

# Simulación

1. Ejemplo e introducción a la simulación
2. Ventajas e inconvenientes de la simulación
3. Tipos de sistemas y simulaciones
4. Elementos de la simulación por eventos discretos
5. Simulación por incremento de tiempo fijo y por eventos
6. Estructura y traza de un modelo de simulación
7. Generación de muestras uniformes
8. Generación de números pseudoaleatorios
9. Generación de distribuciones discretas y continuas

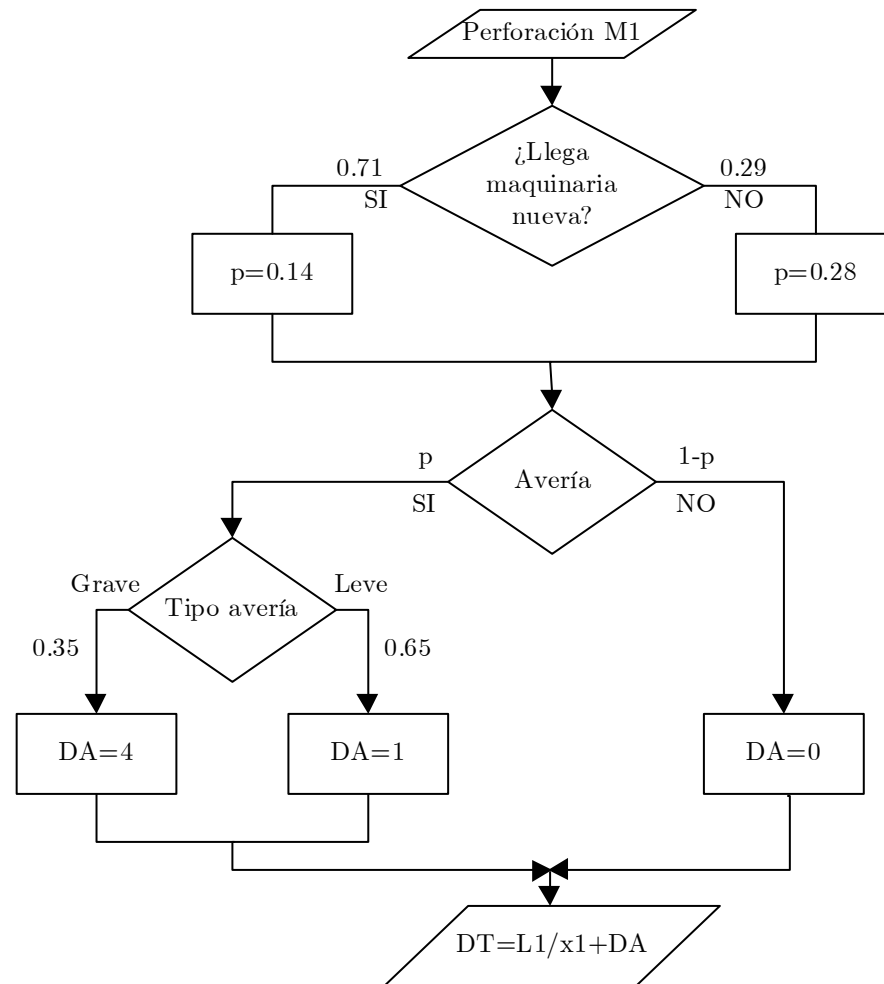
## Ejemplo

Se desea construir una carretera y para ello se ha de hacer un túnel a través de una montaña. Existen dos puntos posibles donde hacer el túnel, correspondientes a dos montañas distintas M1 y M2 cercanas. Supongamos que se tiene que decidir si hacerlo en un punto o en otro. Mediante estudios preliminares se sabe que en el punto M1 la longitud del túnel habría de ser  $L_1$ , y en la montaña M2,  $L_2$ . En la primera de ellas, por las características del terreno, se perforaría a razón de  $x_1$  unidades por jornada de trabajo, mientras que en la otra montaña sería a razón de  $x_2$  unidades.

La empresa debe recibir una maquinaria nueva con una probabilidad 0.71. La probabilidad de que la nueva maquinaria se averíe en M1 es 0.14 y en M2 es 0.16. Para la maquinaria vieja estas probabilidades son, respectivamente, 0.28 y 0.19.

Las averías pueden ser de dos tipos: graves con probabilidad 0.35 y 4 jornadas de trabajo de reparación o leves con 1 jornada de trabajo de reparación.

¿Dónde se debe perforar el túnel para tardar lo menos posible en la construcción de la carretera?



## **Modelo de simulación**

1. Se definen los parámetros aleatorios (llegada de nueva maquinaria, avería, gravedad de la avería)
2. Se simulan o muestrean los parámetros aleatorios
3. Se recopilan los resultados
4. Se detiene el proceso de muestreo

# Introducción

Sistema:

*Elementos (componentes)* cuyo comportamiento se representa en términos de variables para cada estado posible

*Relaciones (reglas de operación lógicas)* entre elementos

Simulación:

Proceso de *diseño* de un modelo estocástico de un sistema real y realización de *experimentos* para reproducir su funcionamiento (*muestreo estocástico de un modelo*)

Objetivo:

Modelado y simulación de *sistemas estocásticos* donde los *eventos* que ocurren en el sistema se generan mediante *distribuciones de probabilidad*

## **¿Cuándo conviene utilizar simulación?**

- No existe formulación matemática conocida del problema (líneas de espera)
- Los modelos analíticos no son eficientes (sistemas complejos)
- Se desea experimentar con el sistema antes de su uso (para formar a pilotos mediante los simuladores de vuelo)
- Se desea experimentar con el sistema para mejorarlo, detectando eventualidades (simuladores de vuelo del espacio)
- Para reducir escalas de tiempo porque la evolución del sistema sea muy lenta (disminución de la capa de ozono, talas de árboles)

## **Inconvenientes de la simulación**

- Incluso la construcción del modelo puede ser compleja y costosa (aunque esto siempre ocurre en el desarrollo de cualquier modelo; por ejemplo se pueden olvidar relaciones entre los elementos del sistema a modelar)
- Dificultad de establecer la precisión de los resultados, pues los resultados son muestras y se deben utilizar técnicas estadísticas

# Tipos de sistemas

## 1. CONTINUOS

Las variables de estado cambian de forma continua con el tiempo. Por ejemplo, el movimiento de un tren por una vía, en donde las variables de estado son la posición, la velocidad y la aceleración

## 2. DISCRETOS

Las variables de estado cambian en ciertos instantes de tiempo. Por ejemplo, un sistema de atención a clientes atendido por un solo servidor, la variable de estado es el número de clientes que hay en el centro de servicio.



