ANALISIS DE COLAS DETERMINISTAS CON DATOS FUZZY

PARDO SANTIAGO, María José
CASTRO IÑIGO, Belén

Departamento Economía Aplicada I
Universidad del País Vasco

Abstract

El objetivo del trabajo es combinar la capacidad de los Conjuntos Borrosos para representar situaciones prácticas con las aproximaciones tradicionales a los modelos deterministas de la Teoría de Colas.
El análisis desarrolla un modelo de línea de espera para el cual los tiempos entre llegadas son constantes y los tiempos de servicio se comportan de manera incierta, y pueden ser representados mediante una distribución de posibilidad acorde a un número borroso trapezoidal. El modelo se obtiene tomando como punto de partida el estudio clásico del modelo D/D/1.
Los resultados teóricos se aplican, en un ejemplo numérico, para apoyar la toma de decisiones para el establecimiento de un sistema de citas. Se añade una comparación de los resultados obtenidos en la aplicación con un estudio de simulación.

1. Introducción.

El modelo Determinista de la Teoría de Colas es el más elemental y simple: los tiempos entre sucesivas llegadas de las unidades al sistema y los tiempos de servicio, de un único servidor, son constantes. De aquí que no haya recibido apenas atención por parte de los investigadores de la Teoría de Colas, de tal forma que no se han desarrollado de manera exhaustiva las expresiones de las medidas de efectividad que caracterizan a estos modelos, ni se ha incorporado en sus supuestos la posibilidad de datos inciertos mediante la Teoría de los Conjuntos Borrosos.

El presente trabajo plantea un doble objetivo: por un lado, desarrollar todas las medidas de efectividad de los modelos de colas deterministas -número de clientes en cola y tiempo de permanencia en cola- y por otro lado, en base a dichas expresiones, incorporar a los modelos la capacidad de los conjuntos borrosos para describir situaciones inciertas, con el estudio de un modelo de línea de espera para el cual los tiempos entre llegadas son constantes pero los tiempos de servicio se comportan de manera incierta, siendo representados mediante una distribución de posibilidad.

Para el desarrollo del modelo clásico se toma como punto de partida los estudios realizados por *Saaty* (1967) y *Gross* y *Harris* (1985) sobre los modelos de colas deterministas puros D/D/1. Para el tratamiento del modelo con datos inciertos se sigue el trabajo llevado a cabo por *Li* y *Lee* (1989) para analizar los casos más simples de colas fuzzy a través del principio de extensión de *Zadeh*, junto con el tratamiento de la incertidumbre realizado por *Kaufmann* y *Gil Aluja* (1987) los cuales introducen la Teoría de los Conjuntos Borrosos en modelos matemáticos de gestión.

Los resultados teóricos se aplican a un ejemplo numérico para apoyar la toma de decisiones en el establecimiento de un sistema de citas. El objetivo es determinar el número de clientes a citar en un tiempo de servicio prefijado a fin de que sea máximo, con la condición de que dada una capacidad limitada del sistema, ningún cliente sea rechazado del mismo.

El modelo se completa realizando una comparación de los resultados obtenidos en la aplicación con un estudio de simulación de la cola determinista con datos fuzzy. Se sigue el procedimiento de *Chanas* y *Nowakowski* (1988) para generar un valor único de una variable fuzzy.

2. Modelo clásico de colas determinista puro D/D/1.

2.1. Descripción del modelo.

En el modelo de colas clásico determinista los clientes acceden al sistema con tiempo entre llegadas constante y son servidos con tiempo de servicio también constante. Se supone un sistema de colas organizado de tal manera que las llegadas ocurren a intervalos regulares y son servidas regularmente, en donde los clientes llegan a un único canal, es decir sólo hay un servidor, con una disciplina de cola ordenada de forma que el primero en llegar es el primero en ser atendido y con una fuente de entrada ilimitada.

En este modelo de colas, dependiendo de la relación del tiempo entre llegadas con el tiempo de servicio, se tienen dos situaciones posibles:

1. que el tiempo entre llegadas sea mayor que el tiempo de servicio,
2. que el tiempo entre llegadas sea menor que el tiempo de servicio,

cada una de estas situaciones da lugar a una evolución distinta del sistema siendo necesario un estudio por separado. En lo que sigue se considera, sin pérdida de generalidad, que el instante inicial del sistema es el instante 0 y que la llegada del primer cliente al sistema se produce en dicho instante. Se denota por:

a: tiempo entre sucesivas llegadas,

b: tiempo de servicio,

I: número de clientes en el sistema en el instante inicial.

El valor de I estará formado por el cliente que llega en dicho instante primero más los que ya se encontraban en el sistema. Por ello, el caso I = 1 indicará que a la llegada del primer cliente se encuentra el sistema vacío.

2.1.1. Modelo D/D/1 con tiempo entre llegadas mayor que el tiempo de servicio.

Al ser en este modelo $b<a$, el número de clientes que llegan al sistema por unidad de tiempo es menor al número de servicios. Al haber I clientes iniciales, mientras éstos son atendidos nuevos clientes van llegando a la línea de espera. Dado que el sistema es capaz de servir más rápido que lo que tardan los clientes en llegar, se van absorbiendo estos nuevos clientes de manera que se alcanza un instante en el cual el servidor queda ocioso hasta la llegada del siguiente cliente. Desde dicho instante el sistema se encuentra en *estado estacionario* de manera que cada nueva llegada es atendida antes de la llegada del siguiente, siendo desde ese momento los valores del número de clientes en cola y del tiempo de permanencia en cola de cero.

Para caracterizar el modelo es necesario determinar cuándo el sistema alcanza ese instante y cuántas nuevas llegadas se producen hasta ese momento, ya que junto con las I iniciales son los clientes que no encuentran el servidor ocioso a su llegada y tienen que esperar para ser atendidos. Se denota en lo que sigue por:

A: llegadas al sistema, desde el instante inicial, necesarias para alcanzar el estado estacionario,

T: instante de tiempo en el cual el sistema alcanza el estado estacionario.

Se indica a continuación el desarrollo seguido para el cálculo de los valores A y T:

- El sistema alcanza el *estado estacionario* en T, una vez el servidor ha atendido a los I clientes iniciales y a las nuevas llegadas A. Teniendo en cuenta que cada servicio tiene una duración de b, se cumple: $T = (I + A)b$.
- La llegada A+1 será la primera que encuentre al servidor libre y pase directamente a ser atendida. Esta llegada ocurrirá en: $(A+1)a$. De aquí que el momento en que llegue el cliente $(A+1)$ tendrá que ser superior al tiempo de atender a los $I+A$ clientes, de otra forma no sería posible que dicha llegada encontrara el sistema libre, luego:

$$(A+1)a > (I + A)b \Leftrightarrow A > \frac{Ib-a}{a-b}$$

Como A tiene que ser el primer número entero que cumple la expresión anterior, entonces: $A = \left[\frac{Ib-a}{a-b}\right] + 1$, o de forma equivalente $A = \left[\frac{b(I-1)}{a-b}\right]$ donde con $[x]$ se representa la parte entera del número x.

