

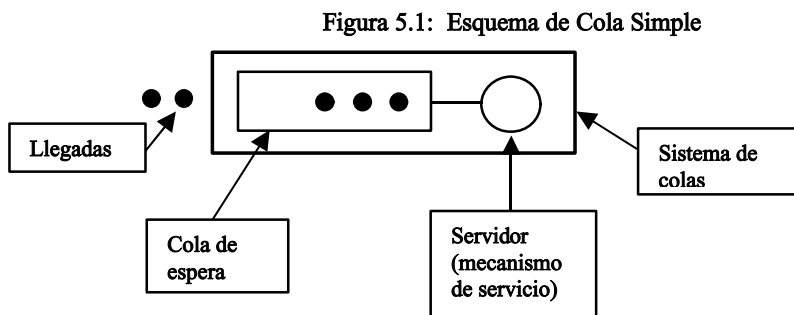
5 Gestión de Colas

En la mayoría de las organizaciones existen ejemplos de procesos que generan colas de espera. Estas colas suelen aparecer cuando un usuario, un empleado, una máquina o una unidad tiene que esperar a ser servidas debido a que la unidad de servicio, operando a plena capacidad, no puede atender temporalmente a este servicio. Un típico ejemplo de colas de espera que ilustra el problema es un viaje en avión. Primero, para comprar el billete podemos tener que hacer cola en la ventanilla correspondiente. Una vez obtenido el billete, tendremos que hacer cola para facturar el equipaje y obtener las tarjetas de embarque. Después hacemos cola para pasar por el detector de metales y finalmente esperamos en cola en la sala de embarque. Una vez dentro del avión, tendremos que esperar a que los pasajeros coloquen sus bolsas de mano para poder llegar a nuestro asiento. Cuando el avión se dirige hacia la pista de despegue puede encontrar con una cola de aviones esperando su turno para despegar. Cuando llega a su destino, puede dar unas cuantas vueltas antes de tener permiso para aterrizar. Y finalmente, cuando se asigna una puerta de desembarque para el avión, tendremos que esperar a que lleguen las maletas. En este viaje, es posible que hayamos sido miembros de por lo menos diez colas. Y eso sin considerar la experiencia en colas de la propia compañía aérea para este mismo viaje. El avión en el cual viajábamos tiene que esperar en cola para repostar, ser inspeccionado, asignarle una puerta determinada, una tripulación, una carga de comidas, una ruta específica, etc. De ahí que las compañías aéreas se preocupan de gestionar sus operaciones lo más eficientemente posible, y tratar de reducir al mínimo el tiempo de espera en realizar dichas operaciones.

Los sistemas sanitarios también se enfrentan a este tipo de problemas. Las listas de espera son muy comunes en muchos procesos quirúrgicos dentro de una red sanitaria, y a nivel ambulatorial es muy común la existencia de personas esperando a ser atendidas en un Centro de Asistencia Primaria. Los sistemas de urgencias muchas veces se ven congestionados siendo el tiempo de espera crucial. Los modelos de gestión de colas intentan simular el sistema en donde puede existir congestión (y por lo tanto, colas) y generan una serie de parámetros –que veremos en este capítulo– que permiten evaluar el sistema actual y evaluar la realización de modificaciones en el servicio en cuestión.

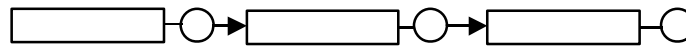
5.1 Descripción de un sistema de colas

Un sistema de colas tiene dos componentes básicos: la cola y el mecanismo de servicio. En la figura 5.1 se presenta un esquema de una cola simple.

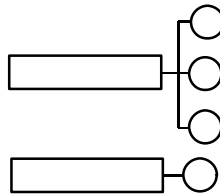


Pueden existir varias configuraciones de colas más complejas. En la Figura 5.2 se exponen otros tipos de configuraciones de sistemas de colas.

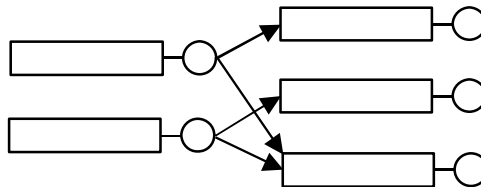
Figura 5.2: Configuraciones de colas



Sistema de colas en paralelo



Sistema de colas en serie



Sistema complejo de colas

El proceso básico en la mayoría de los sistemas de colas es el siguiente. Los clientes que vienen a procurar un determinado servicio se generan a través del tiempo en una fuente de entrada. Estos clientes entran dentro del sistema y se unen a una cola. En un determinado momento, se selecciona uno de los clientes para poder proporcionarle el servicio en cuestión, mediante lo que se denomina la *disciplina de servicio*. Esta disciplina es la que rige el mecanismo de atención. Una vez seleccionado el cliente, este es atendido por el *mecanismo de servicio*. Una vez terminado el servicio, el cliente sale del sistema.

En general, un sistema de colas tiene una *población potencial* infinita. Es decir, que el tamaño de la cola es muy pequeño respecto al potencial de usuarios del sistema. Por ejemplo, un ambulatorio de urgencias en general cubre una región con población grande comparado con las posibles urgencias que se puedan generar. Ahora bien, existen casos en donde la población es finita respecto del tamaño de la cola. Esto puede suceder en la farmacia de un hospital, en donde la población potencial la forma las enfermeras y ATS. En un momento dado puede formarse una cola considerable. Como los cálculos son mucho más sencillos para el caso infinito, esta suposición se emplea casi siempre.

Otro factor a tener en cuenta es el patrón estadístico mediante el cual se generan los clientes a través del tiempo. La suposición normal es que el proceso se genere siguiendo un *proceso de Poisson*, que veremos más adelante en este capítulo. Si el proceso de llegada es Poisson, el tiempo entre cada una de las llegadas sigue una *distribución exponencial*.

Otro factor importante a tener en cuenta en un sistema de colas es la *"fuga"* de algún cliente. Al modelizar la cola hay que considerar si una persona que lleva dentro de la cola un rato, desiste de ser atendida, cansada de esperar, abandonando la cola.

