indefinidamente). En la figura 6 se grafica la ocupación media del servidor en función de la cantidad de arribos. A diferencia del sistema estudiado anteriormente,

Parámetro	Valor estimado
Tiempo medio de un cliente en el sistema [min]	$7{,}017006 \pm 0{,}147429$
Número máximo de clientes en cola	$16{,}336957 \pm 0{,}263592$
Máximo retardo en cola [min]	$14{,}839343 \pm 0{,}261827$
Máximo retardo en el sistema [min]	$15{,}875468 \pm 0{,}280937$
Proporción de clientes con retardo en cola superior a 1 minuto	$0{,}766630 \pm 0{,}00251$
Proporción de clientes con retardo en el sistema superior a 1 minuto	$0{,}879337 \pm 0{,}001863$

TABLA 1. Estimaciones de parámetros del sistema de cola simple (M/M/1 con $\mu = 1$) con sus intervalos de confianza.

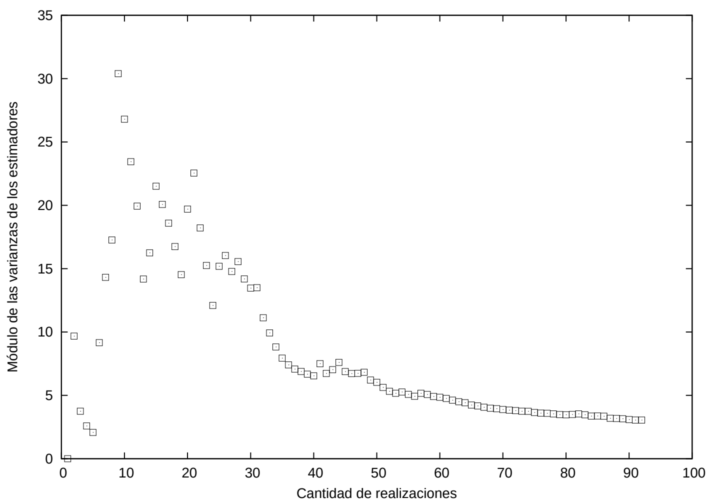

FIGURA 4. Varianza de la norma de los estimadores.

en este caso el servidor no se encuentra casi siempre ocupado, ya que la cola no crece indefinidamente. Estos resultados se condicen con la estabilidad del sistema, dado que la *intensidad de tráfico* es $\rho = 0{,}5 < 1$ (es estable).

En la tabla 2 se presentan las estimaciones de los mismos parámetros poblacionales que en la sección anterior previa, con sus respectivos intervalos de confianza. En este caso, fijando nuevamente la cota del criterio de detención en 0,005, basta con 31 iteraciones.

El tiempo medio de un cliente en cola W_q y en el sistema W se pueden computar analíticamente utilizando las siguientes expresiones:

$$W_q = \frac{\rho}{\mu - \lambda} \tag{5}$$

$$W = \frac{1}{\mu - \lambda} \tag{6}$$

donde λ y μ son las tasas de arribo de clientes al sistema y la inversa del tiempo medio de servicio respectivamente.

Se puede verificar que el intervalo de confianza para el tiempo medio de un cliente en el sistema computado mediante simulaciones coincide con el valor que se obtiene de la ecuación 6.

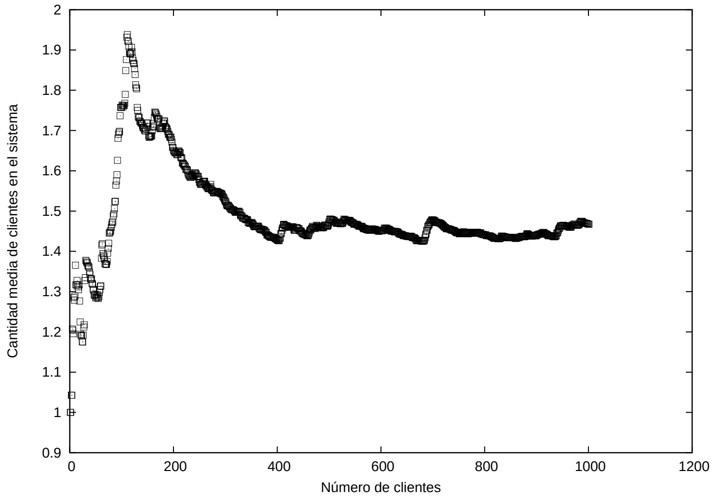

FIGURA 5. Cantidad media de clientes en el sistema en función de la cantidad de arribos.