- El tiempo total hasta que la instalación del servicio esté ociosa por primera vez es:

$$T = (I + A)b = \left(I + \left[\frac{b(I-1)}{a-b}\right]\right)b = \left[\frac{Ia-b}{a-b}\right]b$$

Caracterizado el sistema con los valores de A y de T se pasa a describir alguna de las principales medidas de efectividad del modelo a partir de las cuales se pueden obtener las demás. Se denotarán por n_q^t y W_q^n al número de clientes que hay en la cola en el instante t y al tiempo de permanencia en la cola del cliente n-ésimo, respectivamente.

- **Número de clientes en cola en el instante *t*: n_q^t.**

El número de clientes en cola en el instante t toma el valor:

$$n_q^t = \begin{cases} \left[\dfrac{t}{a}\right] + I - 1 - \left[\dfrac{t}{b}\right] & si \quad 0 \le t < T \\ 0 & si \quad t \ge T \end{cases}$$

Demostración:

1. Sea $0 \le t < T$. El número de clientes en cola en el instante t será la diferencia entre el número de clientes que han llegado al sistema hasta ese momento menos los clientes que en ese mismo espacio de tiempo ya han sido, o están siendo, atendidos. Para un tiempo t, al sistema han accedido $[t/a]$ nuevos clientes más los I iniciales, que en total son: $[t/a] + I - 1$, descontado el cliente que, estando en el sistema, es atendido por el servidor y no forma cola. En ese mismo espacio de tiempo t, se han atendido $[t/b]$ clientes que ya han abandonado el sistema, por tanto quedan en línea de espera: $[t/a] + I - 1 - [t/b]$.
2. Sea $t \ge T$. El sistema ya ha alcanzado el *estado estacionario* y cada nueva llegada es atendida antes de que acceda la siguiente, no formándose línea de espera, y la cola será de 0.

- **Tiempo de permanencia en cola del *n*-ésimo cliente: W_q^n.**

El tiempo de permanencia en cola del n-ésimo cliente toma el valor:

$$W_q^n = \begin{cases} (n-1)b & si \quad n \le I \\ (n-1)b - (n-I)a & si \quad I < n \le I + A \\ 0 & si \quad n > I + A \end{cases}$$

Demostración:

Para un cliente determinado su tiempo de espera en cola vendrá dado por la diferencia entre el momento en que entra a ser servido y su tiempo de llegada a las instalaciones. Si se denota por T_{ll}^n al tiempo de llegada del cliente n, entonces:

$$T_{ll}^n = \begin{cases} 0 & si \quad n \le I \\ (n-I)a & si \quad n > I \end{cases}$$

1. Sea $n \le I$. Su tiempo de espera en cola es el que se tarde en atender a los n-1 clientes anteriores a él: $(n-1)b$.

2. Sea $I < n \leq I + A$. Son los nuevos clientes que no encuentran el sistema libre y tienen que esperar para ser atendidos. El momento en que entra la unidad n-ésima a ser servida es en: $(n-1)b$, que es lo que tarda el sistema en atender a las llegadas anteriores a él. El tiempo de llegada al sistema de la unidad *n*-ésima viene dado por: $(n-I)a$. Por lo que el tiempo de permanencia en cola del cliente *n* queda: $(n-1)b-(n-I)a$.
3. Sea $n > I + A$. Los clientes encuentran al servidor ocioso y son atendidos inmediatamente, con un tiempo de permanencia en cola de 0.

2.1.2. Modelo D/D/1/K con tiempo entre llegadas menor que el tiempo de servicio.

En este modelo se cumple que $a<b$, así el número de clientes que llegan al sistema por unidad de tiempo es mayor al número de servicios. Si este sistema no se controla nunca alcanzará un *estado estacionario*, puesto que los clientes se acumularán en sus instalaciones, creciendo la línea de espera de forma ilimitada. Debido a ello cada cliente deberá esperar para ser servido más que sus predecesores. Para prevenir esto y asegurar el funcionamiento del sistema, se establece una capacidad máxima de *K* clientes en las instalaciones. Cuando se alcanza esta cantidad de *K* clientes ninguno más es admitido hasta que el número de clientes en el sistema disminuye a *K*-1, lo que ocurre cuando un cliente ha completado su servicio y lo abandona. Se fija para el desarrollo del modelo la restricción $I < K$.

El sistema alcanza el *estado estacionario* cuando en el mismo se acumulan *K* clientes. Para caracterizarlo se necesita calcular el instante en el que por primera vez se produce esta circunstancia y el cliente que en ese momento es rechazado. Se denota en lo que sigue por:

n_{t_k} : primera llegada que se rechaza en el sistema,

t_k : instante a partir del cual se rechazan clientes en el sistema y se alcanza el *estado estacionario*.

Se pasa a exponer el desarrollo del cálculo de los valores de n_{t_k} y t_k :

- Por una parte, el tiempo máximo de espera en este sistema corresponde a la de aquel cliente que a su llegada ocupe la posición *K*-ésima en las instalaciones del sistema. Su tiempo de espera coincidirá con el servicio completo de los *K*-1 clientes anteriores a él: $(K-1)b$.
- Por otra parte, el tiempo de permanencia en cola del *n*-ésimo cliente es la diferencia entre el momento de empezar a ser servido, $(n-1)b$ (empieza a ser servido cuando completa su servicio el anterior), y el instante de acceder al sistema, $T_{ll}^{n} = (n-I)a$. Así, el cliente *n* permanece en cola: $(n-1)b-(n-I)a$. Esta diferencia es creciente en *n*, cada nuevo cliente tiene que esperar en línea más que el anterior.

- Aquel cliente n_{t_k} que cumple por primera vez que su tiempo de permanencia en cola es superior a lo máximo que se puede esperar en el sistema tendrá que ser rechazado. Esta condición indica que a la llegada del cliente *n*-ésimo al sistema en él hay ya *K* clientes y que esto sucede por primera vez. Por tanto:

$$\left(n_{t_k} - 1\right)b - \left(n_{t_k} - I\right)a > (K-1)b \Leftrightarrow n_{t_k} > \frac{Kb - Ia}{b-a}$$

n_{t_k} tiene que ser el primer número entero que cumple la expresión anterior, entonces:

$$n_{t_k} = \left[\frac{Kb - Ia}{b-a}\right] + 1$$

- Como t_k es el instante en que se rechaza al cliente n_{t_k}, su valor es el de la llegada del mismo:

$$t_k = T_{ll}^{n_{t_k}} = \left(n_{t_k} - I\right)a$$

Con los valores de n_{t_k} y t_k se tiene caracterizado el sistema. Se pasa a describir las medidas de efectividad para el modelo.

