

TEMA 6: SIMULACIÓN

INTRODUCCIÓN: Se puede definir *simulación* como un proceso numérico diseñado para experimentar el comportamiento de cualquier *sistema*, conjunto de componentes y acciones que interaccionan, en una computadora a lo largo del tiempo. El comportamiento del sistema se presenta a base de *modelos matemáticos y lógicos*, diseñados para tal fin. El proceso de simulación, será de utilidad cuando se dificulta o imposibilita la resolución del modelo analítico o numérico requerido en un determinado problema.

Los métodos de simulación, permiten estudiar el sistema real sin deformarlo. Los modelos analíticos y numéricos requieren la simplificación del sistema real de estudio, a fin de que se apegue a las condiciones que fundamentan la teoría del modelo en uso, por eso, finalmente muchos modelos analíticos y numéricos resuelvan un sistema deformado muy lejano al sistema real bajo estudio. En contra, los procesos de simulación no producen resultados óptimos, sino simplemente buenos. Son aproximaciones sujetas al error aleatorio subyacente en el modelo.

EJEMPLO: Considérese un ordenador simplificado, compuesto por un sistema de entrada –salida (E/S) y una unidad central de proceso (CPU). El ordenador falla cuando lo hace alguno de los dos componentes. Estudiar el tiempo medio hasta el fallo del sistema.

SOLUCIÓN: Si denotamos por X_1 , tiempo hasta el fallo de E/S, X_2 el tiempo hasta el fallo de la CPU y T el tiempo hasta el fallo del ordenador, se tiene que:

$$T = \min\{X_1, X_2\}$$

Habitualmente, habrá incertidumbre sobre los tiempos anteriores por lo que consideremos que X_1 , X_2 y T son variables aleatorias. Por simplicidad, supóngase que las variables X_1 y X_2 siguen modelos exponenciales independientes con parámetros μ_1 y μ_2 , respectivamente, es decir, con función de densidad y distribución siguientes:

$$f_{X_i} = \mu_i e^{-\mu_i x}, \quad F_{X_i}(x) = P(X_i \leq x) = 1 - e^{-\mu_i x}, \quad i = 1, 2$$

Así pues, se quiere calcular:

$$E[T] = \int_0^{\infty} \int_0^{\infty} \min\{x_1, x_2\} \mu_1 e^{-\mu_1 x_1} \mu_2 e^{-\mu_2 x_2} dx_1 dx_2$$

Aproximación analítica: Teniendo en cuenta la independencia de las variables, así como su distribución, se tiene que:

$$P(T > t) = P(X_1 > t, X_2 > t) = P(X_1 > t)P(X_2 > t) = e^{-(\mu_1 + \mu_2)t}$$

es decir, T sigue un modelo exponencial con parámetro $\mu_1 + \mu_2$, por tanto:

$$E[T] = \frac{1}{\mu_1 + \mu_2}$$

Si $\mu_1 = 1$ y $\mu_2 = 2$, entonces el tiempo medio de fallo es $E[T] = 1/3$. Se obtiene la solución analítica exacta, pero resulta esencial emplear hipótesis de independencia y exponencialidad, de manera que el resultado es poco robusto. Si cambiamos alguna de estas hipótesis, el cálculo exacto puede resultar mucho más complicado, e incluso, imposible de obtener mediante un método analítico.

Aproximación numérica: Una posibilidad, en tal caso, es apelar a algún procedimiento de integración numérica. Para ello, un paso inicial será reparametrizar el problema de manera que la región de integración sea acotada. Bastará realizar el cambio $x_i = -(\ln(1 - u_i)) / \mu_i$, obteniéndose:

$$E[T] = \int_0^1 \int_0^1 \min\left\{-\frac{\ln(1 - u_1)}{\mu_1}, -\frac{\ln(1 - u_2)}{\mu_2}\right\} du_1 du_2$$

Se puede por ejemplo aplicar la regla del trapecio. En nuestro caso con 15 nodos, $\mu_1 = 1$ y $\mu_2 = 2$, se obtiene 0.317 como aproximación del tiempo esperado hasta el fallo. Obsérvese que la estrategia numérica adoptada es más robusta, en el sentido de que si sustituimos alguna de las dos hipótesis, siguen siendo aplicable este método. El inconveniente estará en que el aumentar la dimensión de la región de integración (variables), la aproximación resultará cada vez más ineficaz.