## Tipos de simulación

- Según la *evolución del tiempo*
  - *Estáticos*: representan un sistema en un instante particular. A menudo se denomina simulación de Monte Carlo.
  - *Dinámicos*: representan un sistema que evoluciona con el tiempo.
- Según la *aleatoriedad*
  - *Deterministas*: no incluyen variables aleatorias. Dados unos datos de entrada, existe un único conjunto posible de datos de salida.
  - *Probabilistas* o *estocásticos*: contienen variables aleatorias, siendo las salidas también aleatorias (estimaciones de las verdaderas características).
- Según las *variables de estado*
  - *Continuos*: si todas las variables de estado cambian de forma continua con el tiempo.
  - *Discretos*: si todas las variables de estado cambian en determinados instantes de tiempo.
  - *Híbridos* o *combinados*: si incluyen variables de estado continuas y discretas.

Nos centraremos en los modelos de simulación *dinámicos*, *estocásticos* y *discretos*, denominados modelos de *simulación de eventos discretos*.

# Elementos de la simulación de eventos discretos

## 1. ESTADOS Y EVENTOS

*Estado* del sistema definido por el estado de cada uno de sus componentes.

*Eventos*: sucesos que producen un cambio en el estado del sistema.

## 2. MECANISMO DE TRANSICIÓN ENTRE ESTADOS POR OCURRENCIA DE EVENTOS

Ejemplo: Sistema de colas con 1 servidor

**Estado**: número de clientes en el sistema (en cola o siendo atendidos)  $N(t)$  en el instante  $t$ .

**Eventos**: llegada de un cliente o final de un servicio

**Mecanismo de transición**: 
$$N(t) = \begin{cases} N(t) + 1 & \text{si llegada cliente} \\ N(t) - 1 & \text{si final de servicio cliente} \end{cases}$$

### 3. RELOJ DE SIMULACIÓN

Mide el *tiempo simulado* para el sistema. Mecanismos de avance:

- INCREMENTO DE TIEMPO FIJO (*time step*)

Se avanza un intervalo fijo (pequeño) de tiempo y se actualiza el estado del sistema (se comprueba si ha ocurrido algún evento en el sistema). Se considera que los eventos ocurren al final del intervalo. Problemas asociados a simultaneidad de eventos (ver transparencia 14 cuando  $P_e = P_s$ ).

- INCREMENTO POR EVENTO (*event step*)

Se avanza el tiempo hasta el próximo evento y se actualiza el estado del sistema.

Se genera el intervalo de tiempo hasta el próximo evento. En todos los intervalos hay un evento luego más eficiente que incremento por tiempo fijo (en donde si el incremento del tiempo es pequeño, puede no haber ningún evento).

#### 4. LISTA DE EVENTOS

Necesaria para saber cuáles son los eventos futuros que se deben incluir en la actualización del sistema (ya sea con incremento de tiempo fijo o por evento) y en qué instantes se van a dar, al menos los más inmediatos

#### 5. CONTADORES ESTADÍSTICOS

Variables utilizadas para almacenar información sobre el comportamiento del sistema y que al final, mediante algún cálculo matemático, darán respuesta al objetivo del estudio.

## Sistema de colas M/M/1

Tasa media de llegada  $\lambda$ : 3 clientes/hora con distribución de probabilidad exponencial

Tasa media de servicio  $\mu$ : 5 clientes/hora con distribución de probabilidad exponencial

$t$  tiempo del sistema

$N(t)$  estado del sistema (número de clientes en el sistema = clientes en cola + clientes en servicio)

Eventos: llegada de un cliente, final del servicio a un cliente

Transición entre estados

$N(t) = N(t) + 1$  llega un cliente al sistema en el tiempo  $t$

$N(t) = N(t) - 1$  sale un cliente del sistema en el tiempo  $t$

Estado inicial: ningún cliente en el sistema

## Simulación por incremento de tiempo fijo

$$\Delta t = 0.1 \text{ h}$$

### Hipótesis

No hay simultaneidad de eventos. Sólo hay una llegada o salida en cada intervalo  $[t, t+\Delta t]$

Si  $T$  es una exponencial( $\lambda$ ) que mide el tiempo de llegadas, la probabilidad de llegada en  $\Delta t$  supuesto que el tiempo es mayor que  $t$

$$P_e = P\{T \leq t + \Delta t \mid T > t\} = 1 - e^{-\lambda \Delta t} = 1 - e^{-0.3} = 0.26$$

Si  $T$  es una exponencial( $\mu$ ) que mide el tiempo de salidas, la probabilidad de salida en  $\Delta t$

$$P_s = P\{T \leq t + \Delta t \mid T > t\} = 1 - e^{-\mu \Delta t} = 1 - e^{-0.5} = 0.39$$

Estas probabilidades permiten determinar qué evento ocurre en cada intervalo  $[t, t+\Delta t]$ :

- Generador de evento  $\Rightarrow$  generador número aleatorio uniformemente distribuido  $U[0,1]$

- |               |                    |                |                       |
|---------------|--------------------|----------------|-----------------------|
| • $U < 0.26$  | ocurre una llegada | $U \geq 0.26$  | no ocurre una llegada |
| • $U' < 0.39$ | ocurre una salida  | $U' \geq 0.39$ | no ocurre una salida  |

Nótese que para que no haya simultaneidad de eventos es necesario  $P_e \neq P_s$ .