- **Número de clientes en cola en el instante *t*: n_q^t.**

Sean $0 \le r < a; 0 \le r_1 < b$, tal que $t \equiv r \mod(a)$[1]; $t \equiv r_1 \mod(b)$, se cumple que:

$$n_q^t = \begin{cases} \left[\frac{t}{a}\right] + I - 1 - \left[\frac{t}{b}\right] & si \quad 0 \le t < t_k \\ K-2 & si \quad t \ge t_k \ ; r > r_1 \\ K-1 & si \quad t \ge t_k \ ; r \le r_1 \end{cases}$$

Demostración:

1. Sea $0 \le t < t_k$. No se ha alcanzado todavía el *estado estacionario* y el número de clientes en cola será igual al total de llegadas hasta *t* menos el total de servicios completados hasta ese mismo momento, excluyendo al cliente que se encuentra en el servicio en ese instante. Por el modelo anterior se tiene: $[t/a] + I - 1 - [t/b]$.
2. Sea $t \ge t_k$. El sistema se encuentra en *estado estacionario* oscilando el número de clientes en cola entre *K*-1 y *K*-2.

[1] Por brevedad en las expresiones, se representa por $x \equiv r \mod(y)$ con $0 \le r < y$ al valor de $r = x - [x/y]y$. Si los valores de *x* y de *y* fuesen números enteros, *r* correspondería al resto de la división de *x* entre *y*. En nuestro caso los valores quizá no sean enteros por lo que se realiza un abuso en la notación con el objetivo de facilitar la lectura de las expresiones.

- Como se produce la salida de un cliente de las instalaciones a intervalos de amplitud *b*, todo instante *t* estará entre dos salidas consecutivas, que se denotarán salida *m* y salida *m*+1. Se cumple entonces que $mb \leq t < (m+1)b$, luego:

$$\left.\begin{array}{r} mb \leq t \Leftrightarrow m \leq \dfrac{t}{b} \\ t < (m+1)b \Leftrightarrow m > \dfrac{t}{b} - 1 \end{array}\right\} \Rightarrow \frac{t}{b} - 1 < m \leq \frac{t}{b}$$

Dado que $m \in N^+$, es $m = [t/b]$ de forma que la salida anterior al instante *t*, más próxima a *t*, se produce en el momento $mb = [t/b]b$.

- Como se produce la llegada de un cliente al sistema a intervalos de amplitud *a*, todo instante *t* estará entre dos llegadas consecutivas, que se denotarán por llegada n_1 y llegada n_1+1. Por ello se cumple que $n_1 a \leq t < (n_1 + 1)a$. Despejando como en el caso anterior se obtiene: $n_1 = [t/a]$, y por tanto la llegada anterior a *t*, más próxima a *t*, se produce en el instante $n_1 a = [t/a]a$.

De esta forma:

I. Hay *K*-2 clientes únicamente desde que se produce una salida hasta la llegada siguiente. Es decir, si entre la salida anterior a *t* y el instante *t* no se produce ninguna llegada entonces en *t* hay *K*-2 clientes en cola, lo que equivale a que la llegada más próxima a *t* sea anterior a la salida más próxima: $[t/a]a < [t/b]b$. Si $r = t - [t/a]a$ y $r_1 = t - [t/b]b$, entonces la condición anterior se traduce en $r > r_1$.

II. Hay *K*-1 clientes si *t* es igual o posterior a una llegada, es decir, si hay una llegada entre *t* y la salida anterior a *t*. De acuerdo al apartado I, en este caso se cumple: $[t/a]a \geq [t/b]b$ o lo que es equivalente $r \leq r_1$.

- **Tiempo de permanencia en cola del *n*-ésimo cliente: W_q^n.**

El cliente *n*-ésimo es aceptado en el sistema si $n < n_{t_k}$ o, si $n > n_{t_k}$ y cumple que $0 \leq r < a$, para $(n-I)a \equiv r \text{ mod}(b)$ con $0 \leq r < b$. En caso de ser aceptado, su tiempo de permanencia en cola es:

$$W_q^n = \begin{cases} (n-1)b & si \quad n \leq I \\ (n-1)b - (n-I)a & si \quad I < n < n_{t_k} \\ (K-1)b - r & si \quad n > n_{t_k} \end{cases}$$

Si por el contrario $n \geq n_{t_k}$ y $a \leq r < b$, entonces el cliente es rechazado.

Demostración:

1. Sea $n \leq I$. El cliente es uno de los iniciales llegados al sistema y tendrá que esperar a que atiendan a los *n*-1 anteriores a él: $(n-1)b$.
2. Sea $I < n < n_{t_k}$. Dado que n_{t_k} es el primer cliente rechazado, *n* es aceptado en el sistema teniendo que permanecer en cola el tiempo que transcurre desde su llegada a las instalaciones hasta que empieza a ser servido, que coincide con la salida del cliente *n*-1. Se conoce que: $T_{ll}^{n} = (n-I)a$ y $T_{s}^{n-1} = (n-1)b$, entonces el tiempo de permanencia en cola del cliente *n* es: $(n-1)b-(n-I)a$.
3. Sea $n > n_{t_k}$, ($n = n_{t_k}$ es el primer cliente rechazado). El cliente *n*-ésimo accede al sistema cuando éste ya ha alcanzado el *estado estacionario* con *K* clientes, pudiendo ser aceptado o rechazado. Será rechazado si a su llegada se encuentran *K*-1 clientes en cola. Por el contrario, si a su llegada la línea de espera a descendido a *K*-2, por haber terminado su servicio un cliente, el *n*-ésimo será aceptado. Esto último ocurre cuando entre la salida más próxima a la llegada del cliente *n* y el instante de llegada de dicho cliente no ha tenido lugar ninguna otra llegada. Es decir, se sabe que las salidas se producen de forma regular a intervalos de amplitud *b*, así que la llegada del cliente *n* se encuentra entre dos salidas que se denotarán *m* y *m*+1:

$$mb \leq T_{ll}^{n} < (m+1)b \Rightarrow mb \leq (n-I)a < (m+1)b$$

Despejando *m* se llega a: $m = [(n-I)a/b]$ con lo que $mb = [(n-I)a/b]b$ es el momento en que se produce la salida más cercana o más próxima a la llegada de *n*. Si desde *mb* hasta la llegada de *n* transcurre un tiempo superior a *a*, entonces ha tenido que acceder otro cliente al sistema y no puede ser la cola de *K*-2 clientes. Por tanto, para que cuando llega *n* sea aceptado, su instante de entrada en el sistema tiene que estar comprendido en el intervalo $[mb, mb+a)$, es decir: $mb \leq (n-I)a < mb+a$; y será rechazado si: $mb+a \leq (n-I)a < (m+1)b$. Sustituyendo el valor de *m* en estas expresiones:

$$\left[\frac{(n-I)a}{b}\right]b \leq (n-I)a < \left[\frac{(n-I)a}{b}\right]b + a$$

$$\left[\frac{(n-I)a}{b}\right]b + a \leq (n-I)a < \left(\left[\frac{(n-I)a}{b}\right]+1\right)b$$

Sea $(n-I)a \equiv r \bmod(b)$ con $0 \leq r < b$, entonces la condición para que el cliente *n* sea aceptado en el sistema se traduce en que $0 \leq r < a$, y la condición para ser rechazado se traduce en que $a \leq r < b$.