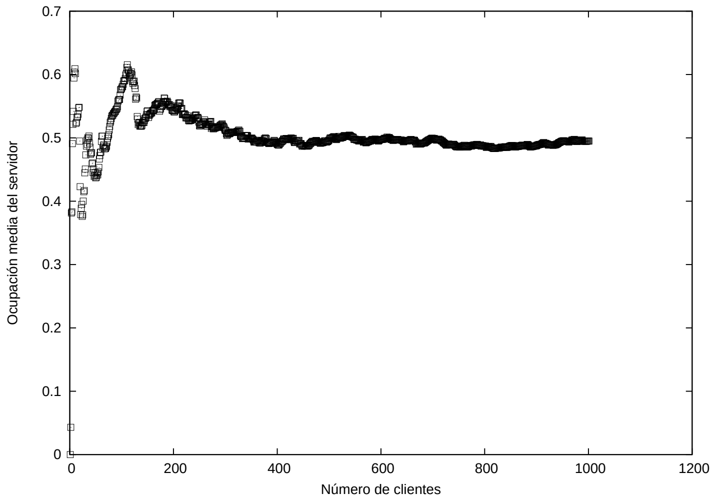

FIGURA 6. Ocupación media del servidor en función de la cantidad de arribos.

Parámetro	Valor estimado
Tiempo medio de un cliente en el sistema [min]	$1{,}006330 \pm 0{,}020162$
Número máximo de clientes en cola	$5{,}129032 \pm 0{,}202543$
Máximo retardo en cola [min]	$3{,}204629 \pm 0{,}131135$
Máximo retardo en el sistema [min]	$3{,}849013 \pm 0{,}136589$
Proporción de clientes con retardo en cola superior a 1 minuto.	$0{,}185484 \pm 0{,}010229$
Proporción de clientes con retardo en el sistema superior a 1 minuto.	$0{,}365917 \pm 0{,}010162$

TABLA 2. Estimaciones de parámetros del sistema de cola simple (M/M/1 con $\mu = 2$) con sus intervalos de confianza.

Parámetro	Valor estimado
Tiempo medio de un cliente en el sistema [min]	$1{,}258275 \pm 0{,}031788$
Número máximo de clientes en cola	$4{,}232143 \pm 0{,}075973$
Máximo retardo en cola [min]	$2{,}496609 \pm 0{,}053087$
Máximo retardo en el sistema [min]	$6{,}230678 \pm 0{,}153826$
Proporción de clientes con retardo en cola en cola superior a 1 minuto.	$0{,}103036 \pm 0{,}010740$
Proporción de clientes con retardo en el sistema superior a 1 minuto.	$0{,}479076 \pm 0{,}009895$

Tabla 3. Estimaciones de parámetros del sistema de cola simple (M/M/2 con $\mu_1 = \mu_2 = 1$) con sus intervalos de confianza.

3.2. Cola simple con dos canales. Se desea estudiar si una cola simple con dos facilidades es más eficiente que una cola simple con un único canal doblemente rápido. Para ello, se simula una cola simple con 2 canales con tiempos de servicio $\mu_1 = \mu_2 = 1$.

En la tabla 3 se presentan los resultados de la simulación. La cota para el criterio de detención se fija en 0,0005 resultando 56 iteraciones. No se obtienen los mismos resultados que en el modelo con una única facilidad. Por ejemplo, el tiempo medio de un cliente en el sistema es mayor en este caso. En consecuencia, la proporción de clientes con demora mayor a un minuto también es mayor. También se puede observar que el retardo máximo en el sistema es prácticamente el doble que en el modelo con un único canal. Este resultado se debe a que la varianza del tiempo medio de servicio en este caso es mayor. En consecuencia, los tiempos de servicio se encuentran más dispersos, generando retardos máximos más grandes. También se puede observar que en este caso son menores tanto el número máximo de clientes en cola, como el máximo retardo en la misma y la proporción de clientes con demora superior al minuto. Es decir que en este modelo, los clientes esperan menos en cola pero más en el servidor.

4. Conclusiones

En un sistema modelado como cola simple, la naturaleza de estabilidad es exclusivamente dependiente de la *intensidad de tráfico*. El tiempo medio de un cliente en el sistema es mayor en una cola simple con dos canales que en una cola simple con una única facilidad doblemente rápida. En este sistema con un único canal, los clientes permanecen un mayor tiempo en cola y un menor tiempo en servicio que en el sistema de dos canales.

Referencias

[Devore, 2003] Devore, Jay L. (2003), *Probabilidad y estadística para ingeniería y ciencias.* 5º edición. Thomson Learning.