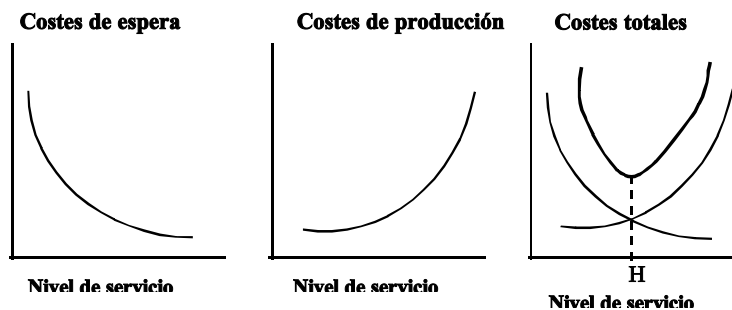
Como hemos mencionado anteriormente, la disciplina de la cola rige el sistema de entrada en el mecanismo de servicio. La mayoría de los sistemas utiliza el método “First In First Out”, conocido como FIFO. Otros sistemas pueden ser de tipo aleatorio, o de acuerdo con un sistema de prioridad previamente establecido.

El mecanismo de servicio consiste en una o más instalaciones de servicio, con cada una de ellas con uno o más canales de servicios, llamados *servidores*. Los clientes son atendidos en estos servidores. El tiempo que transcurre desde el inicio del servicio para un cliente hasta su terminación se llama el *tiempo de servicio* (o duración del servicio). Un modelo de sistema de colas tiene que especificar la distribución de probabilidad de los tiempos de servicio de cada servidor (y tal vez para distintos tipos de clientes), aunque normalmente se supone la misma distribución para todos los servidores. Una vez más, la distribución exponencial es la más empleada en los tiempos de servicio.

5.2 Objetivos de la gestión de colas

En los modelos de colas existen dos objetivos: por un lado la minimización del tiempo de espera y por el otro la minimización de los costes totales de funcionamiento del sistema. Estos objetivos suelen ser conflictivos, ya que para reducir el tiempo de espera se necesitan poner más recursos en el sistema, con el consiguiente aumento de los costes de producción. En muchos casos el tiempo de espera es difícil de determinar, sobretodo cuando se trata de un sistema en donde seres humanos están implicados. En la Figura 5.3 podemos ver la disyuntiva entre el coste de espera y el coste de producción.

Figura 5.3: Costes de un sistema de colas



Si pudiéramos sumar ambos costes, el coste total alcanzaría su mínimo en el punto H. En este punto el nivel de servicio es óptimo. Sin embargo, en muchos casos la obtención “objetiva” de este resultado puede ser muy complicada ya que, como se ha indicado anteriormente, la cuantificación del tiempo de espera en valores monetarios puede ser harto complicada y subjetiva.

Por lo tanto, en general se intenta llegar a una solución que sea “lógica” en función de los valores que adopten los diferentes parámetros del modelo. En la sección siguiente se examinan estos parámetros.

5.3 Medidas del sistema

Existen dos tipos de medidas para poder valorar un sistema en donde pueden aparecer colas: medidas “duras” y medidas “blandas”. Estas últimas están relacionadas con la calidad del servicio. Por ejemplo, no es lo mismo esperar 15 minutos de pie haciendo cola en un ambulatorio sin refrigeración y poco ventilado que esperar el mismo tiempo en una sala de espera con butacas confortables, revistas, aire acondicionado y música clásica de fondo. El paciente valorará mucho más 1 minuto de espera en el primer caso ya que representa un coste mucho más elevado en términos de confort. En otras palabras, seguramente un minuto de cola en el ambulatorio equivale a muchos minutos de espera en la sala de espera confortable. La gestión cuantitativa de las colas no se ocupa de estos aspectos cualitativos (que no por ello dejan de ser importantes) sino que da valores a una serie de medidas “frías” o “duras”. Las medidas duras más utilizadas en los modelos de gestión de colas y su notación estándar son las siguientes:

- Tasa media de llegada, $\bar{\lambda}$
- Tasa media de servicio, μ
- Tiempo medio de espera en la cola, W_q
- Tiempo medio de estancia en el sistema, W_s
- Número medio de personas en la cola, L_q
- Número medio de personas en el sistema, W_s
- Porcentaje de ocupación de los servidores, P_w
- Probabilidad de que hayan x personas en la en el sistema, P_x

En los siguientes apartados iremos examinando estos conceptos.

5.4 Un sistema de colas elemental: tasa de llegada y de servicio constantes

Supongamos que tenemos un sistema en donde tanto la tasa de llegada (en personas por unidad de tiempo) como el tiempo de servicio son constantes. En este caso, podemos tener las tres situaciones siguientes:

5.4.1 No hay cola, tiempo ocioso del servidor

Supongamos que tenemos un sistema en donde cada 6 minutos, exactamente, llega una persona a un ambulatorio. O, en otras palabras, la tasa de llegada es exactamente de 10 personas por hora. Supongamos que la tasa de servicio del médico (del servidor en términos técnicos) es de 12 personas por hora siempre, ni una más ni una menos. En esta situación nunca se formará una cola porque el servidor puede manejar perfectamente las llegadas. Incluso ya sabemos que el servidor estará ocioso un 16,6% de su tiempo, ya las llegadas necesitan únicamente de 10/12, o 83,33% de la capacidad de servicio.

5.4.2 No hay cola ni tiempo ocioso del servidor.

Siguiendo el ejemplo anterior, supongamos que la tasa de servicio pasa a ser igual a 10 personas por hora, es decir, exactamente igual que la tasa de llegada. En esta situación es imposible que se forme una cola, pero por otro lado el servidor estará ocupado 100% de su tiempo y trabajará a plena capacidad.

5.4.3 Formación de cola y sin tiempo ocioso en el servidor

Ahora supongamos que la tasa de servicio pasa a ser igual a 8 personas por hora, mientras que siguen llegando pacientes cada 6 minutos exactamente. En esta situación se formará una cola que irá creciendo, ya que el servidor no puede absorber toda la demanda de servicio y los pacientes se irán acumulando. La cola de llegadas no servidas inmediatamente irá creciendo a una tasa de 2 personas por hora, es decir el exceso de llegadas partido por las personas servidas. Por ejemplo, al cabo de ocho horas, tendríamos 16 personas en la cola.

El hecho de que hayamos asumido unas tasas de llegada y de servicio constantes hasta ahora facilita los cálculos para obtener información sobre el sistema. Pero la situación se complica si nos trasladamos a la situación más realista, en donde las tasas de llegada y de servicio no son constantes, sino que siguen una determinada distribución probabilística. Por ejemplo, si las llegadas y los tiempos de servicio estuviesen distribuidos aleatoriamente a lo largo de la jornada, aunque la capacidad de los servidores sea suficiente para absorber la demanda, puede pasar que un grupo de pacientes llegue en bloque y formen durante un tiempo una cola. Y, por otro lado, si durante un tiempo no llegan más pacientes, la cola puede ser reducida por el mecanismo de servicio. En las siguientes secciones examinaremos algunos de estos casos.