Aproximación basada en simulación: La aproximación basada en simulación consistirá en construir un programa que describa el comportamiento del sistema y realizar experimentos con él

$$X_1 \mapsto \exp\{\mu_1\}$$

$$X_2 \mapsto \exp\{\mu_2\}$$

n : repeticiones del sistema, $(X_{1i}, X_{2i}) \rightarrow T_i = \min\{X_{1i}, X_{2i}\}, i = 1, \dots, n$

Programa de simulación:

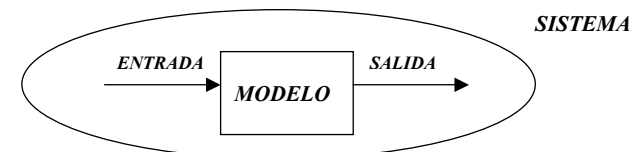
```
tiempo <- function(n, nu1, nu2) {X1 <- rexp(n, nu1);
                               X2 <- rexp(n, nu2);
                               apply(cbind(X1, X2), 1, min) }
```

```
res <- tiempo(10000, 1, 2); mean(res)
```

Aproximaremos $E[T] \cong \frac{\sum_{i=1}^n T_i}{n}$. Si consideramos $n=10000$ obtenemos que

$E[T] = 0.33669$. Darse cuenta que si cogemos otra replicación del experimento se obtiene una estimación diferente. Así mismo, “mejor” será la aproximación si simulamos más repeticiones del sistema.

COMENTARIO: Teniendo en cuenta lo anterior, podemos concluir que la simulación es una metodología de análisis de sistemas basada en la construcción de un modelo, típicamente implementado en un ordenador, que describe el comportamiento del sistema y que permite generar observaciones, dadas ciertas entradas. Tales observaciones se analizan estadísticamente para estimar medidas del comportamiento del sistema de interés.



Los modelos de simulación, pueden diferir en gran medida, dependiendo de si el espacio de estado es discreto o continuo, las observaciones estáticas o dinámicas; como función continua o discreta del tiempo. A pesar de la gran variedad de posibilidades existentes, la mayoría de los experimentos de simulación, una vez construido el correspondiente modelo, se adaptan al siguiente esquema:

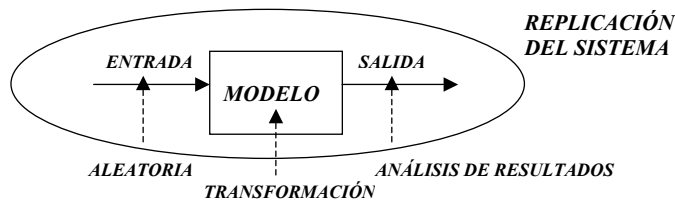
- *Generación de entradas* al modelo según las especificaciones del mismo. Modelos de probabilidades asociados.
- *Transformación de las entradas*, a través del modelo, en salidas.
- *Calcular estadísticas* a partir de las salidas, con el fin de describir el comportamiento del sistema y estimar las medidas de comportamiento de interés.

ETAPAS BÁSICAS EN UN PROCESO DE SIMULACIÓN: Como etapas básicas en todo proceso de simulación, destacan la modelización del sistema, replicación del modelo mediante simulación y análisis de resultados. A continuación, se detalla brevemente cada una de ella.

CONSTRUCCIÓN DEL MODELO, MODELIZACIÓN: El primer paso en un estudio de simulación es elaborar un modelo que represente el sistema que se va a investigar. Es decir, al modelizar, se caracterizan matemáticamente las relaciones que gobiernan las interacciones de los componentes del sistema y de sus actividades. La descripción del sistema se debe hacer de forma modular y por bloques. El conjunto de todos los módulos es el sistema bajo estudio. El sistema como un todo se modela matemáticamente de acuerdo a la interconexión de sus bloques. Cada bloque debe contener únicamente la información relevante.

REPLICACIÓN DEL PROCESO MEDIANE SIMULACIÓN: Una vez determinado el modelo que describa fielmente el sistema y con el fin de obtener una replicación del mismo, se tendrá que determinar la distribución de probabilidad de los estados del sistema de tal manera que el generar la evolución del sistema a partir de este estado se obtenga una representación fiel de su comportamiento. Todo ello se llevará a cabo en un computador mediante un programa diseñado para tal fin.

ANÁLISIS DE RESULTADOS: Iterando replications del proceso, se obtiene una colección de salidas del sistema recogidas sistemáticamente mediante simulación. A partir de esta información y mediante métodos estadísticos, obtendremos conclusiones del comportamiento del sistema.