En el caso en que aun siendo  $P_e \neq P_s$  se produzca una llegada y una salida en  $[t, t+\Delta t]$ , se puede proceder:

- Obviando cada caso con simultaneidad de eventos (se vuelve a generar  $U[0,1]$ ).
- Dado que la probabilidad  $P_e \cdot P_s$  de que se produzca simultaneidad tiende a 0 con  $\Delta t$  mucho más rápidamente que  $P_e$  o  $P_s$ , decrementando  $\Delta t$ . El problema es que si se decremента en exceso pudiera no ocurrir nada en  $[t, t+\Delta t]$  (dado que  $P_e$  y  $P_s$  pequeños).

## **Simulación por eventos**

- Se generan los tiempos de ocurrencia de eventos de llegada o salida en función de sus distribuciones de probabilidad mediante números aleatorios.
- Se avanza el tiempo de reloj hasta el siguiente evento (mínimo de los tiempos futuros más cercanos)

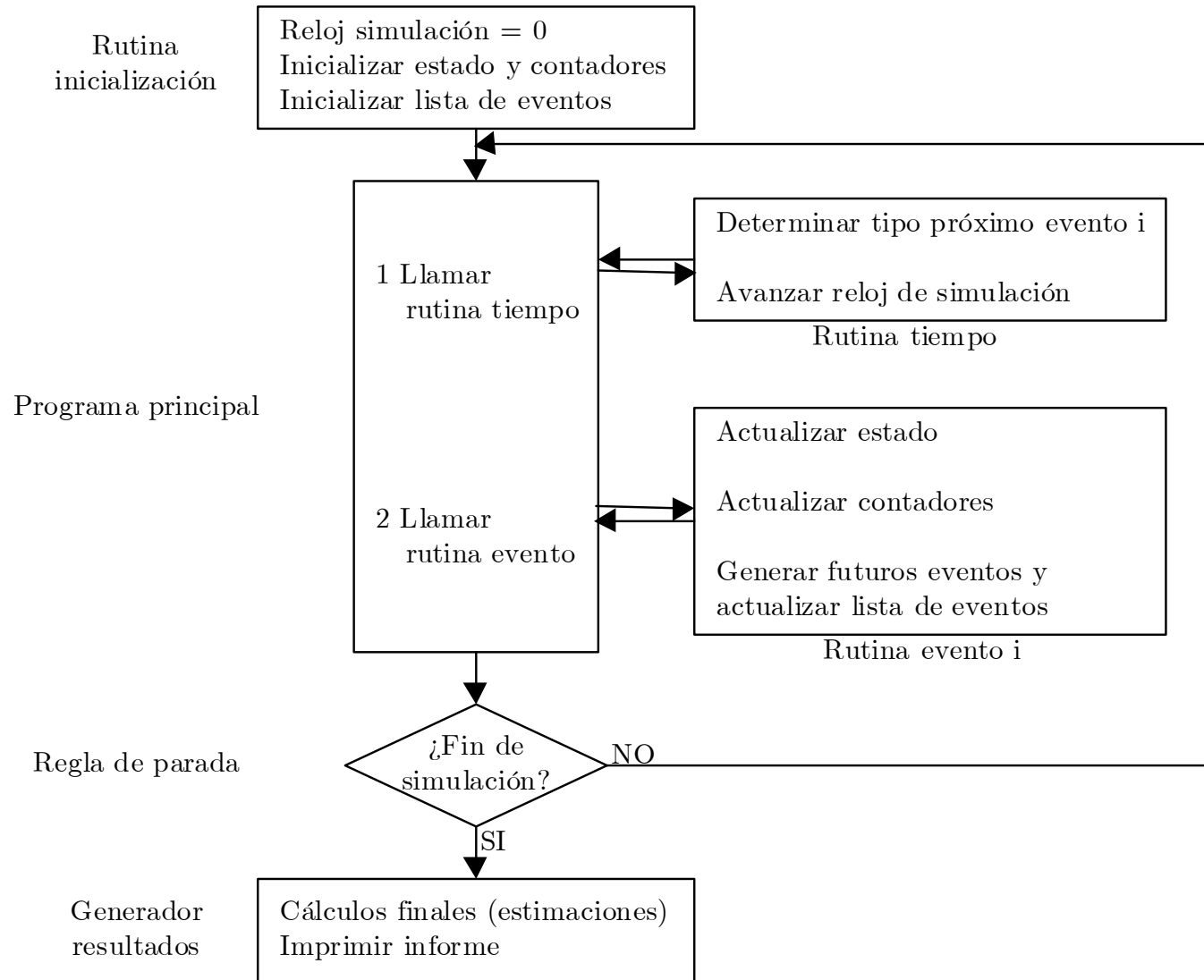
## **Resultados de la simulación**

- Se van calculando para cada instante discreto las medidas del sistema (por ejemplo en un modelo de colas las medidas de eficiencia: clientes en cola, clientes en servicio, etc.)



## **Estructura (informática) de un modelo de simulación por eventos**

- Rutina de tiempo  
Determina el siguiente evento y avanza el reloj de simulación
- Rutina de evento  
Actualiza el estado del sistema. Una por cada tipo de evento.
- Generador de informes o resultados  
Realiza los cálculos y valores estadísticos de las características que se desean medir.
- Programa principal  
Enlaza las rutinas anteriores.



## Ejemplo: Sistema de una línea de espera con un servidor

Objetivo: estimar el número medio de clientes en el sistema.

Hipótesis y datos:

- Tiempos entre llegadas de clientes según distribución F
- Tiempos de servicio según distribución G
- Tiempos independientes
- T tiempo máximo de simulación

Variable de estado:  $N$  el número de clientes en el sistema.