5.5 Las distribuciones de Poisson y Exponencial

5.5.1 La distribución de Poisson

Esta distribución es muy frecuente en los problemas relacionados con la investigación operativa, sobre todo en el área de la gestión de colas. Suele describir, por ejemplo, la llegada de pacientes a un ambulatorio, las llamadas a una centralita telefónica, la llegada de coches a un túnel de lavado, el número de accidentes en un cruce, etc. Todos estos ejemplos tienen un punto en común: todos ellos pueden ser descritos por una variable aleatoria discreta que tiene valores no-negativos enteros (0,1,2,3,4...). El número de pacientes que llegan al ambulatorio en un intervalo de 15 minutos puede ser igual, a 0, 1, 2 3...

Sigamos con el ejemplo del ambulatorio. La llegada de pacientes se puede caracterizar de la forma siguiente:

1. El número medio de llegadas de los pacientes para cada intervalo de 15 minutos puede ser obtenido a través de datos históricos.
2. Si dividimos el intervalo de 15 minutos en intervalos mucho más pequeños (por ejemplo, 1 segundo), podemos afirmar que:
 - 2.1 La probabilidad de que exactamente un único paciente llegue al ambulatorio por segundo es tiene un valor muy reducido y es constante para cada intervalo de 1 segundo.
 - 2.2 La probabilidad de que 2 o más pacientes lleguen dentro del intervalo de 1 segundo es tan pequeña que podemos decir que es igual a 0.
 - 2.3 El número de pacientes que llegan durante el intervalo de 1 segundo es independiente de donde se sitúa este intervalo dentro del periodo de 15 minutos.
 - 2.4 El número de pacientes que llegan en un intervalo de 1 segundo no depende las llegadas que han sucedido en otro intervalo de 1 segundo

Si al analizar un proceso de llegada este cumple estas condiciones, podemos afirmar que su distribución es de Poisson.

La fórmula para obtener la probabilidad de que un evento ocurra (que lleguen 3 pacientes, por ejemplo) es la siguiente:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

En donde x representa en número de llegadas, λ la tasa media de llegadas y $P(x)$ la probabilidad de que el número de llegadas sea igual a x .

5.5.2 La distribución Exponencial

Mientras que la distribución de Poisson describe las llegadas por unidad de tiempo, la distribución exponencial estudia el tiempo entre cada una de estas llegadas. Si las llegadas son de Poisson, el tiempo entre ellas es exponencial. Mientras que la distribución de Poisson es discreta, la distribución exponencial es continua, porque el tiempo entre llegadas no tiene por qué ser un número entero.

Esta distribución se utiliza mucho para describir el tiempo entre eventos, más específicamente, la variable aleatoria que representa el tiempo necesario para servir a la llegada. Ejemplos típicos de esta situación son el tiempo que un médico dedica a una exploración, el tiempo de servir una medicina en una farmacia, o el tiempo de atender a una urgencia.

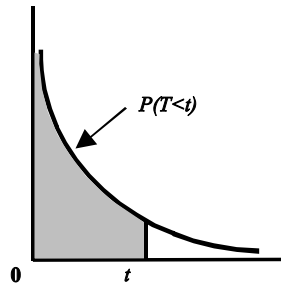
El uso de la distribución exponencial supone que los tiempos de servicio son aleatorios, es decir, que un tiempo de servicio determinado no depende de otro servicio realizado anteriormente, ni de la posible cola que pueda estar formándose. Otra característica de este tipo de distribuciones es que no tienen “edad”, o en otras palabras, “memoria”. Por ejemplo, supongamos que el tiempo de atención de un paciente en una sala quirúrgica sigue una distribución exponencial. Si el paciente ya lleva 5 horas siendo operado, la probabilidad de que esté una hora más es la misma que si hubiera estado 2 horas, o 10 horas o las que sea. Esto es debido a que la distribución exponencial supone que los tiempos de servicio tienen una gran variabilidad. A lo mejor el próximo paciente operado tarda 1 hora porque su cirugía era mucho más simple que la del anterior.

La función de densidad de la distribución exponencial es la siguiente:

$$p(t) = \mu e^{-\mu t}$$

En donde t representa el tiempo de servicio y μ la tasa media de servicio (pacientes servidos por unidad de tiempo). La densidad exponencial se presenta en Figura 5.4. En general nos interesará encontrar $P(T < t)$, la probabilidad de que el tiempo de servicio T sea inferior o igual a un valor específico t . Este valor es igual al área por debajo de la función de densidad.

Figura 5.4: Distribución Exponencial



Si, por ejemplo, queremos saber cual es la probabilidad de que el tiempo de servicio sea de 2 o menos horas cuando el tiempo medio es de 3 horas (una tasa de servicio de 1/3), podemos aplicar la fórmula siguiente:

$$p(T \leq t) = 1 - e^{-\mu t}$$

En este caso, $P(T \leq 2) = 0,486$, casi un 50% de probabilidad.

5.6 Modelo de colas simple: Llegadas en Poisson y tiempos de servicio exponencialmente distribuidos.

El modelo que presentaremos a continuación tiene que cumplir las condiciones siguientes:

1. El número de llegadas por unidad de tiempo sigue una distribución de Poisson
2. Los tiempos de servicio siguen una distribución exponencial
3. La disciplina de la cola es de tipo FIFO
4. La población potencial es infinita
5. Existe un único canal de servicio
6. La tasa media de llegadas es menor que la tasa media del servicio
7. El tamaño potencial de la cola es infinito

Si estas condiciones se cumplen y si conocemos la tasa media de llegada λ , y la tasa media de servicio μ , las ecuaciones para obtener valores de las medidas descritas anteriormente son:

- Número medio en la cola:

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

- Número medio en el sistema:

$$L_s = \frac{\lambda}{(\mu - \lambda)}$$

- Tiempo medio de espera en la cola:

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

- Tiempo medio en el sistema:

$$W_s = \frac{1}{(\mu - \lambda)}$$

- Factor de utilización:

$$P_w = \frac{\lambda}{\mu}$$

Un ejemplo

Consideremos el caso de un gran laboratorio farmacéutico que tiene en su almacén un único estacionamiento de carga de que sirve a todas las farmacias de una región, y existe un único trabajador para buscar los medicamentos del pedido de cada furgoneta y cargarlos en ella. Se observa que de vez en cuando las furgonetas de transporte se acumulan en el estacionamiento formando cola, y de vez en cuando el trabajador está ocioso. Después de un estudio del sistema observamos que éste cumple las condiciones expuestas anteriormente. Después de examinar las llegadas de las camionetas durante varias semanas, se determina que la tasa media de llegada es de 4 camionetas por hora, y que la tasa de servicio es de 6 camionetas por hora. Los gestores del almacén están considerando el añadir un trabajador adicional, o incluso dos de ellos, para aumentar la tasa de servicio. El problema consiste en evaluar estas opciones diferentes.