GENERACIÓN DE NÚMEROS ALEATORIOS, VARIABLES ALEATORIAS: Los números aleatorios son utilizados para introducir el comportamiento estocástico en el sistema bajo estudio. Un *generador de números aleatorios* es un algoritmo que produce sucesiones de números que siguen una distribución uniforme y que posee la apariencia de aleatorios. La referencia a una sucesión de números, significa que el algoritmo produce muchos números aleatorios en forma serial. Los números generados en una computadora, no son totalmente aleatorios, porque se obtienen de fórmulas preestablecidas. Por esta razón, a estos números se les denominan *pseudoaleatorios*. La mayoría de los métodos para generar números aleatorios son iterativos, donde un número pseudoaleatorio se genera del anterior. El *periodo* del método es el número de generaciones que se debe esperar hasta repetirse la secuencia. Es deseable que este periodo sea lo mayor posible.

Un buen número de técnicas de generación de números aleatorios pertenecen a la familia de los *métodos multiplicativos congruentes*, donde siguen la siguiente fórmula recursiva:

$$x_{n+1} = (ax_n + b) \text{ mod } m. \text{ Uniforme en } \{0, \dots, m-1\}$$

$$u_n = \frac{x_n}{m} \in [0,1], \text{ Uniforme en el intervalo } [0,1]$$

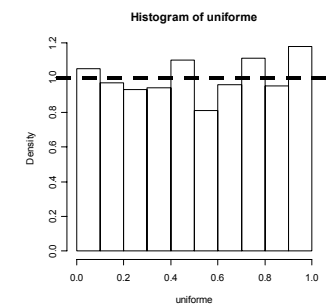
para un multiplicador a , sesgo b , módulo m y semilla x_0 , con $a, b \in \{0, \dots, m-1\}$. Si $b=0$ se denominan métodos congruenciales multiplicativos. A pesar de su simplicidad y previsibilidad, una selección cuidadosa de los parámetros (a, b, m) permiten obtener de manera eficiente sucesiones de números suficientemente largas y aleatorias para muchos propósitos.

Uno de los generadores de números aleatorios basados en el método congruencial multiplicativo más utilizado dada la arquitectura de 32 bits de las computadoras, es debido a Larnmouth-Lewis, donde $m = 2^{31} - 1$, $a = 7^5$ y $b = 0$. Este generador posee las siguientes características:

- Periodo máximo
- Salida parecida a aleatoria
- Se implementa de forma eficiente en aritmética de 32 bits.

EJEMPLO:

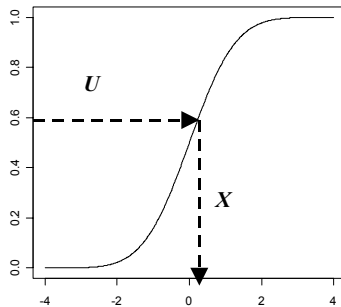
`runif(1000, 0, 1)`



A partir de la simulación de números aleatorios, se obtendrán números generados a partir de las principales distribuciones de probabilidad. Para ellos distinguiremos entre variables aleatorias continuas y discretas. Uno de los métodos más utilizado es el denominado *método de inversión*.

MÉTODO DE INVERSIÓN: Sea X una variable aleatoria con función de distribución $F(x) = P(X \leq x)$. Si X es una *variable aleatoria continua tal que exista F^{-1}* , inversa de la función de distribución, entonces se pueden generar valores aleatorios de la distribución F a partir de una variable aleatoria uniforme en el intervalo $[0,1]$, pues:

$$F(X) = U \Rightarrow X = F^{-1}(U)$$



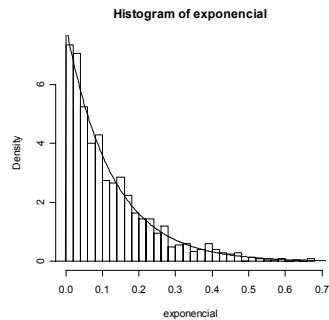
EJEMPLO: Considérese una variable aleatoria X exponencial con parámetro λ , es decir $F(x) = 1 - e^{-\lambda x}$, entonces:

$$u = 1 - e^{-\lambda x} \Rightarrow x = -\frac{1}{\lambda} \ln(1 - u)$$

Programa de simulación:

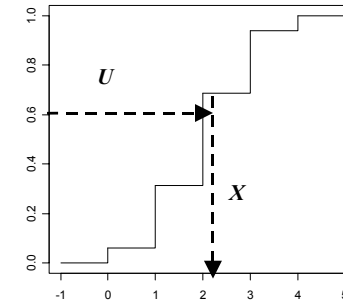
```
rexponencial<-function(n,landa){-(1/landa)*log(1-runif(n)) }
```

```
rexponencial(10000,8)
```

