2. GENERACIÓN NUMEROS ALEATORIOS

- GENERAR números aleatorios de una cierta distribución de probabilidad significa obtener VALORES DE LA DISTRIBUCIÓN que se puedan considerar INDEPENDIENTES.

- En un proceso en tres fases:
  1. Obtención de los valores
  2. Comprobación de que pertenecen a la distribución elegida (test de hipótesis)
  3. Comprobación de que pueden considerarse independientes (test de hipótesis)

GENERACIÓN NUMEROS ALEATORIOS

- La obtención de los valores se hace en dos pasos:
  1. Obtención de valores de una distribución U(0,1)
  2. Transformación de los valores anteriores en valores de la distribución elegida.
- Estudiaremos las dos fases por separado.

GENERACIÓN NUMEROS ALEATORIOS U(0,1)

- Por orden cronológico:
  - Extracción con reemplazamiento de bolas de una urna o similares.
  - Método cuadrado medio (1940).
  - Métodos congruenciales lineales (1951).
  - Métodos congruenciales multiplicativos módulo primo (1962).
  - Métodos múltiplemente recursivos (1990).
  - Combinación de métodos de los dos últimos tipos.

GENERACIÓN NUMEROS ALEATORIOS U(0,1)

- Método más antiguo:
  - Elección, con reemplazamiento, de números del 0 al 9 sacando bolas numeradas de una bolsa.
  - Si se quieren, por ejemplo, 4 decimales, se extraen 4 bolas y se pone un punto decimal delante.

Ventaja: los números obtenidos son totalmente independientes.

Inconveniente: la generación es muy lenta. Se buscan métodos mecanizados.

GENERACIÓN NUMEROS ALEATORIOS U(0,1)

- Método cuadrado medio (1940) Von Newmann:
  - Se elige un entero de cuatro cifras, Z_0.
  - Cada valor Z_{i+1} se obtiene a partir de Z_i tomando los cuatro dígitos centrales de Z_i^2. Si Z_i^2 no tiene ocho dígitos se completa con ceros a su izquierda hasta que sí los tenga.
  - Se pone un punto decimal delante obteniéndose el valor U_i, número aleatorio de una U(0,1)

MÉTODO CUADRADO MEDIO

Z_i	Z_i^2	U_i
7182	51581124	0.5811
5811	33767721	0.7677
7677	58936329	0.9363

- **Inconveniente**: la generación no es independiente en absoluto y la secuencia presenta una fuerte tendencia a 0.

GENERACIÓN NUMEROS ALEATORIOS U(0,1)

- TODOS los generadores que se usan tienen este inconveniente: los números NO son independientes.
- Sin embargo, si se generan con cuidado, pueden obtenerse números que PARECEN independientes (pasan los test de hipótesis)
- Por este motivo, los números no se llaman aleatorios sino PSEUDOALEATORIOS.

MÉTODOS CONGRUENCIALES LINEALES (1951; Lehmer)

- Se eligen cuatro enteros iniciales: Z_0 (semilla), m (módulo), a (multiplicador) y c (incremento); $0 < Z_0, a, c < m$.
- Se usa la fórmula recursiva:

$$Z_{i+1} = \left(aZ_i + c\right) \mathbf{mod}\, m$$

- Se define U_i como: $$U_i = \frac{Z_i}{m}$$

OBJECCIONES A LOS MÉTODOS CONGRUENCIALES LINEALES

- Cada valor de la secuencia está totalmente determinado desde el principio a partir de Z_0, m, a y c:

$$Z_i = \left[a^i Z_0 + \frac{c\left(a^i - 1\right)}{a - 1} \right] \mathbf{mod}\, m$$

- Los valores U_i solamente pueden ser 0, 1/m, 2/m,..., (m-1)/m y si m es pequeño hay muchos números del intervalo [0,1] que no pueden salir. Esto se resuelve tomando m grande.
- Ejemplo con Z_0= 7, m = 16, a = 5 y c = 3.

i	Z_i	U_i	i	Z_i	U_i
0	7		11	0	0
1	6	0.375	12	3	0.188
2	1	0.063	13	2	0.125
3	8	0.500	14	13	0.813
4	11	0.688	15	4	0.250
5	10	0.625	16	7	0.438
6	5	0.313	17	6	0.375
7	12	0.750	18	1	0.063
8	15	0.938	19	8	0.500
9	14	0.875	20	11	0.688
10	9	0.563	21	10	0.625

OBSERVACIONES

- El comportamiento cíclico de este ejemplo es INEVITABLE.
- La longitud del ciclo se llama PERIODO del generador. Claramente, el periodo siempre es menor o igual que m.
- Si el periodo es m se dice que el generador tiene PERIODO COMPLETO.
- Como en una simulación se usan cientos de miles de números aleatorios interesa que los generadores tengan periodo lo más largo posible y, si se puede, que tengan periodo completo.