Eventos:

- Llegada de un cliente al sistema
- Final de servicio de un cliente

El mecanismo de transición se define como

$$N(t) \leftarrow \begin{cases} N(t) + 1 & \text{si es llegada de un cliente} \\ N(t) - 1 & \text{si es final de servicio} \end{cases}$$

## Ejemplo: Sistema de una línea de espera con un servidor (cont.)

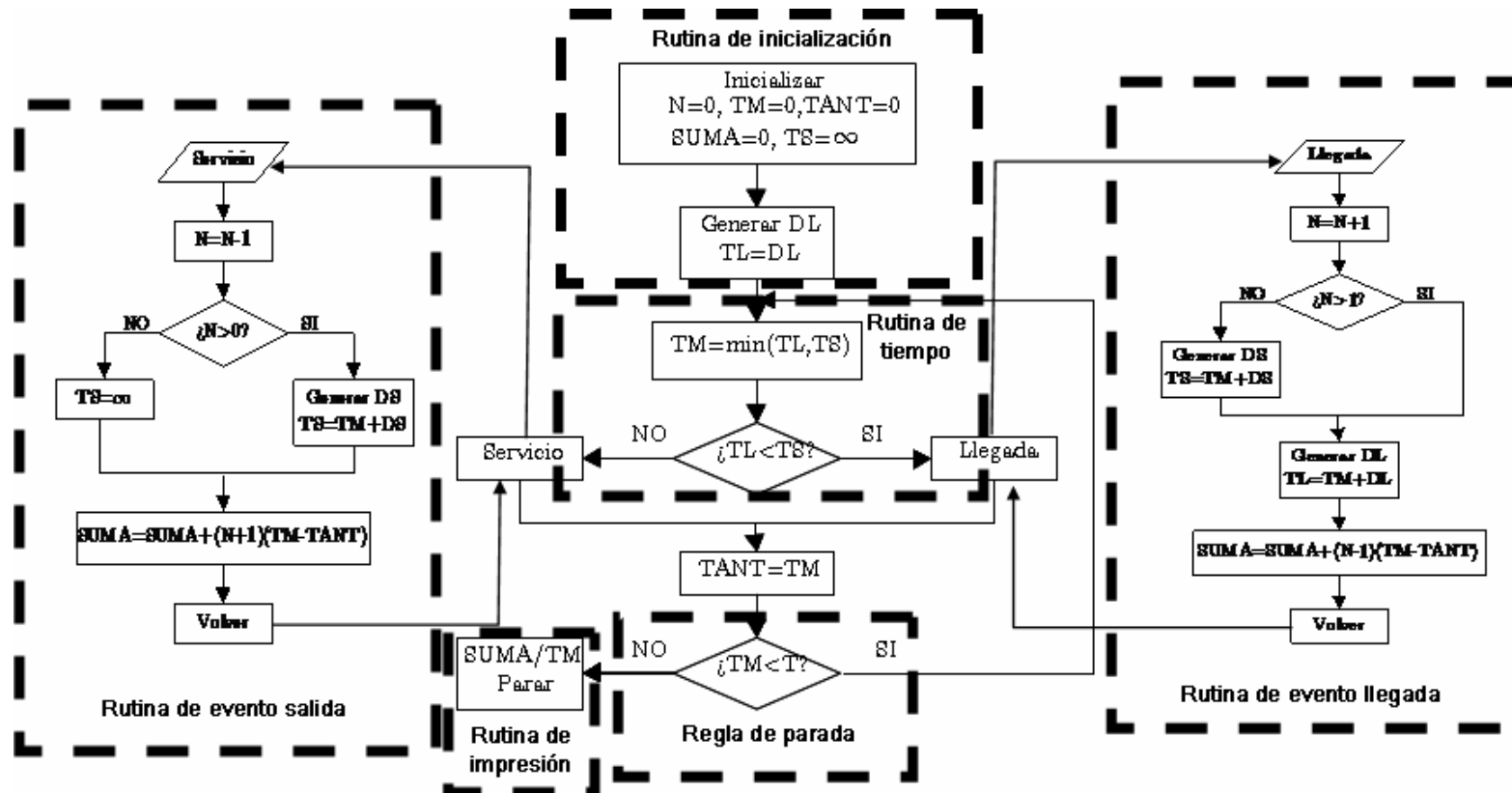
Variables principales son:

- N número de clientes en el sistema
- TM reloj de simulación
- DL tiempo entre llegadas  $= F^D$
- DS tiempo de servicio  $= G^D$

Otras variables auxiliares son:

- TL instante de la próxima llegada
- TS instante del próximo final de servicio
- TANT variable auxiliar (instante del último evento)
- SUMA contador acumulando suma de tiempos de permanencia de clientes

## Diagrama del modelo



## **Traza del modelo**

Tabla en la que se recogen los valores de las variables que intervienen en el modelo en varias iteraciones.

- Permite detectar la necesidad de variables auxiliares.
- Es una forma de verificar la programación posterior.