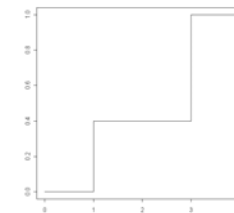


Si la variable X es *discreta*, entonces F es escalonada y se genera la valores como sigue:

$$\min\{X : F(X) \geq U\}$$



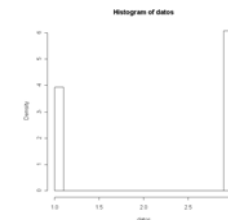
EJEMPLO: Considérese una variable aleatoria X dicotómica con valores 1 y 3 y probabilidades 0.4 y 0.6 respectivamente. Su función de distribución tiene la siguiente forma:



Programa de simulación:

```
rnp<-function(n, estados, prob){
  comparacion <- function(a, estados, prob) {
    cumprob <- c(0, cumsum(prob))
    estados[sum(cumprob <= a)]
  }
  apply(cbind(runif(n)),1,comparacion,estados, prob)
}
```

```
rnp(10000,c(1,3),c(.4,.6))
```



Una distribución de gran interés, es la *distribución normal*. Para esta distribución, apoyándose en el *Teorema Central del Límite*, permite obtener otra forma de simulación, como sigue:

$$U \mapsto \text{Uniforme}[0,1] \Rightarrow E[U]=1/2, \text{Var}[U]=1/12$$

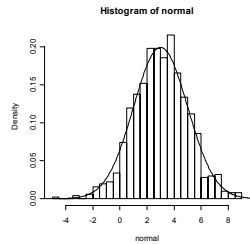
$$\text{Teorema Central del Límite } \frac{\sum_{i=1}^N U_i - N/2}{\sqrt{N/12}} = X \mapsto N(0,1)$$

Es suficiente considerar $N=12$, obteniéndose: $\sum_{i=1}^{12} U_i - 6 = X \mapsto N(0,1)$. Además si $Y \mapsto N(\mu, \sigma^2)$, entonces $Y = \sigma X + \mu$.

Programa de simulación:

```
rnormal<-function(n,nu=0,var=1){A<-matrix(runif(12*n),n,12)
  sqrt(var)*(apply(A,1,sum)-6)+nu}
```

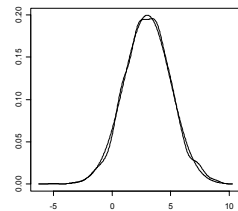
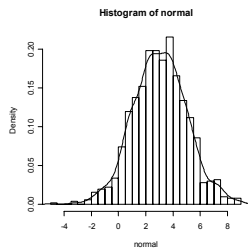
```
rnormal(1000,3,4)
```



NOTA: Una vez generado un conjunto de datos a partir de cierta variable aleatoria X , sería conveniente comprobar que efectivamente estos datos se pueden suponer extraídos aleatoriamente según el modelo de probabilidad definido por X . Una primera comprobación descriptiva, consistirá en comparar visualmente el histograma con la función de densidad, caso continuo, o función de probabilidad, caso discreto, teórica. Inclusive se puede obtener la función de densidad estimada a partir del conjunto de datos.

NORMAL

HISTOGRAMA CON FUNCIÓN DE DENSIDAD ESTIMADA FUNCIÓN DE DENSIDAD TEÓRICA



Finalmente, con el fin de “reafirmar” nuestra sospecha, utilizaremos un test de hipótesis:

Contraste χ^2.Bondad de ajuste:	Cualquier tipo de distribución (Discretas)
$H_0 : F = F_0$	$H_1 : F \neq F_0$
Kolmogorov-Smirnov:	Distribuciones continuas
$H_0 : F = F_0$	$H_1 : F \neq F_0$
Shapiro-Wilk:	Distribuciones Normales
$H_0 : F = F_0$	$H_1 : F \neq F_0$

EJEMPLO:

```
Shapiro-Wilk normality test      Kolmogorov-Smirnov test
data: normal                              data: normal
W = 0.998, p-value = 0.2879      D = 0.0205, p-value = 0.7955
Kolmogorov-Smirnov test
data: exponencial
D = 0.0221, p-value = 0.7146
```

A la luz de estos resultados, se concluye que....

ANÁLISIS DE RESULTADOS DE LA SIMULACIÓN: Una vez construido el modelo que describe el sistema y obtenido un conjunto de datos a través de simulación del modelo mediante la generación de distribuciones de probabilidades adecuadas, nos centraremos en analizar los resultados obtenidos. El objetivo es proporcionar información sobre una medida θ del comportamiento del sistema a partir de los datos recogidos.