CARACTERÍSTICAS DE UN BUEN GENERADOR

- Los números generados deben de parecer independientes y venir de U(0,1) (pasar los test).
- Deben de ser rápidos y necesitar poca memoria (almacenamiento).
- Deben de poder producir secuencias largas (periodo máximo m) para asegurar que en una secuencia de m números aleatorios, cada uno sólo se repite una vez.
- Deben de poder reproducir la misma secuencia de números aleatorios.
- Debe de poder producir secuencias independientes para poder modelizar las distintas fuentes de aleatoriedad del sistema.

MÉTODOS CONGRUENCIALES LINEALES CON PERIODO COMPLETO

- **<u>TEOREMA:</u>** (Hull-Dobell, 1962) un método congruencial lineal tiene periodo completo m sí y solo sí:
  1. El único entero positivo que divide de manera exacta a m y c es 1 (m.c.d.{m,c}=1).
  2. Si q es un número primo que divide a m, entonces q tiene que dividir a (a-1).
  3. Si 4 divide a m, entonces 4 divide a (a-1).

MÉTODOS CONGRUENCIALES LINEALES CON PERIODO COMPLETO

- La condición 1 del teorema hace que sea diferente el trato de los métodos en los que $c = 0$ o $c \neq 0$.
- Si $c \neq 0$: MÉTODOS MIXTOS: se puede verificar la condición 1 del teorema y obtener periodo completo.
- Si $c = 0$: MÉTODOS MULTIPLICATIVOS: no se puede obtener periodo completo pero son más sencillos de implementar.
- Se estudiaron en primer lugar métodos mixtos (puede obtenerse periodo completo) pero los números obtenidos no tenían buenas propiedades estadísticas.

MÉTODOS CONGRUENCIALES LINEALES MULTIPLICATIVOS

- Elección de m:
  - dividir por m para obtener el resto es una operación relativamente lenta. Es deseable no tener que hacer la división en forma explícita.
  - Si b = longitud de palabra del ordenador, eligiendo $m = 2^b$ se puede evitar la división utilizando el overflow de enteros.

MÉTODOS CONGRUENCIALES LINEALES MULTIPLICATIVOS

- Ejemplo:
  - $Z_0 = 5$, b = 4 ($m = 2^4 = 16$), a = 5 y c = 3.
  - ¿Cómo obtener $Z_1 = 12$ sin hacer la división entre m = 16?
  - $5\ Z_0 + 3 = 28$. Su representación binaria es 11100.
  - Como solo se pueden almacenar 4 dígitos, se pierde el de la izquierda, resultando 1100, que es la representación binaria de $Z_1 = 12$.

MÉTODOS CONGRUENCIALES LINEALES MULTIPLICATIVOS

- No se puede obtener periodo completo m pero con una elección adecuada de m y a se puede obtener periodo m-1 (casi completo).
- La elección de $m = 2^b$ parece adecuada pero solo se obtiene un periodo máximo de $2^{b-2} =$ $= 2^b /4$ como mucho (sólo se obtiene la cuarta parte de los valores que podrían obtenerse en (0,1)).
- Consecuencia: no se obtienen valores de una U(0,1).
- Se observó que si se elige m primo, se obtienen mejores resultados. Esto llevó a los métodos congruenciales lineales multiplicativos módulo primo.

MÉTODOS CONGRUENCIALES LINEALES MULTIPLICATIVOS MÓDULO PRIMO

TEOREMA (Hutchinson-Lehmer, 1966):

1. Si m es primo
2. Elegimos a de manera que el menor entero, k, para el cual $a^k - 1$ es divisible por m es k = m-1
3. $Z_0 <$ m-1, entero,

   entonces, el método congruencial multiplicativo resultante tiene periodo (m-1).

MÉTODOS CONGRUENCIALES LINEALES MULTIPLICATIVOS MÓDULO PRIMO

- El caso de longitud de palabra b = 32 bits ha sido muy estudiado porque muchos ordenadores y compiladores en uso funcionan con esta longitud de palabra.
- Para este caso, tomando $m = 2^{31} - 1$ hay 534 millones de números que cumplen la condición para a. Funcionan bien $a = 7^5$ o
  a = 630.360.016.
- Todos los lenguajes de simulación tienen implementados métodos congruenciales lineales multiplicativos módulo primo.

MÉTODOS CONGRUENCIALES MÚLTIPLEMENTE RECURSIVOS

- Son métodos propuestos en la década de los 90. La secuencia obtenida satisface la relación recursiva:

$$Z_{i+1} = \left(a_0 Z_i + a_1 Z_{i-1} + a_2 Z_{i-2} + \ldots + a_k Z_{i-k}\right) \mathbf{mod}\, m$$

- Donde m, k, a_i enteros positivos, con
  $0 < a_i < m-1$.
- El máximo periodo conseguido con estos generadores es $m^k - 1$, que se alcanza cuando m es primo y y el polinomio característico de la relación recursiva anterior verifica una serie de condiciones.