Datos de tiempo entre llegadas DL: 3, 2, 5, 1, 2, 6, 6, 2, 8

Datos de tiempo de servicio DS: 4, 1, 3, 1, 3, 2, 3, 5

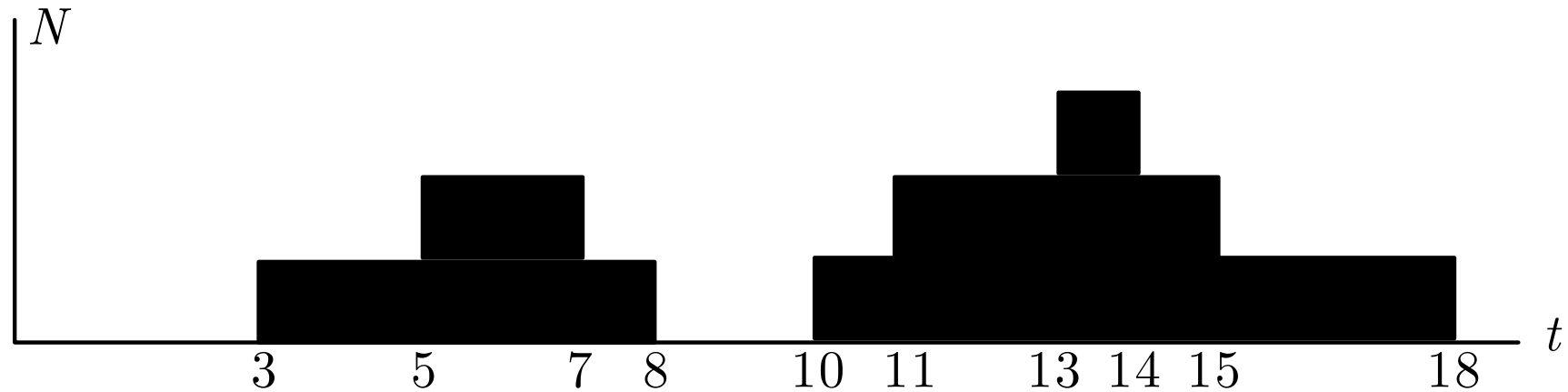
Tiempo de simulación T= 35

Nº evento	Reloj Simulación	Tipo evento	N	TL	TS	SUMA
0	0	Inicio	0	3	$\infty$	0
1	3	Llegada	1	5	7	$0+0\cdot3=0$
2	5	Llegada	2	10	7	$0+1\cdot2=2$
3	7	Servicio	1	10	8	$2+2\cdot2=6$
4	8	Servicio	0	10	$\infty$	$6+1\cdot1=7$
5	10	Llegada	1	11	14	$7+0=7$
6	11	Llegada	2	13	14	$7+1\cdot1=8$
7	13	Llegada	3	19	14	$8+2\cdot2=12$
8	14	Servicio	2	19	15	$12+3\cdot1=15$
9	15	Servicio	1	19	18	$15+2\cdot1=17$
10	18	Servicio	0	19	$\infty$	$17+3\cdot1=20$
11	19	Llegada	1	25	21	$20+0=20$
12	21	Servicio	0	25	$\infty$	$20+1\cdot2=22$
13	25	Llegada	1	27	28	$22+0=22$

14	27	Llegada	2	35	28	$22+1\cdot2=24$
15	28	Servicio	1	35	33	$24+2\cdot1=26$
16	33	Servicio	0	35	$\infty$	$26+1\cdot5=31$
17	35	Final				$31+0\cdot2=31$

Número medio de clientes en el sistema:  $\hat{E}[N] = \frac{31}{35} = 0.89$





Si el tiempo máximo de simulación hubiera sido  $T = 18$

Número medio de clientes en el sistema:  $\hat{E}[N] = \frac{20}{18} = 1.11$

El tiempo de simulación influye en los resultados.

## Ejemplo disponibilidad dinámica de un sistema paralelo

- El sistema está indisponible cuando *todos* los componentes están indisponibles.
- Cuando se repara un componente vuelve a funcionar como nuevo.

Datos:

- Distribución de probabilidad de tiempo entre fallos del componente  $i$   $TF_i(t)$
- Distribución de probabilidad de tiempo de reparación del componente  $i$   $TR_i(t)$
- Tiempo máximo de simulación  $T$

Variables de estado:

- $x_i \begin{cases} 1 & \text{componente en funcionamiento} \\ 0 & \text{componente en reparación} \end{cases}$
- $ES$  estado del sistema  $\begin{cases} 1 & \text{sistema en funcionamiento} \\ 0 & \text{sistema en fallo} \end{cases}$

Variables principales:

- $TM$  tiempo del reloj de simulación
- $TF_i$  muestra de la distribución de tiempo entre fallos del componente  $i$
- $TR_i$  muestra de la distribución de tiempo de reparación del componente  $i$

Variables auxiliares:

- $TP_i$  instante en el que se produce el siguiente evento del componente  $i$
- $TPS$  tiempo (no instante) total de parada del sistema
- $TI$  instante de la última parada del sistema

## Proceso de simulación

### 1. Inicialización de variables

- $TM = 0$
- $TPS = 0$

### 2. Definir estado inicial

- $x_i = 1$  todos los componentes están disponibles
- $ES = 1$  el sistema está disponible

### 3. Generar muestra de distribución de tiempo entre fallos $TF_i$

### 4. $TP_i = TF_i$

### 5. Rutina de tiempo: Determinar mínimo de $TP_i$ (cuándo), el componente al que afecta (a quién), y el evento de fallo o reparación (qué)

- Actualizar el tiempo de simulación  $TM = TP_i$
- Dependiendo de la respuesta a “quié” y a “qué” (la respuesta a qué se deduce del valor que tenga  $x_i$ ) vamos al paso 6 o al 7.
-

6. Rutina de evento de fallo del componente  $i$ : En caso de fallo del componente  $i$
- $x_i = 0$
  - Y estar fallados el resto de componentes se declara indisponible el sistema  $ES = 0$  y  $TI = TM$
  - Generar muestra de distribución de tiempo de reparación  $TR_i$
  - $TP_i = TP_i + TR_i$
7. Rutina de evento de fin de reparación del componente  $i$ : En caso de reparación del componente  $i$
- $x_i = 1$
  - Si el sistema estaba indisponible ( $ES = 0$ ) se contabiliza el tiempo de parada del sistema  $TPS = TPS + (TM - TI)$  y se declara disponible el sistema  $ES = 1$
  - Generar muestra de distribución de tiempo entre fallos  $TF_i$
  - $TP_i = TP_i + TF_i$
8. Comprobar que  $TM < T$ . En caso contrario ir a 10
9. Continuar la simulación. Ir a 5
10. Final. Índice de disponibilidad  $= 1 - TPS / TM$

# Generación de números aleatorios

## Generación de muestras uniformes

Se generan valores de variables aleatorias uniformes en  $(0,1)$  dado que los métodos para generar variables aleatorias son transformaciones  $U(0,1)$

### *Método de la ruleta (método físico)*

Ruleta dividida en 10 partes iguales, con valores del 0 al 9, y una flecha fija que señala.

Para generar valores de una uniforme en  $(0,1)$ , se puede agrupar los valores de  $k$  en  $k$  y considerar que cada grupo son las cifras decimales de una realización de  $U(0,1)$ .

Para considerar informativos los resultados de la simulación es necesario simular muchas veces, más cuanto más complicado es el modelo, siendo necesario el uso del ordenador.

Existen *tablas de números aleatorios* obtenidos por el método de la ruleta y otros métodos físicos, pero no es un buen método para su uso en ordenador. Por ello se crearon métodos aritméticos particulares adaptados al ordenador, aunque con un cierto deterioro de la aleatoriedad, denominándose *números pseudoaleatorios*.