En primer lugar, sería interesante representar gráficamente el conjunto de datos, por ejemplo un histograma con el fin de poder determinar, si es posible, el modelo de probabilidad subyacente, utilizando para ello la estimación de la función de densidad y los test de hipótesis, para tal fin.

A continuación, se obtendrá mediante métodos inferenciales estadísticos, estimaciones de la característica θ de interés, tanto puntuales como intervalos de confianza, incluso llegando el caso, utilizar test de hipótesis apropiados.

Una característica comúnmente estudiada es la media: $\theta = E[X]$, donde X representa la variable aleatoria salida del sistema. Es bien conocido, que un estimador con buenas propiedades, es el estimador media muestral, es decir, dado n realizaciones del sistema $\{x_1, \dots, x_n\}$, se define como:

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i}{n}$$

Así mismo, se pueden obtener intervalos de confianza al nivel $1-\alpha \in (0,1)$, siguientes:

$$\theta = E[X] \in \left(\hat{\theta} - t_{n-1,\alpha} \sqrt{S^2/n}, \hat{\theta} + t_{n-1,\alpha} \sqrt{S^2/n} \right)$$

donde S^2 es el estimador varianza muestral y $t_{n-1,\alpha}$ es el cuantil $1-\alpha/2$ de la distribución *T-Student* con $n-1$ grado de libertad, es decir:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \hat{\theta})^2}{n-1} \text{ y } P(T > t_{n-1,\alpha}) = \alpha/2, \text{ donde } T \mapsto T\text{-Student}(n-1)$$

EJEMPLO: Consideremos el ejemplo de determinar el tiempo medio de fallo de un computador, donde: $X_1 \mapsto \exp\{\mu_1\}$, $X_2 \mapsto \exp\{\mu_2\}$ y se realizan n : repeticiones del sistema, siendo $(X_{1i}, X_{2i}) \rightarrow T_i = \min\{X_{1i}, X_{2i}\}, i=1, \dots, n$. Es conocido, que en esta situación, $T \mapsto \exp\{\mu_1 + \mu_2\}$ y $E[T] = \frac{1}{\mu_1 + \mu_2}$.

Como $\mu_1 = 1$ y $\mu_2 = 2$, entonces $T \mapsto \exp\{3\}$ y $E[T] = \frac{1}{3} = 0.3333$. Vamos a obtener estas conclusiones mediante simulación. Para ello replicamos 10.000 veces el sistema implementado obteniéndose los siguientes resultados:

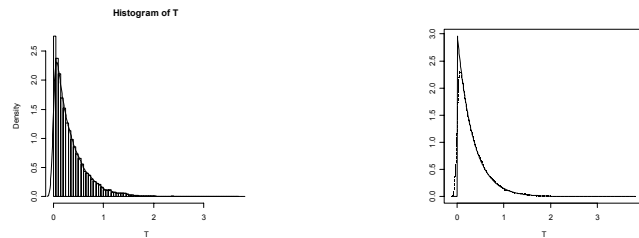
Primeras simulaciones: 0.36420232, 0.07388238, 0.16026809, 0.61732036, 0.20432946, 0.01899736, 0.58828299, 0.50356364, 0.04770036, 0.84546350,....

Histograma con función de densidad estimada, se contrasta con la teórica de T:

```
hist(T, prob=T, br=100)
```

```
lines(density(T))
```

Histograma con función de densidad estimada Función de densidad teórica y estimada



Test de hipótesis para contrastar la distribución de los datos observados:

```
ks.test(T, "pexp", 3)
```

```
Kolmogorov-Smirnov test
```

```
data: T, D = 0.0082, p-value = 0.5064
```

Estimación de la media con intervalo de confianza al 95%:

$E[T] \cong \hat{\theta} = 0.33669$, $Var[T] \cong S^2 = 0.1126527$, entonces el intervalo de confianza al 95% es:

$$0.3333 = \theta = E[X] \in \left(\hat{\theta} - t_{n-1,\alpha} \sqrt{S^2/n}, \hat{\theta} + t_{n-1,\alpha} \sqrt{S^2/n} \right) = (0.3301116, 0.3432683)$$

Darse cuenta que aumentando el n mejor será la aproximación. Asimismo, podríamos estudiar el tiempo medio de fallo si el parámetro $\mu_i \in [1,2]$, obteniéndose lo siguiente:

