Simulación

Simulación de eventos discretos

Jorge de la Vega Góngora

Departamento de Estadística
Instituto Tecnológico Autónomo de México

21 de agosto de 2018

Modelo de Colas

Ejemplo 1: Línea de espera

- Consideren una unidad de servicio con un sólo servidor (cajero, taquilla, una pista de aeropuerto, etc.)
- Los clientes se forman para recibir el servicio.

Se desea estimar el tiempo promedio de permanencia en la fila, medido como el promedio del tiempo que transcurre desde que llega a la fila el total de clientes hasta que comienza a recibir el servicio.
Las variables que determinan el estado del sistema son:

- Estado del servidor (libre, ocupado)
- El número de clientes esperando en la fila $\{0, 1, 2, \ldots\}$
- El tiempo entre llegadas de clientes a la fila
- El tiempo de servicio a cada cliente

En este sistema hay dos tipos de eventos relevantes:

- Llegada de un cliente
- Completar el servicio (Partida de un cliente)

Ejemplo 1: Línea de espera

Definiciones:

$t_i =$ Tiempo de llegada del i-ésimo cliente.

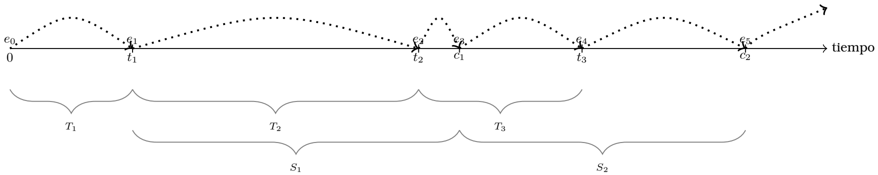
$T_i = t_i - t_{i-1} =$ tiempo entre las llegadas del $i-1$ e i cliente (tiempo de interarribo).

$S_i =$ tiempo de servicio al cliente i.

$D_i =$ tiempo de espera en cola del cliente i.

$c_i = t_i + D_i + S_i =$ tiempo transcurrido entre la llegada del cliente i y su salida del sistema.

$e_i =$ tiempo de ocurrencia del i-ésimo evento de cualquier tipo (reloj de simulación).



Medidas de desempeño del sistema I

- Supongamos que el sistema se ejecuta hasta que n clientes esperan en fila: la simulación termina cuando el cliente n entra a servicio.
- Bajo este supuesto la simulación tiene longitud aleatoria.
- Es posible simular de otro modo. Por ejemplo, podemos suponer que el sistema se observa un tiempo fijo y entonces lo que es aleatorio es el total de clientes que se consideraron en el sistema.

Bajo las condiciones mencionadas, consideremos tres medidas de desempeño del sistema, una para el cliente, y dos para el sistema.

- El *tiempo promedio de espera en cola* de los n clientes:

$$\bar{D}_n = \frac{\sum_{i=1}^{n} D_i}{n}$$

(los clientes que llegan cuando no hay nadie en la fila tienen $D_i = 0$).

Medidas de desempeño del sistema II

- El *número promedio de clientes en la cola*, q_n. Esta variable aleatoria depende del tiempo aleatorio necesario para observar n clientes.
  Para tomar en cuenta ese tiempo aleatorio, consideremos $Q(t) =$ número de clientes en la cola al tiempo t, $Q(t) \in \{0, 1, 2, \ldots\}$. Sea T_n el tiempo total que se requiere observar el sistema para observar las n variables $\{D_1, \ldots, D_n\}$. Entonces

$$q_n = \sum_{i=0}^{\infty} i \frac{T_i}{T_n}$$

  donde T_i es el tiempo del sistema que la cola tiene longitud i (notar que $T_n = \sum_{i=0}^{\infty} T_i$). Entonces:

$$\hat{q}_n = \frac{\sum_{i=0}^{\infty} iT_i}{T_n} = \frac{\int_0^{T_n} Q(t)\,dt}{T_n}$$

  La primera parte de la fórmula nos ayudará a calcular el estimador $\hat{q}_n$ como la suma de áreas de rectángulos.