## ***Generación de números (pseudo)aleatorios***

Producir una secuencia de números que siguen una distribución de probabilidad específica y tienen *apariencia* y *propiedades* de aleatoriedad. El procedimiento debe ser cómodo para su implementación en un ordenador.

*Características* de los números pseudoaleatorios:

1. Generan realizaciones de distribuciones uniformes (0,1).
2. Uniformidad e independencia en las secuencias<sup>1</sup>.
3. Reproducibilidad.
4. Capaces de producir diferentes secuencias de números.
5. Rápidos de generar y no requerir almacenamiento.
6. Tener un ciclo no repetitivo tan largo como se desee.

Los siguientes métodos generan a gran velocidad y ocupan poca memoria de ordenador, puesto que para generar un valor sólo necesito el último valor generado. También son reproducibles (con una misma semilla se obtiene una misma secuencia). No obstante, pueden tener problemas respecto a las características 2 y 6.

---

<sup>1</sup>Los generadores tienen que cumplir propiedades estadísticas (característica 2): uniformidad e independencia de las secuencias. Para ello se han generado distintas secuencias y se les ha hecho pasar contrastes de bondad de ajuste para la uniformidad y la independencia.

### Método de los cuadrados medios

Se toma un número al azar,  $x_0$ , de  $2n$  cifras. Se eleva al cuadrado y se toman de este resultado las  $2n$  cifras centrales siendo el nuevo número  $x_1$  a partir del cual se repite el proceso.

Ejemplo:

$x_0=4122$	$x_0^2=16 9908 84$
$x_1=9908$	$x_1^2=98 1684 64$
$x_2=1684$	$x_2^2=2 8358 56$

La secuencia 4122, 9908, 1684, ... puede ser considerada, al menos a partir de un cierto intervalo, como una secuencia de números pseudoaleatorios. Cada número puede considerarse los decimales de un número uniforme (0,1).

El principal problema de este método es que los números pueden repetirse a partir de una secuencia muy corta. Por ejemplo, si  $x_0 = 3708$  se tiene que  $x_4 = 6100 = x_8$ , lo cual no es controlable (no cumple con la característica 6).



### Método de Lehmer

Sea  $x_0$  un número al azar de  $n$  cifras. Se multiplica por otro  $k'$  (fijo del generador) de  $k$  cifras, dando lugar a uno de  $n+k$ . Se quitan las  $k$  cifras de la izquierda, obteniendo uno de  $n$  cifras al que se resta el de  $k$  cifras que se había separado siendo el nuevo número  $x_1$  a partir del cual se repite el proceso.

Ejemplo:

$x_0=4122$	$k'=76$	$4122 \cdot 76=31 3272$	$3272-31=3241$
$x_1=3241$	$k'=76$	$3241 \cdot 76=24 6316$	$6316-24=6292$

Cada número puede considerarse los decimales de un número uniforme (0,1). Este método acaba degenerando a 0 (no cumple con la característica 2).

### Método congruencial mixto

Obtiene una secuencia de números pseudoaleatorios en el conjunto  $\{0, \dots, m-1\}$  calculados a partir del anterior.

$$x_{k+1} = (ax_k + b) \bmod m$$

$x_0$  semilla conocida       $m$  longitud del ciclo       $a, b, m$  enteros positivos       $a, b < m$

Si  $b = 0$  el método se denomina *congruencial multiplicativo*

Sean  $x, y, m \in \mathbb{Z}$ . Se dice que  $x$  es congruente con  $y$  en módulo  $m$  si y sólo si  $x$  e  $y$  dan el mismo resto al dividir por  $m$ . Se representa por  $x \equiv y \pmod{m}$ .

Ejemplo:       $m = 9, a = 5, b = 1, x_0 = 1$       Secuencia generada: 6, 4, 3, 7, 0, 1

$$y_1 = 5 \cdot 1 + 1 = 6 \quad x_1 = 6$$

$$y_2 = 5 \cdot 6 + 1 = 31 \quad x_2 = 4$$

$$y_3 = 5 \cdot 4 + 1 = 21 \quad x_3 = 3$$

$$y_4 = 5 \cdot 3 + 1 = 16 \quad x_4 = 7$$

$$y_5 = 5 \cdot 7 + 1 = 36 \quad x_5 = 0$$

$$y_6 = 5 \cdot 0 + 1 = 1 \quad x_6 = 1 = x_0$$

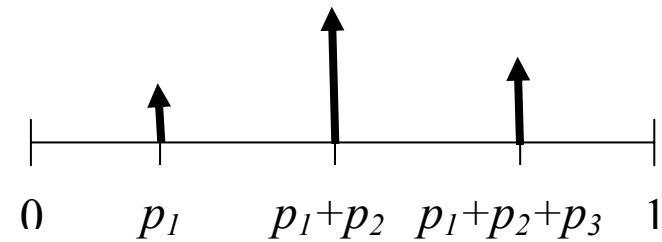
Respecto al ciclo no repetitivo (característica 6) en el mismo ejemplo se puede ver que es muy corto. De hecho, la máxima longitud que se puede lograr sin repetir es de  $m$ , ya que es la mayor cantidad de números diferentes que se puede obtener y una vez que se obtiene un número repetido toda la secuencia se repite. Para seguir generando números aleatorios sin ciclos se puede modificar el valor de  $a$  y  $b$  en el momento de la repetición.

Se puede afirmar que los dos generadores multiplicativos con  $m = 2^{31} - 1$  y multiplicador  $a = 16807$  o bien  $a = 63036016$  generan valores con la propiedad de la uniformidad y la independencia y con ciclo no repetitivo maximal (el ciclo con mayor número de elementos): El primero es más rápido y tiene menor riesgo de desbordamiento de memoria. El segundo tiene mejores propiedades estadísticas.

Muestras uniformes en el intervalo  $(0,1)$  se obtienen con el cociente de dividir cada número generado por el módulo  $m$  ( $u_k = \frac{x_k}{m} \quad \forall n$ ).

## Generación de variables aleatorias discretas

$$X = \begin{cases} x_1 & \text{con prob } p_1 \\ x_2 & \text{con prob } p_2 \\ x_3 & \text{con prob } p_3 \\ \vdots & \end{cases}$$



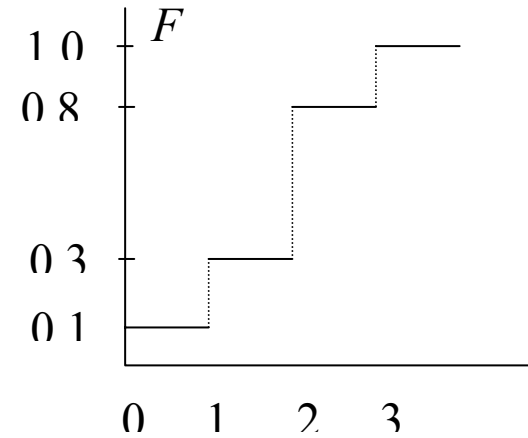
$$\sum_k p_k = 1$$

$$u \in U(0,1)$$

$$X = x_i \quad \text{si} \quad \begin{aligned} &\sum_{k=1}^{i-1} p_k \leq u < \sum_{k=1}^i p_k \\ &F_x(x_{i-1}) \leq u < F_x(x_i) \end{aligned}$$

Ejemplo:

$$X = \begin{cases} 0 & \text{con prob } p_1 = 0.1 \\ 1 & \text{con prob } p_2 = 0.2 \\ 2 & \text{con prob } p_3 = 0.5 \\ 3 & \text{con prob } p_4 = 0.2 \end{cases}$$



$u:$	0.27	0.54	0.06	0.89	0.15
$X:$	1	2	0	3	1

### **Método de la distribución binomial $(n, p)$**

Se basa en la propiedad según la cual la distribución de la suma de v.a.i.i.d. con distribución de Bernoulli de parámetro  $p$  es  $\text{Binomial}(n, p)$ . Luego, para generar valores de una variable con distribución  $\text{Binomial}(n, p)$ , se pueden generar  $n$  Bernoulli de parámetro  $p$  y sumarlas.

Algoritmo:

1.  $x \leftarrow 0$
2. Hacer  $n$  veces

Generar  $u \in U(0,1)$

Si  $u \leq p$   $x \leftarrow x + 1$

3. Salida:  $x$  se distribuye según  $\text{Binomial}(n, p)$

### **Método de la distribución Geométrica<sup>(p)</sup>**

La distribución geométrica corresponde al número de ensayo en que aparece el primer éxito al repetir un experimento de Bernoulli de parámetro  $p$ . Así que para generar valores con esa distribución es posible hacerlo con el método tradicional o aprovechando esta propiedad.

Con el método tradicional, basta aplicarlo con su función de distribución  $[F(x) = 1 - (1 - p)^x \quad x = 1, 2, \dots]$  para ver que la expresión sería:

$$x = 1 + \left\lceil \frac{\ln(1 - u)}{\ln(1 - p)} \right\rceil \stackrel{d}{=} 1 + \left\lceil \frac{\ln(u)}{\ln(1 - p)} \right\rceil$$

Utilizando la propiedad que la relaciona con los experimentos de Bernoulli el algoritmo sería:

1.  $x \leftarrow 0$
2. Hacer hasta que  $u \leq p$

$$x \leftarrow x + 1$$

Generar  $u \in U(0,1)$

3. Salida:  $X$  se distribuye según Geométrica<sup>(p)</sup>

Este método es bueno cuando el valor de  $p$  es grande (en caso contrario poca eficiencia dado que se alargaría en exceso el proceso hasta llegar al éxito).

# Generación de variables aleatorias continuas

## Método de la transformación inversa

$X$  variable aleatoria

$F(x) = P\{x \leq X\}$  función de distribución acumulada

Procedimiento:

1. Generar número aleatorio uniforme entre 0 y 1,  $r$
2.  $F(x) = r$  se determina  $x$

Se puede aplicar a la uniforme, exponencial, Weibull

Ejemplo: *Distribución exponencial*

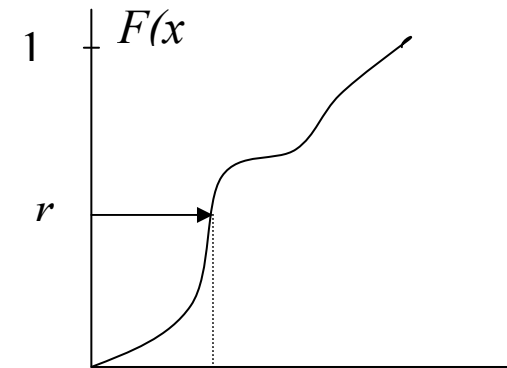
$$F(x) = 1 - e^{-\alpha x} \quad x \geq 0 \quad 1/\alpha \text{ media de la distribución}$$

$$F(x) = r \quad \Rightarrow \quad x = \frac{\ln(1-r)}{-\alpha}$$

Dado que la suma de  $k$  ( $k$  entero) exponenciales

es Gamma (o Erlang) de parámetro  $k$  para su generación basta  $x = -\frac{1}{\alpha} \sum_{i=1, \dots, k} \ln(1-r_i)$

Inconveniente: Es necesario conocer la función de distribución (en ocasiones algo difícil)

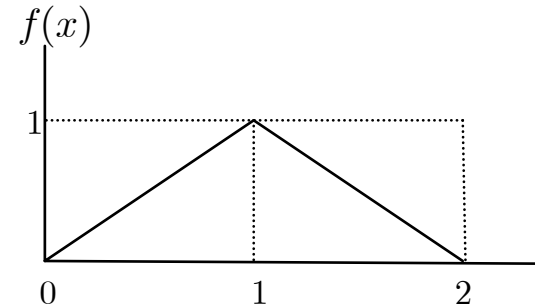




## Método simple de aceptación - rechazo

Distribución triangular (cualquiera dentro de un rectángulo)

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 1 - (x - 1) & 1 \leq x \leq 2 \\ 0 & \text{fuera de } [0,2] \end{cases}$$



Procedimiento:

1. Generar  $r_1 = U(0,1)$  y  $r_2 = U(0,1)$
2.  $x = 2r_1$      $y = r_2$
3. Aceptar  $x$  si  $y \leq f(x)$ , rechazar si  $y > f(x)$  y volver a paso 1

El rectángulo puede tener altura  $c$  igual a  $\text{Max}\{f(x) : x \in (a_1, a_2)\}$  (en el ejemplo  $c=1$ ) o puede ser mayor o igual que dicho valor.