Bajo la condición de que el cliente *n* es aceptado, se pasa a calcular su tiempo de permanencia en cola.

- Si la llegada del cliente *n* se produce en el instante de una salida, entonces en ese momento hay *K*-1 clientes antes que él en las instalaciones. Su espera será la situación de máxima permanencia en cola, y toma el valor: $(K-1)b$.

- Si la llegada de *n* dista una cantidad *x* de la salida anterior, deberá esperar: $(K-1)b - x$, donde *x* es el tiempo que el servidor está atendiendo al cliente siguiente. El valor de *x* viene dado por la diferencia entre el tiempo de llegada de *n* y el instante de la salida más próxima a *n*, que es *mb*:

$$x = T_{ll}^{n} - mb = (n-I)a - \left[\frac{(n-I)a}{b}\right]b = r$$

En el caso en que la llegada se produzca a la vez que la salida, se cumple: $x = T_{ll}^{n} - mb = 0 \Rightarrow r = 0$. En ambos casos el tiempo de permanencia en cola se escribe: $(K-1)b - r$.

3. Modelo de colas con tiempos de servicio fuzzy $D/\tilde{D}/1/K$.

Los modelos de colas deterministas desarrollados anteriormente se ajustan ahora a uno fuzzy que puede expresarse en una notación similar a la de *Kendall* como $D/\tilde{D}/1/K$. *Li-Lee* (1989) consideran que cada modelo de colas fuzzy puede ser considerado como una percepción de un sistema de colas usual al que se le puede denominar el original del modelo de colas fuzzy. El conjunto *Q* de todos los posibles originales del modelo fuzzy $D/\tilde{D}/1/K$ se puede escribir como: $Q = \left\{(D/D/1/K)/b \in \text{soporte}\,\tilde{b}\right\}$. Proponen obtener las distribuciones de posibilidad de las medidas de efectividad de los modelos de colas fuzzy aplicando el principio de extensión de *Zadeh* desde las soluciones de sus modelos originales con los parámetros conocidos de forma precisa, de manera que la función de pertenencia de la cola fuzzy es:

$$\mu_{(D/\tilde{D}/1/K)}(D/D/1/K) = \mu_{\tilde{b}}(b)$$

donde $\mu_{\tilde{b}}(b)$ representa la función de pertenencia del número borroso $\tilde{b}$. En general, todas las funciones de parámetro fuzzy $\tilde{b}$ pueden ser definidas por:

$$\mu_{f(\tilde{b})}(y) = \sup_{x\in R}\left\{\mu_{\tilde{b}}(x) / y = f(x)\right\}$$

Sea un modelo de línea de espera al que los clientes acceden a una tasa de 1/*a* clientes por unidad de tiempo y en el que los servicios se producen de forma incierta y pueden aproximarse mediante un número borroso trapezoidal que se representa por $\tilde{b} = [b_1, b_2, b_3, b_4]$, con $b_2 \le a \le b_3$, y se produce sin iteración con las llegadas. Se fija para el desarrollo del modelo la restricción $I < K$.

Por ser en este modelo $b_2 \le a \le b_3$, la evolución del sistema en los casos extremos puede ser o bien encontrarse en un instante determinado el servidor ocioso por servir a todos los clientes de forma más rápida que el ritmo de llegadas, o bien acumularse los clientes en cola por atenderlos con un tiempo superior al tiempo entre llegadas, hasta alcanzar la capacidad máxima de *K* clientes en las instalaciones. Se caracteriza el sistema calculando el momento primero en que se pueden dar estas situaciones, es decir, el instante en que el servidor puede quedar ocioso por primera vez y cuántos nuevos clientes llegarán a las instalaciones hasta ese momento, o el instante en el que el sistema puede rechazar al primer cliente y cuál puede ser ese primer cliente rechazado. Se denota en lo que sigue por:

$\tilde{A}$: nuevas llegadas al sistema antes de poder quedar el servidor ocioso,

$\tilde{T}$: instante de tiempo a partir del cual puede quedar el servidor ocioso,

$\tilde{n}_{t_k}$: primer cliente que puede ser rechazado en el sistema,

$\tilde{t}_k$: instante de tiempo en el que se pueden rechazar clientes en el sistema.

Aplicando el principio de extensión de *Zadeh* se pueden obtener los resultados para el modelo fuzzy con parámetro $\tilde{b}$, pero dada la dificultad de aplicación del mismo se ha comprobado que las expresiones de los modelos originales de colas deterministas *A*, *T*, n_{t_k} y t_k cumplen ser, para *A* y *T*, funciones crecientes respecto a la variable *b*, y funciones decrecientes para n_{t_k} y t_k, de forma que todas ellas pueden ser representadas mediante el uso de α-cortes. Sea $b_\alpha = \left[\underline{b}_\alpha, \overline{b}_\alpha\right]$ el α-corte de $\tilde{b}$, entonces:

$$A_\alpha = \left[\left[\frac{\underline{b}_\alpha (I-1)}{a - \underline{b}_\alpha}\right], \infty\right) \qquad T_\alpha = \left[\left(I + \underline{A}_\alpha\right)\underline{b}_\alpha, \infty\right)$$

$$n_{t_k \alpha} = \left[\left[\frac{K\overline{b}_\alpha - Ia}{\overline{b}_\alpha - a}\right] + 1, \infty\right) \qquad t_{k\alpha} = \left[\left(\underline{n}_{t_k \alpha} - I\right)a, \infty\right)$$

Las expresiones anteriores sólo van a estar acotadas inferiormente ya que, por una parte, puede suceder que el sistema nunca tenga al servidor ocioso si evoluciona con un tiempo de servicio superior al tiempo entre llegadas o, por otra parte, también puede suceder que el sistema nunca rechace a ningún cliente (y todos son aceptados en las instalaciones) en el caso de que su evolución sea la contraria. De aquí que el extremo superior en cada una de ellas va a ser infinito y solamente es necesario calcular la primera vez que se pueden dar estas situaciones.