Si se añade un trabajador, el sistema seguirá siendo de cola simple, porque una única camioneta puede cargarse a la vez. Si usamos dos trabajadores, la tasa de servicio será igual a 12. Si utilizamos tres trabajadores, la tasa de servicio será igual a 18.

En el Cuadro 5.1 se han utilizado las ecuaciones expuestas anteriormente para obtener las medidas de eficiencia del sistema. Hemos supuesto que la capacidad de trabajo es proporcional al número de trabajadores.

Cuadro 5.1: Resultados del modelo simple de colas

		Trabajadores		
		1	2	3
Número medio de camionetas en la cola	L_q	1,333	0,167	0,063
Número medio de camionetas en el sistema	L_s	2,000	0,500	0,286
Tiempo medio de la camioneta en cola	W_q	0,333	0,042	0,016
Tiempo medio de la camioneta en el sistema	W_s	0,500	0,125	0,071
Ocupación del servicio	P_w	0,667	0,333	0,222

Supongamos que los costes de operación de cada camioneta por hora son de 2000 ptas y los trabajadores cobran 1800 ptas por hora de trabajo y que estos trabajan 8 horas al día. En el Cuadro 5.2 se presentan los costes asociados. Al interpretar los tiempos, hay que ir con cuidado ya que estos están en fracciones de hora.

Cuadro 5.2: Costes de operación del sistema

Trabajadores	Coste de Camioneta por día	Coste de mano de obra por día	Coste total Por día
1	320.000	144.000	464.000
2	80.000	288.000	368.000
3	46.000	432.000	478.000

Los gestores tendrían que añadir un nuevo trabajador al sistema ya que esto representará una reducción de los costes totales operacionales, aunque el factor de utilización pasará a ser de 33%. Es decir, que los dos trabajadores tendrán 5 horas y 20 minutos para dedicarse a otras tareas dentro del laboratorio farmacéutico.

Extensión del modelo simple a colas con capacidad limitada

Existen casos en los que el sistema (cola más servicio) tiene una cierta capacidad. Si un cliente llega cuando hay M o más personas en el sistema, el cliente se va inmediatamente y no vuelve. Este tipo de modelo es característico de los problemas de colas que se pueden encontrar en algunos servicios. Por ejemplo, un restaurante con un estacionamiento limitado. En este caso, las ecuaciones del modelo son:

- Probabilidad de 0 personas en el sistema:

$$P_0 = \frac{1 - (\lambda/\mu)}{1 - (\lambda/\mu)^{M+1}}$$

- Factor de utilización:

$$P_w = 1 - P_0$$

- Proporción de clientes perdidos porque el sistema está lleno:

$$P_M = (\lambda/\mu)^M P_0$$

- Número medio en el sistema:

$$L_s = \frac{P_w - M(\lambda/\mu)P_M}{1 - (\lambda/\mu)}$$

- Número medio en la cola:

$$L_q = L_s - \frac{\lambda(1 - P_M)}{\mu}$$

- Tiempo medio de espera en el sistema:

$$W_s = \frac{L_s}{\lambda(1-P_M)}$$

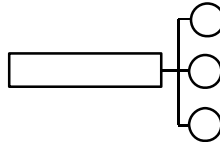
- Tiempo medio en la cola:

$$W_q = W_s - \frac{1}{\mu}$$

5.7 Modelo múltiple de colas: Llegadas en Poisson y tiempos de servicio exponencialmente distribuidos.

En muchos casos podemos tener situaciones en donde existe más de un servidor en el sistema. A medida que van llegando los clientes, los servidores se van ocupando y cada vez que uno de ellos acaba su servicio, el primero de la cola lo vuelve a ocupar. El sistema está representado en la Figura 5.5.

Figura 5.5: Sistema múltiple de colas



En este tipo de modelos la tasa de llegada siempre tiene que ser inferior a la tasa agregada de servicio, que no es más que la tasa de servicio individual multiplicada por el número de canales. En este modelo se supone, además de las condiciones expuestas anteriormente, que la tasa individual de cada canal es la misma. Las expresiones matemáticas para la obtención de las medidas de eficiencia del sistema dependen de P_0 que es la probabilidad de que no haya nadie en el sistema (cola más servicio). En el disquete (archivo colas.xls) se pueden calcular los valores de P_0 . Los valores de las medidas de eficiencia son función de P_0 y se obtienen a partir de las siguientes fórmulas:

- Factor de utilización:

$$P_w = \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k \frac{k\mu}{k\mu - \lambda} P_0$$

- Número medio en el sistema:

$$L_s = \frac{\lambda\mu(\lambda/\mu)^k}{(k-1)!(k\mu - \lambda)^2} P_0 + \frac{\lambda}{\mu}$$

- Número medio en la cola:

$$L_q = L_s - \frac{\lambda}{\mu}$$

- Tiempo medio de espera en el sistema:

$$W_s = \frac{L_s}{\lambda}$$

- Tiempo medio en la cola:

$$W_q = \frac{L_q}{\lambda}$$

Tenemos que tener presente que en estas ecuaciones λ representa la tasa de servicio por canal. En la página web (archivo colas.xls) se pueden calcular estos valores tanto para sistemas múltiples de colas con capacidad infinita como con capacidad limitada.

Un ejemplo

El ambulatorio de una región tiene dos médicos de cabecera que atienden a los pacientes que van llegando. En general los pacientes tienen que esperar a ser atendidos y la gerencia está estudiando la posibilidad de contratar un nuevo médico para aligerar el sistema. Como es muy difícil estimar en términos monetarios el coste de espera de los pacientes, la gerencia realizará la nueva contratación si se consiguen reducir los tiempos totales del servicio (espera más atención) a la mitad. Después de observar y recoger datos sobre las llegadas y sobre el tiempo de servicio, la gerencia calcula que en media llegan 8 pacientes por hora, y que cada uno de los médicos puede atender 5 pacientes por hora.