- La *utilización esperada del servidor*, u_n. Es la proporción del tiempo de simulación que el servidor estuvo ocupado. Si definimos $B(t) = I(\text{servicio al tiempo } t)$, entonces:

$$\hat{u}_n = \frac{\int_0^{T_n} B(t)\,dt}{T_n}$$

La simulación nos tiene que dar los estimadores, y de ser posible, debería permitirnos visualizar el comportamiento de las funciones $Q(t)$ y $B(t)$.

Componentes de un modelo de simulación de eventos discretos I

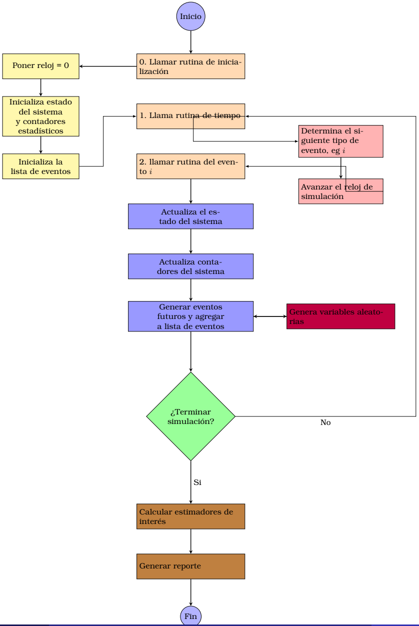
Los modelos de simulación discreta tienen componentes comunes, que facilitan su implementación computacional. Los componentes que comunmente se encuentran en este tipo de modelo son los siguientes:

- *Estado del sistema*: Es la colección de variables de estado necesarias para describir el sistema en particular.
- *Reloj de simulación*: la variable que lleva el tiempo simulado.
- *Lista de eventos*: Una lista con los tiempos en los que ocurrirá cada tipo de evento.
- *Contadores estadísticos*: Variables que almacenan la información sobre el desempeño del sistema.
- *Rutina de inicialización*: Un subprograma que inicializa el modelo de simulación (sus variables, listas, etc.) en el tiempo 0.
- *Rutina de tiempo*: Un subprograma que determina el siguiente evento de la lista de eventos y avanza el reloj de simulación al tiempo en que ocurre ese evento.
- *Rutina de evento*: Un subprograma que actualiza el estado del sistema cuando ocurre cierto tipo de evento (puede haber una rutina para cada tipo de evento).

- *Rutina de bibliotecas*: Un conjunto de subprogramas usadas para generar observaciones aleatorias con las distribuciones especificadas en el modelo de simulación.
- *Generador de reportes*: Un subprograma que calcula las medidas de desempeño elegidas para reportar y produce un reporte cuando termina la simulación.
- *Programa principal*: Un subprograma `mm1` que llama a los diferentes subprogramas conforme se requiere (típicamente llama primero a la rutina de inicialización y posteriormente de manera repetitiva a la rutina de tiempo y a la rutina de evento).

El código de colores corresponde a las actividades que se muestran en el siguiente diagrama de flujo, que conecta las relaciones lógicas entre los componentes del modelo.