A partir de $\tilde{A}$, $\tilde{T}$, $\tilde{n}_{t_k}$, $\tilde{t}_k$ que caracterizan el sistema de colas fuzzy se desarrollan las medidas de efectividad $\tilde{n}_q^t$ y $\tilde{W}_q^n$. Para la aplicación del principio de extensión de *Zadeh* es necesario determinar el modelo de colas original al modelo con datos fuzzy, pero para las expresiones de las medidas de efectividad $\tilde{n}_q^t$ y $\tilde{W}_q^n$ existen dos originales al modelo fuzzy, dependiendo de la relación entre el tiempo entre llegadas y el tiempo de servicio, por lo que no es posible la aplicación de este principio.

Teniendo en cuenta que todo número borroso $\tilde{C}$ puede ser escrito de la forma: $\tilde{C} = \bigcup_{\alpha} \alpha C_\alpha$ con $C_\alpha = \left[\underline{C}_\alpha, \overline{C}_\alpha\right]$ el α-corte de $\tilde{C}$, $\forall\alpha \in [0,1]$, las medidas de efectividad quedan:

$$\tilde{n}_q^t = \bigcup_{\alpha} \alpha n_{q\alpha}^t \text{ con } n_{q\alpha}^t = \left[\underline{n}_{q\alpha}^t, \overline{n}_{q\alpha}^t\right], \quad \tilde{W}_q^n = \bigcup_{\alpha} \alpha W_{q\alpha}^n \text{ con } W_{q\alpha}^n = \left[\underline{W}_{q\alpha}^n, \overline{W}_{q\alpha}^n\right], \quad \forall\alpha \in [0,1]$$

de manera que si se conocen los α-cortes $n_{q\alpha}^t$ y $W_{q\alpha}^n$ se conocen las medidas de efectividad del modelo con datos borrosos.

Los valores de $n_{q\alpha}^t$ y $W_{q\alpha}^n$ son los intervalos, dado α fijo, del número de clientes en cola en el instante *t* y el tiempo de permanencia en cola del cliente *n*-ésimo cuando el tiempo de servicio $\tilde{b}$ toma sus valores en el mismo α-corte, es decir, cuando el tiempo de servicio se mueve en el intervalo $\left[\underline{b}_\alpha, \overline{b}_\alpha\right]$, por lo que se pasa a calcular las expresiones del número de clientes en cola en *t* y tiempo de permanencia en cola del cliente *n*-ésimo, para el modelo de colas con tiempo de servicio comprendido en el intervalo $\left[\underline{b}, \overline{b}\right]$ y tiempo entre llegadas *a*, constante , modelo que siguiendo la notación de *Kendall* se denotará D/DI/1/K. En este caso se caracteriza el modelo con:

$$\underline{A} = \left[\frac{\underline{b}(I-1)}{a-\underline{b}}\right] \qquad \underline{T} = \left(I + \underline{A}\right)\underline{b}$$

$$\underline{n}_{t_k} = \left[\frac{K\overline{b} - Ia}{\overline{b} - a}\right] + 1 \qquad \underline{t}_k = \left(\underline{n}_{t_k} - I\right)a$$

Con los valores de $\underline{A}$, $\underline{T}$, $\underline{n}_{t_k}$ y $\underline{t}_k$ se sabe que si al sistema llegan los clientes a una tasa constante de 1/*a* clientes por unidad de tiempo y el tiempo de servicio oscila en el intervalo $\left[\underline{b}, \overline{b}\right]$, con $\underline{b} \le a \le \overline{b}$, el sistema evoluciona de forma que con menos de $\underline{A}$ nuevas llegadas o antes del instante $\underline{T}$ no es posible que un cliente encuentre el sistema vacío a su llegada, y todos los clientes llegados antes del cliente $\underline{n}_{t_k}$ o del instante $\underline{t}_k$ son aceptados en las instalaciones.

- **Intervalo del número de clientes en cola en el instante *t*: $\left[\underline{n}_q^t, \overline{n}_q^t\right]$.**

El número de clientes que hay en la cola en el instante *t* puede variar en el intervalo $\left[\underline{n}_q^t, \overline{n}_q^t\right]$, donde:

$$\underline{n}_q^t = \begin{cases} \left[\dfrac{t}{a}\right] + I - 1 - \left[\dfrac{t}{\underline{b}}\right] & si \quad 0 \le t < \underline{T} \\ 0 & si \quad t \ge \underline{T} \end{cases}$$

$$\overline{n}_q^t = \begin{cases} \left[\dfrac{t}{a}\right] + I - 1 - \left[\dfrac{t}{\overline{b}}\right] & si \quad 0 \le t < \underline{t}_k \\ K - 1 & si \quad t \ge \underline{t}_k \end{cases}$$

Demostración:

1. El valor mínimo del intervalo del número de clientes en cola en *t* se obtiene:
   - Sea $0 \le t < \underline{T}$. Inicialmente en el sistema hay I clientes, y en el mejor de los casos la cola irá descendiendo hasta ser de *cero* en el instante $\underline{T}$, luego si *t* es anterior a $\underline{T}$ la cola está formada por el número de llegadas al sistema hasta dicho instante, $[t/a] + I$, menos el número de servicios que se han completado, si se realizan con un tiempo de servicio de $\underline{b}$, $[t/\underline{b}]$, excluyendo al cliente que se encuentra en servicio. Así: $[t/a] + I - 1 - [t/\underline{b}]$.
   - Sea $t \ge \underline{T}$. El valor mínimo del número de clientes en cola ya es de 0 dado que si se sirve a un ritmo de $\underline{b}$, en $\underline{T}$ el servidor queda ocioso entre servicios y no se forma cola en el sistema.
2. El valor máximo del intervalo del número de clientes en cola en *t* se obtiene:
   - Sea $0 \le t < \underline{t}_k$. Si los servicios se realizan más despacio que las llegadas, se van acumulando los clientes en el sistema hasta alcanzar en $\underline{t}_k$ la capacidad máxima, *K*. Hasta ese momento, el número de clientes en cola está formado por el número de llegadas al sistema hasta dicho instante, $[t/a] + I$, menos el número de servicios que se han completado si se realizan con un tiempo de servicio de $\overline{b}$, $[t/\overline{b}]$, excluyendo al cliente que se encuentra en el servicio. Así: $[t/a] + I - 1 - [t/\overline{b}]$.
   - Sea $t \ge \underline{t}_k$. El sistema puede haber evolucionado hasta tener en cola los *K*-1 permitidos, cantidad que se puede dar en todo *t* como valor máximo, porque en este modelo se cumple que en todo intervalo de una salida hay al menos la llegada de un cliente, de forma que cuando la cola puede descender a *K*-2 porque se completa un servicio es posible que un nuevo cliente acceda al sistema en el mismo instante con lo que se mantiene la cola en *K*-1.