En el cuadro 5.3 se presentan los resultados después de aplicar las fórmulas del modelo con dos y tres médicos.

Cuadro 5.3: Resultados del modelo múltiple de colas

		Médicos	
		2	3
Probabilidad de que todos los médicos estén libres	P_0	0,111	0,190
Probabilidad de que todos los médicos estén ocupados	P_w	0,710	0,278
Número medio de pacientes en el sistema	L_s	4,442	1,918
Número medio de pacientes en cola	L_q	2,842	0,318
Tiempo medio de un paciente en el sistema	W_s	0,555	0,240
Tiempo medio de un paciente en cola	W_q	0,355	0,040

En el cuadro podemos observar que si añadimos un médico adicional el tiempo de espera de cada paciente en el sistema pasa de 0,555 horas a 0,240 horas. Por lo tanto, el objetivo de la gerencia se cumple al añadir un nuevo médico. También se puede observar que con tres médicos el tiempo de espera en la cola es insignificante.

5.8 Limitaciones de los modelos de gestión de colas

Los dos modelos que hemos presentado en este capítulo son los más comunes cuando se trata de sistemas en donde están implicados seres humanos. Sin embargo, pueden existir casos en donde la población potencial del sistema es finita, la cola de la disciplina no es FIFO, la tasa de servicio depende de las personas en la cola, y las distribuciones de las llegadas no son de Poisson. En estos casos estos modelos son inservibles.

Las distribuciones juegan un papel esencial en estos modelos. Los sistemas en donde las variaciones de las llegadas en diferentes horarios son muy grandes no pueden ser examinados con las formulaciones presentadas. Cuando tenemos sistemas más complejos se utiliza la simulación como método de análisis. En la siguiente sección simularemos un sistema de colas.

5.9 Ejemplo de simulación de un sistema de colas.

La farmacia de un hospital tiene dos personas para atender a 10 enfermeras que vienen a buscar medicamentos para los pacientes. La gerencia observa que de vez en cuando se forman colas para recoger las medicinas y que el servicio resulta un tanto ineficiente. Por otro lado, la población potencial es bastante reducida y no puede considerarse como infinita. Por otro lado la distribución del número de llegadas no es Poisson ni el tiempo de servicio exponencial. Por lo tanto no se puede aplicar el modelo múltiple de gestión de colas.

5.9.1 Recogida de datos

La gerencia observó el funcionamiento de la farmacia durante periodos de 1 hora distribuidos a lo largo de un mes. Estos periodos de una hora fueron aleatoriamente escogidos durante el día para obtener una muestra representativa de la actividad.

Los resultados de la observación se muestran en el Cuadro 5.4.

Cuadro 5.4: Resultados de la muestra

Duración del tiempo de servicio (minutos)	Número de Observaciones
8	15
9	30
10	45
11	60
Total de llegadas	150

Además, la gerencia dividió el tiempo de observación en intervalos de 5 minutos y anotó las llegadas de enfermeras que llegaron durante estos intervalos. Se observó que en promedio llegaba una enfermera cada 5 minutos. Al final del periodo de observación, la gerencia presentó los resultados obtenidos de la siguiente forma:

- Distribución porcentual de los tiempos de servicio:

$$\begin{aligned}
 15/150 &= 10\% \text{ (8 min)} \\
 30/150 &= 20\% \text{ (9 min)} \\
 45/150 &= 30\% \text{ (10 min)} \\
 60/150 &= 40\% \text{ (11 min)}
 \end{aligned}$$

- Media ponderada de los tiempos de servicio:

$$\begin{array}{rcl}
 10\% & 8 \text{ min} = & 0,8 \text{ min} \\
 20\% & 9 \text{ min} = & 1,8 \text{ min} \\
 30\% & 10 \text{ min} = & 3,0 \text{ min} \\
 40\% & 11 \text{ min} = & \underline{4,4 \text{ min}} \\
 \text{Tiempo medio de servicio:} & & 10,0 \text{ min}
 \end{array}$$

Con esta información, la gerencia pudo empezar a realizar la simulación con la ayuda de números aleatorios (ver anexo 2)

5.9.2 Simulación de llegadas.

En primer lugar la gerencia simula las llegadas de las enfermeras a la farmacia. Ésta sabe que las llegadas son aleatorias, aunque en promedio llega una cada cinco minutos. Como los números aleatorios tienen 10 dígitos (0,1,2,3,4,5,6,7,8,9,10), la gerencia escoge (aleatoriamente) el 7 como representativo de una llegada. Si cogemos al azar un número de la tabla, la cantidad de sietes que contenga el número indicará la cantidad de llegadas en un intervalo de 5 minutos. La gerencia simula las llegadas a la farmacia durante 24 periodos de 5 minutos. Quizás no sea una cantidad muy representativa en este caso, pero para efectos de explicación de la simulación es suficiente. En un caso real simularíamos el sistema con muchos más periodos, pero la mecánica seguiría siendo la misma.

Para ilustrar el procedimiento de simulación de llegadas, hemos escogido los 12 primeros números aleatorios del apéndice (por columna) y contado las veces que sale el número 7.

1239650125	0	6749281769	2	0178780337	3
1370937859	2	8912349495	0	9128374452	1
0926561938	0	9172674928	2	4412773934	2
1639438732	1	9916253764	0	0112378549	1

En el Cuadro 5.5 se muestran los resultados para los 24 periodos.

Cuadro 5.5: Simulación del número de llegadas en cada intervalo

Periodo	Cantidad de Llegadas	Periodo	Cantidad de llegadas
1	0	13	0
2	2	14	0
3	0	15	1
4	1	16	4
5	2	17	1
6	0	18	1
7	2	19	1
8	0	20	0
9	3	21	0
10	1	22	1
11	2	23	0
12	1	24	2

5.9.3 Simulación de los tiempos de servicio

Después de obtener los resultados de la simulación de las llegadas, la gerencia tiene que realizar la simulación de los tiempos de servicio. Recordemos la distribución de los tiempos de servicios observada anteriormente:

<u>Minutos</u>	<u>Porcentaje</u>
8	10
9	20
10	30
11	40

Como seguimos utilizando los números aleatorios de 10 dígitos, consideraremos que el 0 representa un tiempo de servicio de 8 minutos, el 1 y el 2 un tiempo de 9 minutos, el 3, 4 y el 5 un tiempo de 10 minutos, y el 6, 7, 8 y 9 un tiempo de 11 minutos. De esta forma podemos representar exactamente la probabilidad de los tiempos de llegada. Por ejemplo, observamos en el Cuadro 5.5 que en el segundo periodo hubieron dos llegadas al servicio. Para simular el tiempo de servicio, escogemos la última fila de números aleatorios de la tabla comenzando por la izquierda. El primer número es 9 y el segundo 8. Esto quiere decir que la primera y la segunda llegada tendrán asociadas un tiempo de atención de 11 minutos. Este proceso se repite para todos los periodos en los que hay llegadas. El resultado se presenta en el Cuadro 5.6.

Cuadro 5.6: Resultados de la simulación de los tiempos de servicio

Número de periodo	Número de Llegadas	Tiempo de servicio de cada una
1	0	
2	2	(1) 11min, (2) 11 min
3	0	
4	1	(3) 10 min
5	2	(4) 11 min, (5) 10 min
6	0	
7	2	(6) 9 min, (7) 10 min
8	0	
9	3	(8) 10 min, (9) 10 min, (10) 11 min
10	1	(11) 11 min
11	2	(12) 10 min, (13) 8 min
12	1	(14) 11 min
13	0	
14	0	
15	1	(15) 11 min
16	4	(16) 10 min, (17) 11 min, (18) 11 min, (19) 11 min
17	1	(20) 9 min
18	1	(21) 8 min
19	1	(22) 11 min
20	0	
21	0	
22	1	(23) 11 min
23	0	
24	2	(24) 9 min, (25) 11 min

5.9.4 Simulación conjunta del sistema

Ahora ya podemos simular el sistema. El objetivo de la gerencia es encontrar el número óptimo de trabajadores en la farmacia de forma a minimizar el coste total del servicio. Se considera que el sistema es FIFO, es decir, primer en llegar, primero en ser atendido. Ahora se tienen que definir los criterios de cada llegada dentro de cada periodo. Se definen los criterios siguientes:

1. Si en un periodo llega una única enfermera, ésta lo hará al principio del periodo
2. Si en un periodo llegan 2 enfermeras, la primera lo hace al principio y la segunda en el tercer minuto
3. Si en un periodo llegan 3 enfermeras, la primera llega al principio, la segunda en el minuto 3, y la tercera en el minuto 5
4. Si en un periodo llegan 4 enfermeras, se asume que llegan en los minutos 2, 3, 4 y 5 respectivamente.

En la Figura 5.6 se presenta el patrón de llegadas de los 24 periodos.

Cada llegada viene indicada por un cuadro conteniendo su correspondiente número de orden, y encima del recuadro se indica el tiempo de espera correspondiente. Por ejemplo, si examinamos el periodo entre las 10:15 y las 10:20, vemos que se producen 5 llegadas (de la 16 a la 20), y, como veremos más adelante, seguramente se producirá una cola considerable. El comportamiento del sistema simulado con dos trabajadores atendiendo a los clientes (nivel de congestión, personas en cola, duración del tiempo de espera) puede representarse tal como se muestra en la Figura 5.7. En la Figura 5.8 se representa la simulación con tres trabajadores. Si se comparan las dos figuras, visualmente se puede observar que el tiempo de espera se reduce considerablemente.

Si se examina desde un punto de vista económico, con dos trabajadores atendiendo a las enfermeras, éstas esperarían un total de 213 minutos, un tiempo medio de espera igual a 8,52 minutos. Para obtener un valor monetario del tiempo de espera, la gerencia considera que el coste por hora de cada trabajador es igual a 7 euros y el coste de cada enfermera es de 12 euros. Si recordamos que el tiempo medio entre cada llegada era de 5 minutos, en media las enfermeras realizarían 96 viajes por día (8 horas diarias por 12 viajes por hora). Y si el tiempo medio de espera es de 8,52 minutos por viaje, el tiempo total de espera es igual a 817,9 minutos (13,63 horas perdidas), lo que representa un coste total de espera de 163,56 euros. Si añadimos el coste de los dos trabajadores (112 euros), el coste diario total es igual a 275,56 euros.

Si realizamos el mismo ejercicio pero modificando el número de trabajadores atendiendo a las enfermeras (Figura 7.6) el tiempo total de espera es igual a 47 minutos, o 1,88 minutos de espera por llegada. Si tenemos 96 llegadas por día, el tiempo total perdido es igual a 180,48 minutos (3 horas diarias). El coste de la espera es igual a 36 euros y el sueldo de los trabajadores igual a 168 euros, lo que da un coste total igual a 204 euros. Por lo tanto, la operación con tres trabajadores parece ser más eficiente en términos monetarios. ¿Qué pasaría si contratáramos un cuarto trabajador que eliminaría completamente el tiempo de espera de las enfermeras? En este caso el único coste sería el sueldo de los trabajadores, que sería igual a 224 euros, superior al coste con tres trabajadores.

Figura 5.6: Representación de las llegadas

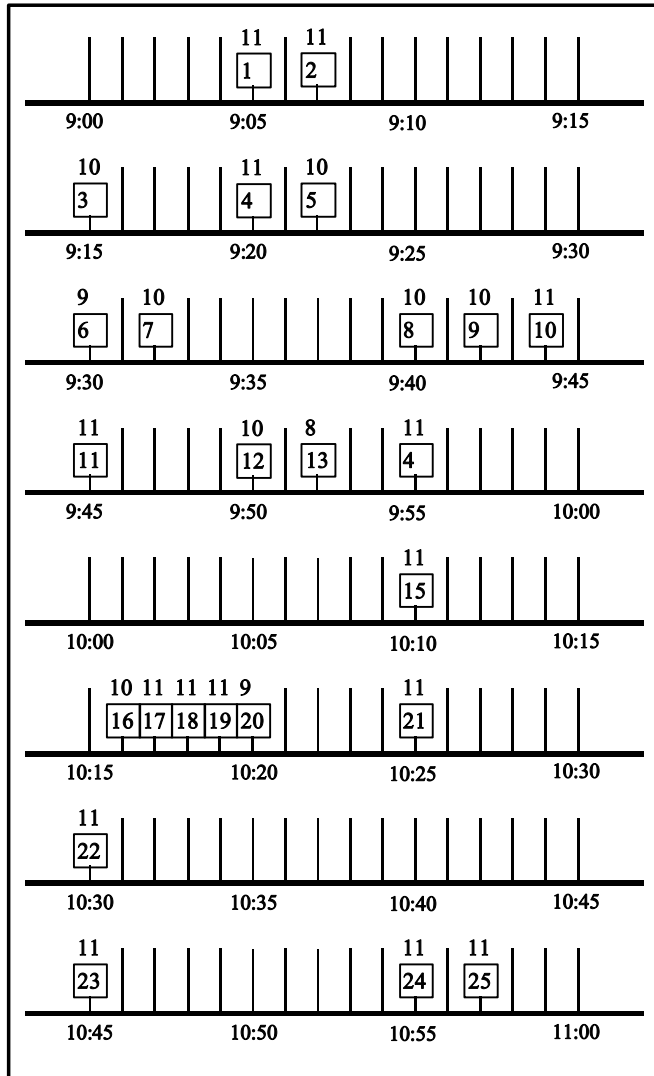


Figura 5.7: Operación de la farmacia con dos trabajadores

← Duración del servicio — Tiempo de espera 1 Llegada

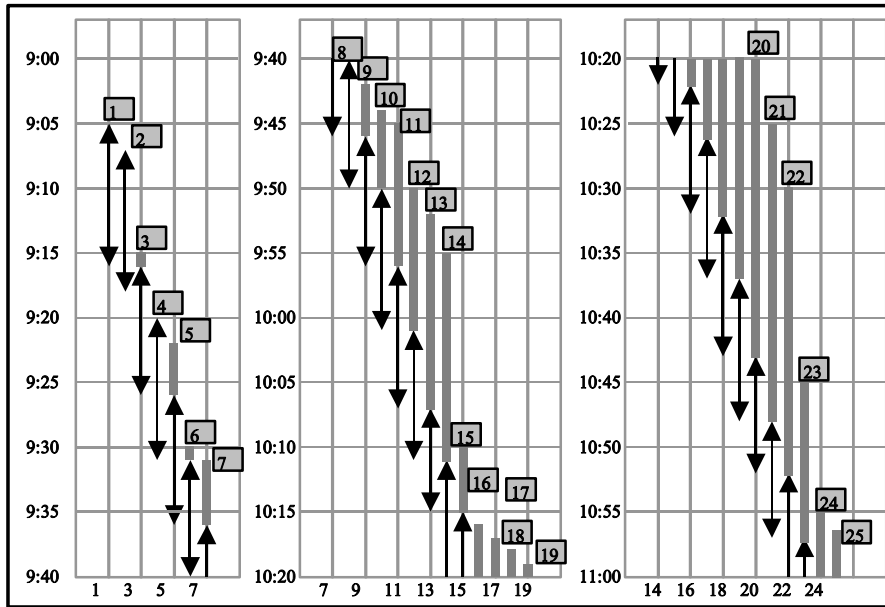
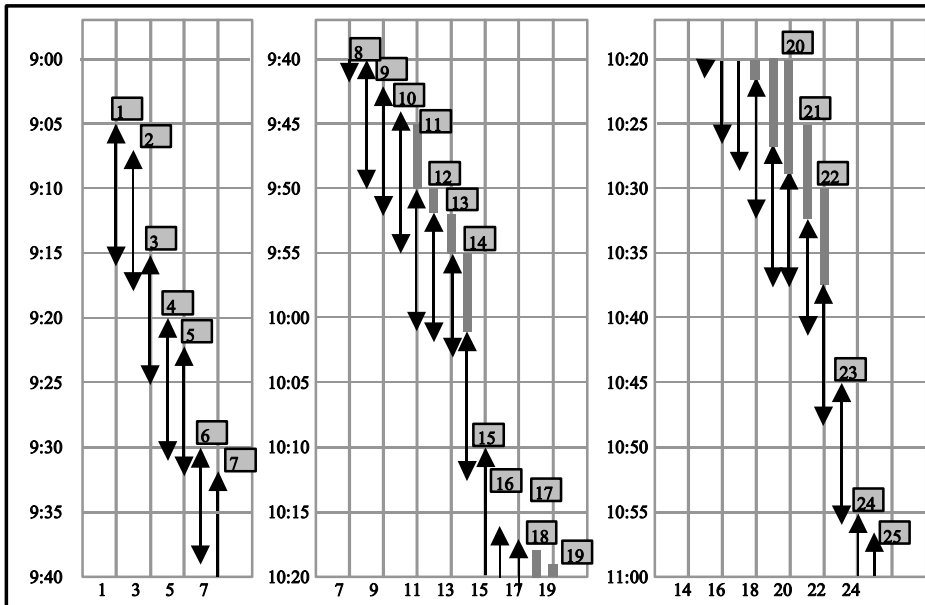


Figura 5.8: Operación de la farmacia con tres trabajadores



5.10 Problemas

5.1. En un ambulatorio se reciben una media de tres pacientes por hora, siguiendo una distribución de Poisson. El único médico que está en el ambulatorio atiende una media de 6 pacientes por hora. Los tiempos de atención siguen una distribución exponencial. La pregunta que se plantea el gestor es si vale la pena contratar a un nuevo médico o no (considerando que éste realizará el mismo número de pacientes).

5.2. La llegada de enfermeras a una farmacia del hospital Todosalud puede describirse a través de una distribución de Poisson. Los tiempos de servicio utilizan una distribución exponencial. La tasa de llegada es de 45 enfermeras por hora, mientras que cada farmacéutico puede atender a 50 enfermeras por hora. El coste de cada enfermera es de 15 euros por hora, mientras que cada farmacéutico gana 10 euros por hora. Encontrar el número óptimo de farmacéuticos a contratar.

5.3. Un centro de atención primaria tiene que administrar la vacuna de poliomelitis a los niños de un barrio. El centro está organizado de forma que los padres van llegando con los niños, formando una cola y siendo atendidos 40 por hora, con una distribución exponencial, por cualquiera de las enfermeras que están de servicio. Este servicio de vacunación se ofrece una vez a la semana, y en este día las llegadas se realizan con una tasa igual a 40 niños por hora. El director del centro sabe que la mayoría de los padres vienen durante sus horas de trabajo y por ello quiere limitar el tiempo total de administración de la vacuna a 15 minutos (incluyendo la espera). ¿Cuántas enfermeras tendrá que utilizar el gerente?