Flujo del proceso

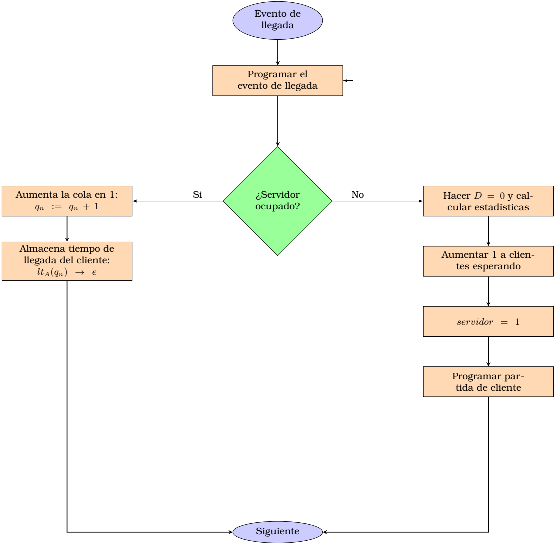


Supuestos del modelo de simulación I

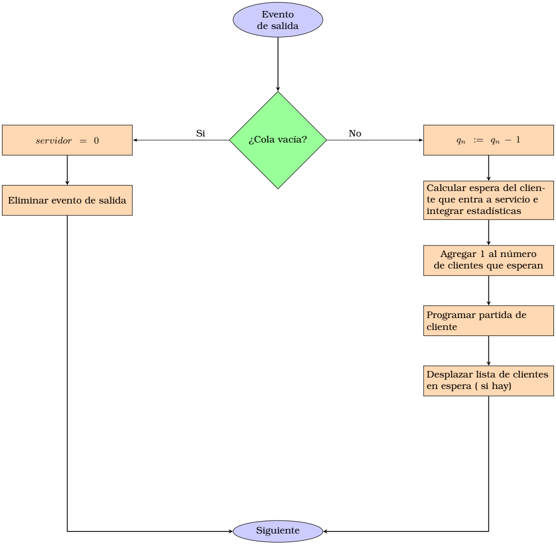
A continuación se resumen todos los supuestos que se utilizarán para el modelo $M/M/1$:

- El sistema se ejecuta hasta que n clientes esperan en fila: la simulación termina cuando el cliente n entra a servicio.
- Los tiempos de interarribo se modelarán como varibles aleatorias iid exponenciales con media λ_A. La distribución exponencial con media λ_A tiene densidad:
$$f(x) = \frac{1}{\lambda_A} e^{-x/\lambda_A} I(x \geq 0)$$
- Los tiempos de servicio se modelarán como variables aleatorias iid exponenciales con media λ_S.
- Hay dos eventos relevantes en la simulación:
  1. Llegada de un cliente al sistema
  2. Salida del cliente del sistemas después de recibir servicio.

Lógica del evento 1: Llegada de un cliente al sistema I



Lógica del evento de salida I



Funciones y Variables relevantes I

Parámetros del modelo:

- `lambdaA` y `lambdaS` son los tiempos promedio de interarribo y de servicio, respectivamente.
- `n` = número de clientes que entraron al sistema. Fijo de antemano.

Variables principales:

- `reloj` = e = reloj de simulación
- `sig_tipo_evento` = siguiente tipo de evento (1= llegada, 2= salida)
- `tiempo_sig_evento` = vector de dimensión 2, con el tiempo de llegada y el tiempo de salida de eventos conforme se registran
- `Di` = D_i tiempo de espera en cola de un cliente
- `total_esperas` = T_n = La suma de los tiempos de espera D_i de clientes en cola conforme se van generando
- `q_t` = longitud de la cola en un tiempo dado

Variables auxiliares:

- `lt_A` = lt_A vector lista de tiempos de arribo (longitud dinámica)
- `le` = lista de eventos ocurridos, con su tiempo de ocurrencia y su tipo y la longitud de la fila en esos tiempos.

Funciones y Variables relevantes II

- `area_q` = variable auxiliar para calcular la longitud promedio de la fila.
- `area_status_servidor` = variable auxiliar para calcular la ocupación del servidor.
- `tiempo_ultimo_evento` = registro del último evento antes de avanzar el reloj de simulación.
- `tiempo_desde_ultimo_evento` = calcula el intervalo de tiempo entre el actual y el tiempo del último evento.
- `clientes_enespera` = Número de clientes que se han formado en la fila.