Inconveniente: Es necesario encontrar un rectángulo envolvente (no es posible para distribuciones con soporte no acotado como la normal)

En general para cualquier función  $f(x)$  con soporte en  $(a_1, a_2)$  el procedimiento sería:

1. Generar  $r_1 = U(0,1)$  y  $r_2 = U(0,1)$

2.  $x = a_1 + (a_2 - a_1)r_1$   $y = cr_2$

3. Aceptar  $x$  si  $y \leq f(x)$ , rechazar si  $y > f(x)$  y volver a paso 1

Siendo  $c \geq \text{Max}\{f(x) : x \in (a_1, a_2)\}$

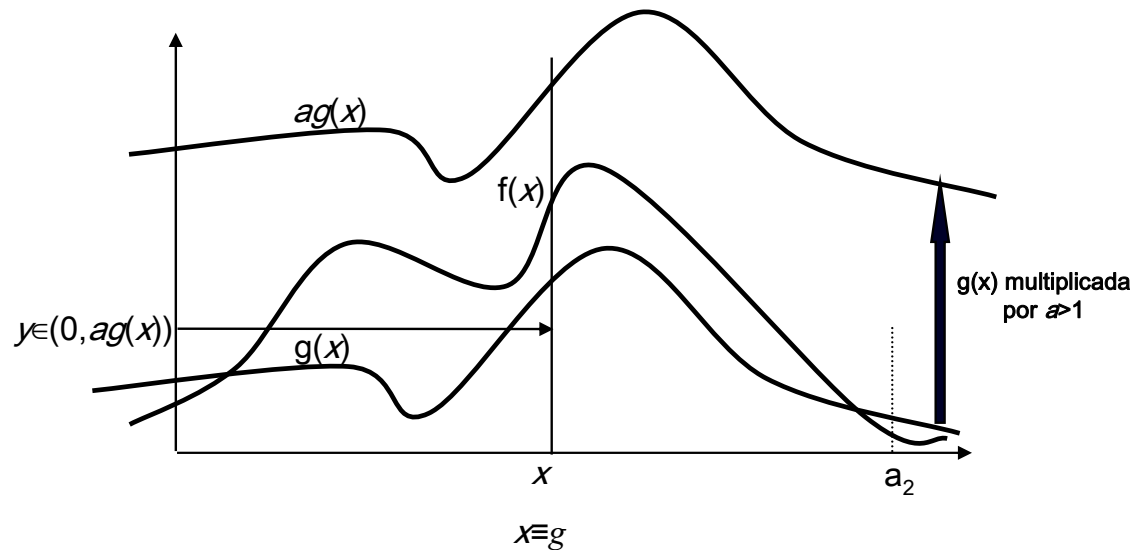
## Método generalizado de aceptación - rechazo

El soporte de la función de densidad  $f(x)$  puede ser no acotado.

Se considera una función (envolvente)  $g(.)$  tal que multiplicada por una constante  $a > 1$  de valores superiores o iguales a los de  $f(.)$  en cada punto.

Procedimiento:

1. Se genera un valor para  $x$  a partir de  $g(.)$
2. Se genera un valor para  $y$  a partir de una  $U(0, a g(x))$
3. Se calcula el valor de  $f(x)$  y si  $y > f(x)$  se vuelve a 1.
4. Si  $y \leq f(x)$  se considera que  $x$  se distribuye según  $f(x)$ .



Nótese que este método generaliza al simple de aceptación-rechazo para funciones de densidad  $f(x)$  acotadas (por ejemplo con soporte  $(a_1, a_2)$ ) sin más que hacer  $g(x) = 1/(a_2 - a_1)$  (es decir,  $g \equiv U(a_1, a_2)$ ) y escogiendo  $a$  tal que  $a \geq \text{Max} \left\{ \frac{f(x)}{g(x)} : x \in (a_1, a_2) \right\}$ , es decir:

$$a \geq (a_2 - a_1) \cdot \text{Max} \{ f(x) : x \in (a_1, a_2) \} \quad (\text{luego } c = a / (a_2 - a_1))$$

Inconveniente: para cada distribución hay que determinar una envolvente apropiada y esto no es nada fácil (ver un ejemplo para la normal en apuntes de la asignatura).

## Método de la distribución normal

No permite aplicar el método de la transformada inversa porque la función de distribución de una normal no tiene expresión analítica exacta (aunque existen aproximaciones analíticas).

### *Método del teorema central del límite*

$X_1, \dots, X_n, \dots$  v.a.i.i.d. con media  $\mu$  y desviación típica  $\sigma$ ,  $\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \rightarrow N(0,1)$  cuando  $n \rightarrow \infty$ .

Caso particular:  $U_i \equiv X_i \equiv U(0,1)$       media  $\mu = 1/2$       desviación estándar  $\sigma = 1/\sqrt{12}$

$$\frac{\sum_{i=1}^n U_i - n/2}{\sqrt{n/12}} \rightarrow N(0,1)$$

Caso ejemplo: tomo  $n=12$        $\left[ \sum_{i=1}^n U_i - 6 \right] \rightarrow N(0,1)$

### *Método del Box-Müller*

Este método se basa en una transformación de variables aleatorias según la cual si  $u_1, u_2$  son v.a.i.i.d., entonces  $x = \sqrt{-2 \ln u_1} \cos(2\pi u_2)$  e  $y = \sqrt{-2 \ln u_1} \sin(2\pi u_2)$  son v.a.i.i.d. con distribución  $N(0,1)$ . Así el procedimiento a seguir sería:

1. Generar  $u_1, u_2 \in U(0,1)$

2. Salida

$x = \sqrt{-2 \ln u_1} \cos(2\pi u_2)$  o bien  $y = \sqrt{-2 \ln u_1} \sin(2\pi u_2)$  son v.a.i.i.d.  $N(0,1)$