5.4. Un fisioterapeuta tiene un pequeño consultorio en donde trabaja con pacientes operados de rodilla. Son en media 8 los pacientes que llegan por las tardes siguiendo una distribución de Poisson. El tiempo de servicio del fisioterapeuta sigue una distribución exponencial con una tasa media de servicio de 9 pacientes por hora. Por las mañanas no está nada ocupado, porque la mayoría de pacientes prefiere venir por las tardes. Después de varios años de experiencia, el fisioterapeuta sabe que, en general, si un paciente llega a su clínica y hay tres otros esperando a ser atendidos, éste cambiará de fisioterapeuta. Para intentar convencer a algunos de sus pacientes a venir por la mañana, decide dar vales para almorzar en el bar de la esquina. Con esta táctica, cree que podrá reducir la tasa de llegadas de la tarde a 5 pacientes por hora. Si la pérdida de un paciente significa una pérdida de 20 euros, ¿cuánto estaría dispuesto a pagar por cada vale considerando que cada mañana y cada tarde es de 4 horas?

Apéndice 5.1: Tabla de números aleatorios

1239650125	1142160902	4527058366	1430875177	2821660618	8377041005
1370937859	8611794741	9815467859	9902648765	8411668944	7446544112
0926561938	1497952703	8609672743	4339955985	7272012830	1644676720
1639438732	2405306350	7914214780	4677121705	7791648017	8176854669
6749281769	7502962776	9671479586	9093449703	5368210111	4339625930
8912349495	7679203454	9981241957	4032184023	4683141252	2008225353
9172674928	7759918425	7230206493	2581236968	3177342357	9568401934
9916253764	2363211025	6282095638	3001769692	8059067287	4185944112
0178780337	8652658498	7424574963	5185675861	9919589016	7193770252
9128374452	9615307086	3542787227	6735835540	6005786976	1738020959
4412773934	5251827422	3399383961	9764534919	7108528926	2628877567
0112378549	2562566485	6582696326	1089265935	7552101346	6433830583
6484118917	4127133914	3687936583	8363383157	9380150844	2984673476
7485396877	1398185872	1298594465	5141064723	5658307618	3659683979
4449420399	3106053098	5647103564	4730042272	9397192556	6329548442
4931720558	5644540248	8583519254	5956489316	6647190158	6024106022
7019256915	8735782233	9315124105	6211324018	2835893430	3460537007
7088642561	3802788585	4015059652	3577040333	5596749322	6100976968
1322899809	8411711735	3383210443	1558138811	8636254576	9483732238
7644596455	7727801919	2425499018	9548164500	3705900236	1713960284
4350250689	5583998402	8094153554	7641705754	1338635552	7597030515
3969625810	7127183350	4809003690	2376632426	7457842856	6287196405
9867466169	3959776552	8642913942	3395147232	1594796171	6959239675
2336529609	8083742448	1499051196	8052499450	8186045265	1772149893
8497899420	6831646060	7649824834	3614066579	5186204268	3416788948
1383109138	8248598021	5880393712	1869970313	3950714593	7974854777
9755836769	8902846338	9965787689	7335817518	7384119119	3462832927
9459378972	9605697733	1001598997	2998482807	2873138839	7637508522
1156270803	1745774801	1544226029	7718016727	7240282667	6018667991
4281042330	4927846359	3928115864	5928523182	1770087219	1465417792

Métodos Cuantitativos para la Toma de Decisiones

Daniel Serra de La Figuera

Octubre 2002

Prólogo

Este libro se refiere al uso de modelos cuantitativos en la resolución de problemas de gestión y administración de sistemas complejos, con especial énfasis en la toma de decisiones. Ha sido escrito para dos tipos de usuarios: el administrador general o administrador en potencia, que puede sacar provecho de su uso en la comprensión y aplicación de los modelos cuantitativos para la toma de decisiones; o para estudiantes de estudios en donde la toma de decisiones juega un papel fundamental, como son los de gestión y administración y economía entre otros.

El libro no pretende ser exhaustivo en cuanto a las técnicas existentes, ya que existe un sinnúmero de excelentes manuales de técnicas cuantitativas y de investigación operativa (algunos de ellos se citan al final del libro). La gran diferencia entre este libro y los manuales clásicos radica en que el nivel de complejidad matemática se mantiene al mínimo nivel posible, y se hace especial énfasis en el planteamiento de modelos y en explicar como algunas de las técnicas existentes pueden ayudar a solucionar problemas que aparecen en cualquier organización. Por ello, para su lectura no se necesita una formación matemática previa; incluso se puede decir que en todo el libro no se utilizan más que las cuatro operaciones aritméticas básicas: sumar, restar, multiplicar y dividir.

También se incluye la posibilidad de obtener todos los problemas resueltos así como una explicación de cómo resolver cualquier formulación presentada en este libro con la popular hoja de cálculo Microsoft Excel 97 en la dirección electrónica siguiente:

<http://www.econ.upf.es/~serra/libro.htm>

Sobre el autor.

Daniel Serra de La Figuera es licenciado en Ciencias Económicas y Empresariales por la Universidad Autónoma de Barcelona, Master en Análisis de sistemas para la toma de decisiones y Doctor (Phd) por la Universidad Johns Hopkins de EEUU. Actualmente es Catedrático de Universidad de Organización de Empresas del departamento de Economía y Empresa en la Universidad Pompeu Fabra. Ha realizado diversos trabajos para el sector público y para el sector privado en el campo de la toma de decisiones y logística, tanto dentro del ámbito sanitario como el de transporte y distribución. Ha publicado varios artículos en prestigiosas revistas científicas internacionales e imparte clases en diversos masters y cursos de postgrado relacionados con la gestión y administración. Actualmente es director del Instituto de Estudios Territoriales de la Universidad Pompeu Fabra, investigador del Centre de Recerca en Economia i Salut y vicerrector de Economía, Promoción y Servicios de la misma universidad.

Agradecimientos

Este libro ha recibido el apoyo de la Fundación Banco Bilbao Vizcaya y del Centre de Recerca en Economia i Salut (CRÉS) de la Universidad Pompeu Fabra. EL autor también agradece a las profesoras Rosa Colomé Perales y Helena Ramalhinho Lourenço por su ayuda y aportaciones al texto.

Prohibida su reproducción sin el consentimiento del autor. Para entrar en contacto, enviar un mensaje a: daniel.serra@econ.upf.es