- **Intervalo del tiempo de permanencia en cola del *n*-ésimo cliente: $\left[\underline{W}_q^n, \overline{W}_q^n\right]$.**

El tiempo de permanencia en cola de *n*-ésimo cliente puede variar en el intervalo $\left[\underline{W}_q^n, \overline{W}_q^n\right]$, donde:

$$\underline{W}_q^n = \begin{cases} (n-1)\underline{b} & si \quad n \le I \\ (n-1)\underline{b} - (n-I)a & si \quad I < n \le I + \underline{A} \\ 0 & si \quad n > I + \underline{A} \end{cases}$$

$$\overline{W}_q^n = \begin{cases} (n-1)\overline{b} & si \quad n \le I \\ (n-1)\overline{b} - (n-I)a & si \quad I < n < \underline{n}_{t_k} \\ (K-1)\overline{b} & si \quad n \ge \underline{n}_{t_k} \end{cases}$$

Demostración:

1. El valor mínimo del tiempo de permanencia en cola del *n*-ésimo cliente se obtiene:
   - Sea $n \le I$. El tiempo mínimo de permanencia en cola es el tiempo que tarda el servidor en atender a los *n*-1 clientes anteriores a *n* en el menor tiempo posible: $(n-1)\underline{b}$.
   - Sea $I < n \le I + \underline{A}$. Tiene que permanecer en cola el tiempo que transcurre desde que llega a las instalaciones en $T_{ll}^n = (n-I)a$, hasta el instante lo más pronto posible en que puede pasar a ser servido, instante que coincide con la salida más temprana del cliente *n*-1, que es $(n-1)\underline{b}$. Así: $(n-1)\underline{b} - (n-I)a$.
   - Sea $n > I + \underline{A}$. Es posible que cuando el cliente *n* accede al sistema encuentre al servidor libre, y no tiene que esperar en cola, y su tiempo mínimo de espera es de cero.
2. El valor máximo del tiempo de permanencia en cola del *n*-ésimo cliente se obtiene:
   - Sea $n \le I$. Permanece en cola el tiempo de atender a los *n*-1 clientes anteriores a *n* con el mayor tiempo posible: $(n-1)\overline{b}$.
   - Sea $I < n < \underline{n}_{t_k}$. Permanece en cola el tiempo que transcurre desde su llegada a las instalaciones hasta el instante lo más tarde posible en que puede pasar a ser atendido, instante que coincide con la salida más tardía del cliente anterior *n*-1, que es en $(n-1)\overline{b}$. Así: $(n-1)\overline{b} - (n-I)a$.
   - Sea $n \ge \underline{n}_{t_k}$. El máximo tiempo de permanencia en cola es el de atender a los *K*-1 permitidos en la línea de espera, con un tiempo de $\overline{b}$, lo cual puede ocurrir si justo la llegada del cliente *n* al sistema coincide con la salida de un cliente del mismo, lo que dadas las características estudiadas para este modelo puede suceder a todo cliente *n*. Luego se tiene: $(K-1)\overline{b}$.

A partir de las expresiones obtenidas con tiempos de servicio en intervalo, se determinan las medidas de efectividad para el modelo con tiempos de servicio fuzzy.

- **Número de clientes en cola en el instante *t*: $\tilde{n}_q^t$.**

El número de clientes en cola en el instante *t*, $\tilde{n}_q^t$ es un número borroso, y si se representa por $\tilde{n}_q^t = \bigcup_\alpha \alpha n_{q\alpha}^t$

con $n_{q\alpha}^t = \left[\underline{n}_{q\alpha}^t, \overline{n}_{q\alpha}^t\right]$ el α-corte $\forall\alpha \in [0,1]$, entonces:

$$\underline{n}_{q\alpha}^t = \begin{cases} \left[\frac{t}{a}\right] + I - 1 - \left[\frac{t}{\underline{b}_\alpha}\right] & si \quad 0 \le t < \underline{T}_\alpha \\ 0 & si \quad t \ge \underline{T}_\alpha \end{cases}$$

$$\overline{n}_{q\alpha}^t = \begin{cases} \left[\frac{t}{a}\right] + I - 1 - \left[\frac{t}{\overline{b}_\alpha}\right] & si \quad 0 \le t < \underline{t}_{k\alpha} \\ K - 1 & si \quad t \ge \underline{t}_{k\alpha} \end{cases}$$

Tomando el intervalo del número de clientes en cola en *t* $\left[\underline{n}_q^t, \overline{n}_q^t\right]$, se obtienen los extremos del α-corte $n_{q\alpha}^t = \left[\underline{n}_{q\alpha}^t, \overline{n}_{q\alpha}^t\right]$ sin más que sustituir el correspondiente α-corte de $\tilde{b}$, $\left[\underline{b}_\alpha, \overline{b}_\alpha\right]$.

- **Tiempo de permanencia en cola del *n*-ésimo cliente: $\tilde{W}_q^n$.**

El tiempo de permanencia en cola del *n*-ésimo cliente, $\tilde{W}_q^n$ es un número borroso, y si se representa $\tilde{W}_q^n = \bigcup_\alpha \alpha W_{q\alpha}^n$ con $W_{q\alpha}^n = \left[\underline{W}_{q\alpha}^n, \overline{W}_{q\alpha}^n\right]$ el α-corte $\forall\alpha \in [0,1]$, entonces:

$$\underline{W}_{q\alpha}^n = \begin{cases} (n-1)\underline{b}_\alpha & si \quad n \le I \\ (n-1)\underline{b}_\alpha - (n-I)a & si \quad I < n \le I + \underline{A}_\alpha \\ 0 & si \quad n > I + \underline{A}_\alpha \end{cases}$$

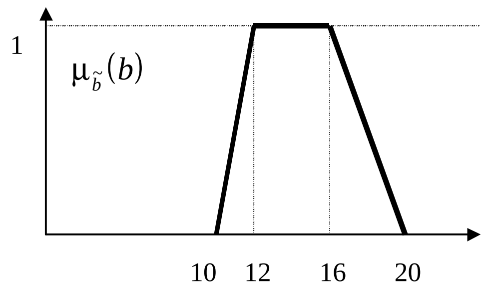
$$\overline{W}_{q\alpha}^n = \begin{cases} (n-1)\overline{b}_\alpha & si \quad n \le I \\ (n-1)\overline{b}_\alpha - (n-I)a & si \quad I < n < \underline{n}_{t_k\,\alpha} \\ (K-1)\overline{b}_\alpha & si \quad n \ge \underline{n}_{t_k\,\alpha} \end{cases}$$

De la misma manera que en la medida de efectividad anterior, con el intervalo del tiempo de permanencia en cola del cliente *n*, $\left[\underline{W}_q^n, \overline{W}_q^n\right]$ se determinan los extremos del α-corte $W_{q\alpha}^n = \left[\underline{W}_{q\alpha}^n, \overline{W}_{q\alpha}^n\right]$ sin más que sustituir el correspondiente α-corte de $\tilde{b}$, $\left[\underline{b}_\alpha, \overline{b}_\alpha\right]$.