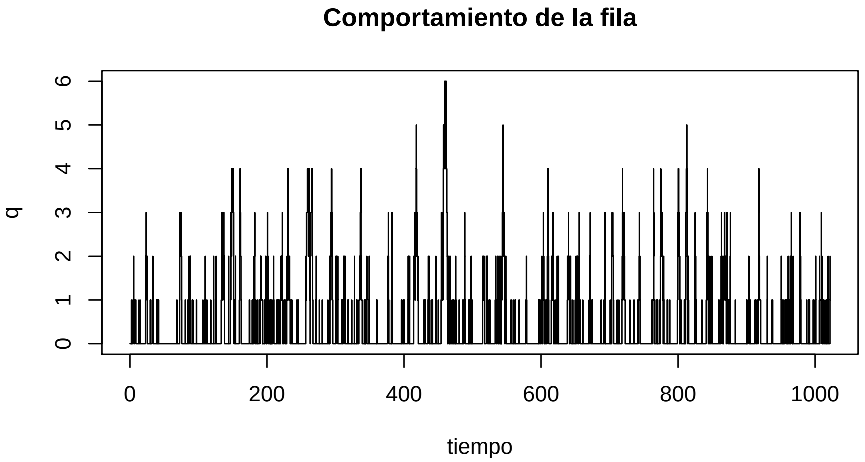
Ejecución del programa I

La siguiente simulación nos da una muestra de tamaño 1 del proceso $M/M/1$. Para sacar conclusiones definitivas, deberíamos ejecutar el modelo de simulación varias veces para obtener una muestra adecuada de las cantidades estimadas.
El programa `Queue.R` lo pueden obtener de la siguiente liga de Piazza

```
source("../scripts/Queue/Queue.R")
a <- mm1(lambdaA = 1, lambdaS = 0.5, n = 1000)

[1] "Promedio de espera en la fila: 0.44 minutos"
[1] "Número promedio de clientes esperando en la fila: 0.43"
[1] "Utilización del servidor: 51 %"
[1] "El tiempo de simulación fue de: 1021.89 minutos"

plot(a$le[,1],a$le[,3],type="s",xlab="tiempo",ylab="q",main="Comportamiento de la fila")
```


Comportamiento de la fila
q
0
1
2
3
4
5
6
0
200
400
600
800
1000
tiempo

De acuerdo a la teoría de los modelos de linea de espera, considerando tiempos de arribo y tiempos de servicio con medias λ_A y λ_S respectivamente, las cantidades que estimamos en el ejemplo anterior tienen los siguientes valores esperados teóricos:

- La utilización esperada del servidor es $u = \frac{\lambda_S}{\lambda_A}$.
- El número promedio de clientes en la cola: $q_n = \frac{\lambda_S^2}{\lambda_A(\lambda_A - \lambda_S)}$
- El tiempo promedio de espera en cola es: $D_n = \frac{\lambda_S^2}{\lambda_A - \lambda_S}$

Modelo de inventario

Ejemplo 2: Simulación de un modelo de inventario I

- Inventario de un producto, por ejemplo, billete de \$500. ¿Cuántos billetes se deben producir para atender la demanda de los próximos n meses?
- La demanda tiene las siguientes características:
  1. $D \sim F_D$ es el tamaño de la demanda (número de unidades demandadas)
  2. $T_i \sim \exp(\lambda)$ es el tiempo entre solicitudes
  3. Costo: Al inicio de cada mes se revisa el inventario y se decide el tamaño de la orden a fabricar. Si se ordenan Z unidades, el costo es

     $$\texttt{Costo} = \begin{cases} 0 & i = 0 \\ K + iZ & \text{en otro caso} \end{cases}$$

     donde K es el costo fijo e i es el costo unitario.
  4. El tiempo de entrega de la orden (conocido como *lead time* tiene distribución uniforme entre a y b
  5. Hay una política de reorden de tipo (s, S), donde s es un indicador para el nivel mínimo del inventario y S es el nivel máximo del inventario:

     $$Z = \begin{cases} S - I & I \leq s \\ 0 & I > s \end{cases}$$

     donde I es el nivel del inventario al inicio del mes
  6. Se supone que la demanda no satisfecha se atiende en el futuro, es decir, el nivel del inventario puede ser negativo.

Ejemplo 2: Simulación de un modelo de inventario II

- Otros costos a considerar en el modelo:
  - Costo de almacenamiento: será de h por unidad almacenada (rentas, seguros, impuestos, etc). Sea $I^+(t) = \max\{I(t), 0\}$ el inventario en el tiempo t. Entonces el costo promedio por mes es $h\bar{I^+}$ donde

$$\bar{I^+} = \frac{1}{n}\int_0^n I^+(t)\,dt$$

  - Costo por atrasos: π por unidad (reputación, registros). Sea $I^-(t) = \max\{-I(t), 0\}$ la demanda no satisfecha. Entonces el costo promedio por mes es $\pi\bar{I^+}$, donde

$$\bar{I^-} = \frac{1}{n}\int_0^n I^-(t)\,dt$$

Las variables de estado del sistema son:

- $I(t)$ el nivel del inventario al tiempo t
- Tamaño de la órden
- Tiempo de los eventos

y los tipos de eventos de este modelo:

1. Llegada de orden del proveedor
2. Solicitud de una orden (Demanda)
3. Fin de la simulación a los n meses
4. Evaluación del estado del inventario al inicio de cada mes.