Los subconjuntos $\tilde{n}_q^t$ y $\tilde{W}_q^n$ son números borrosos, ambos cumplen ser subconjuntos convexos y están normalizados.

4. Ejemplo de un modelo $D/\tilde{D}/1/K$.

A continuación se expone una aplicación del modelo teórico para el caso del establecimiento de un sistema de citas. Se considera un modelo de colas en el que los clientes son citados de forma constante a lo largo del tiempo de servicio de un único servidor. Los servicios se desarrollan siguiendo una distribución de posibilidad acorde a un número borroso, que tiene función de pertenencia trapezoidal $\tilde{b} = [10{,}12{,}16{,}20]$:



$$\mu_{\tilde{b}}(b) = \begin{cases} 0 & si \quad b \le 10 \\ \dfrac{b-10}{2} & si \quad 10 \le b \le 12 \\ 1 & si \quad 12 \le b \le 16 \\ \dfrac{20-b}{4} & si \quad 16 \le b \le 20 \\ 0 & si \quad b \ge 20 \end{cases}$$

El número inicial de clientes en el sistema es I = 3 y la capacidad del sistema de K = 5. El sistema cita a los clientes a lo largo de un tiempo establecido de 300 minutos, a partir del cual ningún cliente nuevo es aceptado en las instalaciones, y el servidor termina de atender a los que permanecen en cola. El objetivo es determinar el tiempo entre sucesivas llegadas, *a*, de manera que con un nivel de aceptación del 90%, en los 300 min. se citen al mayor número posible de clientes a fin de que el servidor se encuentre libre lo mínimo, con la condición de que la longitud de la cola sea menor que la capacidad del sistema para no tener que rechazar a ningún cliente que sea citado previamente.

Para que todos los clientes citados en los 300 min. sean aceptados en las instalaciones es necesario que el primer instante en que un cliente puede ser rechazado en este sistema lo sea con un tiempo posterior a 300 min. Se sabe por el modelo con datos fuzzy que $\tilde{t}_k$ proporciona el instante de tiempo en el que se pueden rechazar clientes en el sistema. Así, se trata de calcular el valor de *a* de forma que se cumpla que $\tilde{t}_k$ sea igual o mayor a 300, y ello con un nivel de aceptación del 90%, es decir, con $\alpha = 0{,}90$ lo que determina un tiempo de servicio $b_{0.9} = \left[\underline{b}_{0,9}, \overline{b}_{0,9}\right] = [11'8{,}16'4]$:

$$t_{k\alpha} = \left[\left(\underline{n}_{t_k\,\alpha} - I\right)a, \infty\right) = \left[\left(\left[\frac{K\overline{b}_\alpha - Ia}{\overline{b}_\alpha - a}\right] + 1 - I\right)a, \infty\right)$$

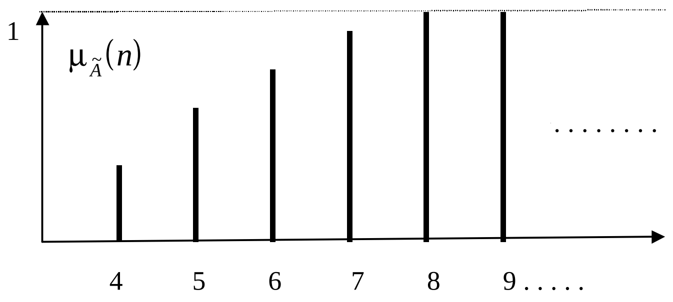
$$t_{k\,0,9} = \left[\left(\left[\frac{5 \cdot 16{,}4 - 3a,}{16{,}4 - a}\right] + 1 - 3\right)a, \infty\right) = \left[\left(\left[\frac{82 - 3a}{16{,}4 - a}\right] - 2\right)a, \infty\right)$$

$$t_{k0,9} \geq 300 \Rightarrow \left[\left(\left[\frac{82-3a}{16{,}4-a}\right]-2\right)a, \infty\right) \geq 300 \Rightarrow \left(\left[\frac{82-3a}{16{,}4-a}\right]-2\right)a \geq 300 \Rightarrow a \geq 14{,}76$$

De esta forma se establece que el tiempo entre sucesivas llegadas que optimiza este modelo de colas fuzzy es a=14,76 con lo que en el tiempo de duración del servicio, de 300 min., el sistema debe seguir una política de citar a 23 clientes.

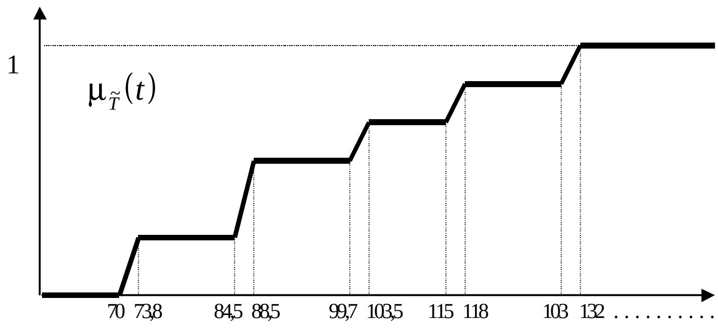
Sustituyendo en las fórmulas presentadas en la sección precedente, se pueden obtener las medidas que caracterizan a este sistema, y las medidas de efectividad del modelo con datos borrosos cuando el tiempo entre sucesivas llegadas es 14,76. Así:

- Nuevas llegadas al sistema antes de poder quedar el servidor ocioso:



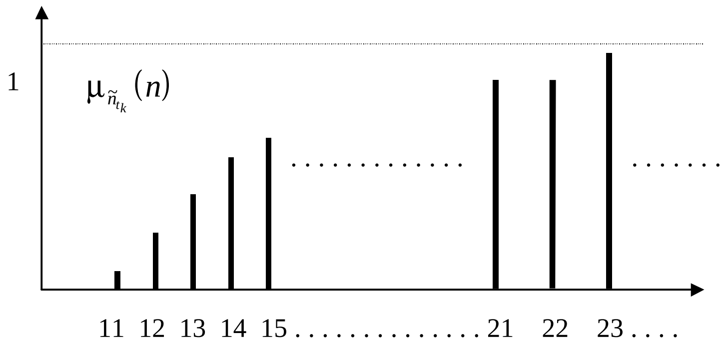
$$A_\alpha = \left[\left[\frac{20+4\alpha}{4{,}76-2\alpha}\right], 23\right)$$

- Instante de tiempo a partir del cual puede quedar el servidor ocioso:



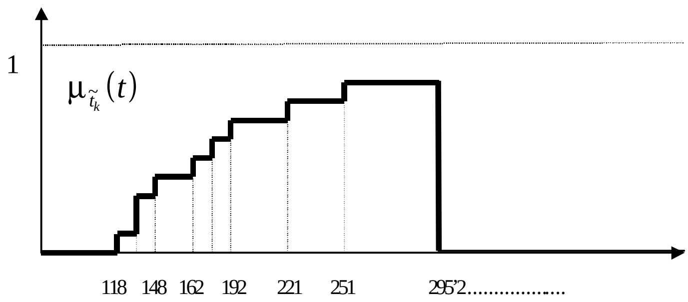
$$T_\alpha = \left[\left(3+\underline{A}_\alpha\right)(10+2\alpha), 460\right)$$

- Primer cliente que puede ser rechazado en el sistema:



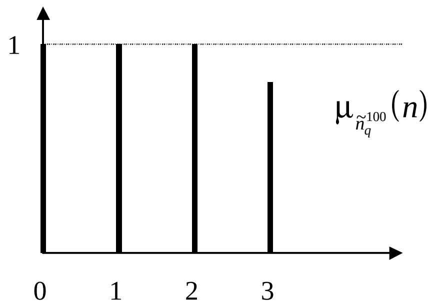
$$n_{t_k\,\alpha} = \left[\left[\frac{55,72-20\alpha}{5,24-4\alpha}\right]+1, 23\right)$$

- Instante de tiempo en el que se pueden rechazar clientes en el sistema:



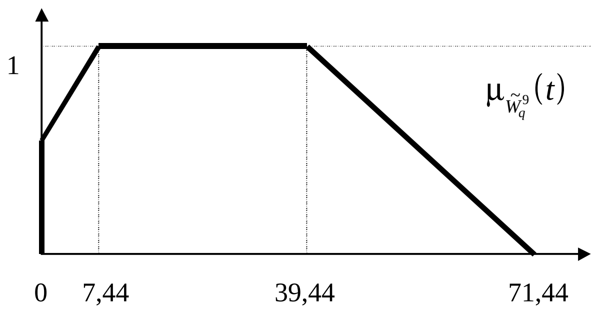
$$t_{k\alpha} = \left[\left(\underline{n}_{t_{k\alpha}}-3\right)14'76, 295'2\right)$$

- Número de clientes en cola en el minuto 100:



$$n_{q\alpha}^{100} = \begin{cases} \left[0, 8-\left[\dfrac{100}{20-4\alpha}\right]\right] & si \quad \alpha \le 0.555 \\ \left[8-\left[\dfrac{100}{10+2\alpha}\right], 8-\left[\dfrac{100}{20-4\alpha}\right]\right] & si \quad \alpha > 0.555 \end{cases}$$

- Tiempo de permanencia en cola del cliente 9:



$$W_{q\alpha}^{9}=\begin{cases}\left[0,\,71'44-32\alpha\right] & si \quad \alpha\le 0'535\\ \left[16\alpha-8'56,\,71'44-32\alpha\right] & si \quad \alpha>0'535\end{cases}$$

5. Simulación.

Con el uso del programa *Arena* se ha efectuado un experimento de simulación del modelo. La variable tiempo de servicio, $\tilde{b}=[10,12,16,20]$ se simula mediante la generación de una variable aleatoria uniforme sobre cada intervalo del tiempo de servicio, $\left[\underline{b}_\alpha,\overline{b}_\alpha\right]$ para los α-cortes de nivel: 0, 0’25, 0’5, 0’75 y 1. Se han realizado 2.000 réplicas del modelo. Los resultados obtenidos son:

- Nuevas llegadas al sistema antes de poder quedar el servidor ocioso:

α	Valor medio	Valor mínimo	Valor máximo
0	22,916	7	23
0,25	22,943	13	23
0,5	22,964	12	23
0,75	22,965	14	23
1,00	22,944	16	23

- Instante de tiempo en el que se pueden rechazar clientes en el sistema:

α	Valor medio	Valor mínimo	Valor máximo
0	298,75	162,36	300
0,25	299,95	250,92	300
0,5	300,00	300,00	300
0,75	300,00	300,00	300

1,00	300,00	300,00	300

- Número de clientes en cola en el minuto 100:

α	Valor medio	Valor mínimo	Valor máximo
0	1,6015	0	3
0,25	1,4760	0	3
0,5	1,3465	0	2
0,75	1,1900	0	2
1,00	1,0375	0	2

- Tiempo de permanencia en cola del cliente 9:

α	Valor medio	Valor mínimo	Valor máximo
0	31,536	7,182	56,786
0,25	29,368	8,821	53,128
0,5	27,442	8,876	45,217
0,75	25,458	11,639	40,422
1,00	23,375	13,431	33,840

La solución analítica obtenida en la sección anterior es comparada de forma favorable con los resultados de la simulación.

6. Bibliografía.

[1] Chanas, S.; Nowakowski, M. (1988) "From fuzzy data to a single action - A simulation approach". *Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision Making*. Eds. Kacprzyk, J. y Fedrizzi, M. Lecture Notes in Economics and Mathematical Systems. Vol. 310, pp 331-341. Ed. Springer-Verlag, Berlín.

[2] Gross, D.; Harris, C.M. (1985) *Fundamentals of Queueing Theory*. Ed. John Wiley and Sons, New York.

[3] Kaufmann, A.; Gil Aluja, J. (1987) *Técnicas operativas de Gestión para el tratamiento de la Incertidumbre*. Ed. Hispano-Europea, S.A., Barcelona.

[4] Li, R.J.; Lee, E.S. (1989) "Analysis of Fuzzy Queues". *Computers and Mathematics with Applications.* Vol. 17, No. 7, pp 1143-1147.

[5] Negi, D.S.; Lee, E.S. (1992) "Analysis and simulation of fuzzy queues". *Fuzzy Sets and Systems*. No. 46, pp 321-330.

[6] Pardo, M. J.; Castro, B. (1995) "Modelo de colas con tiempos de servicio fuzzy y llegadas deterministas". *II Congreso Internacional de Gestión y Economía Fuzzy. SIGEF*. Santiago de Compostela.

[7] Saaty, T.L. (1967) *Elementos de la Teoría de Colas*. Ed. Aguilar, Madrid.