MÉTODOS CONGRUENCIALES MÚLTIPLEMENTE RECURSIVOS

- Con estos métodos se obtienen secuencias mucho mayores que con los métodos congruenciales lineales (se pasa de periodo m-1 a periodo m^k –1)
- Sin embargo, las propiedades estadísticas de la secuencia siguen sin ser óptimas.
- Éstas mejoran si se trabaja con una combinación de métodos conguenciales lineales (MacLaren, 1965))o métodos múltiplemente recursivos (L´Ecuyer, 1996).
- Pascal utiliza una combinación de métodos congruenciales lineales para generar números aleatorios de U(0,1).

TEST DE HIPÓTESIS

- Una vez generada la secuencia de números aleatorios es FUNDAMENTAL verificar que cumple las propiedades estadísticas requeridas:
  - Los números obtenidos son independientes: TEST DE INDEPENDENCIA.
  - Los números aleatorios vienen de U(0,1): TEST DE DISTRIBUCIÓN (contraste de Kolmogorov- Smirnov)

GENERACIÓN NUMEROS ALEATORIOS

- Vimos que la obtención de números aleatorios de una distribución se hace en dos pasos:
  1. Obtención de valores de una distribución U(0,1) (YA HECHO)
  2. Transformación de los valores anteriores en valores de la distribución elegida.
- En la fase 2, hay métodos específicos para cada distribución.
- Sólo estudiaremos los que nos hacen falta para realizar el trabajo: distribuciones exponencial y Erlang (continuas) y distribución discreta cualquiera.

Distribución exponencial

- Método de la transformada inversa: Si U es una v.a. con distribución U(0,1) y X es una v.a. con función de distribución F(x), entonces, la v.a. definida por $F^{-1}(U)$ tiene función de distribución F(x).
- Si se desean generar un número aleatorio de una exponencial con parámetro β,x,
  1. Generar un número aleatorio de U(0,1), u, según el método elegido (en Pascal, u:= random);
  2. Aplicar la fórmula :

$$x := \frac{-1}{\beta}\ln(u)$$

Distribución Erlang

- Reproductividad: si X_i , i=1,..,n, son v.a. exponenciales de parámetro β, independientes

$$X_1 + X_2 + ... + X_n \sim \gamma(n, \beta) = Erlang(n, \beta)$$

- Si se desea generar un número aleatorio de una Erlang (n,β),x,
  1. Generar n nº aleatorios de U(0,1), u_1, u_2, ..., u_n, según el método elegido (en Pascal, u_i: = random, i = 1,..,n);
  2. Aplicar la fórmula :

$$x := \frac{-1}{\beta} \sum_{i=1}^{n} \ln u_i = \frac{-1}{\beta} \ln \prod_{i=1}^{n} u_i$$

Distribución discreta

- Se desea generar un número aleatorio,x, de una variable discreta que toma n valores,

$$\{x_1, x_2, ..., x_n\}, x_1 < x_2 < ... < x_n$$

con probabilidades $\{p_1, p_2, ..., p_n\}$

Paso 1.- Dividir el intervalo [0,1] en n subintervalos cada un de longitud p_i (cerrados por la izquierda y abiertos por la derecha).

Paso 2.- Generar un nº aleatorio de una U(0,1), u, según el método elegido.

Paso 3.- Si u cae en el intervalo i-ésimo, hacer $x = x_i$.

Ejemplo

- Se desea generar un número aleatorio,x, de una variable discreta que toma 4 valores,

  $\{0,1,2,3\}$ con probabilidades $\left\{\frac{1}{6},\frac{1}{3},\frac{1}{4},\frac{1}{4}\right\}$

  Paso 1.- El intervalo [0,1] se divide en los subintervalos [0,1/6), [1/6,1/2), [1/2, 3/4) y [3/4, 1).

  Paso 2.- Se generan valores de una U(0,1), u_i:

  u_1= 0.2365, u_2= 0.9763, u_3= 0.1517,....

  Paso 3.- Si u cae en el intervalo i-ésimo, hacer $x = x_i$.Entonces, x_1=1, x_2= 3, x_3=1,....

Distribución discreta

- **Para el trabajo:** se necesita generar un número aleatorio,x, de una variable discreta que toma dos valores, NODO 2 y NODO 3 con probabilidades 0.4 y 0.6, respectivamente:

  1.- Generar un número aleatorio de U(0,1) (en Pascal, u:= random);

  2.- Si u < 0.4 , hacer x:= NODO 2,

  si u > = 0.4, hacer x:= NODO3.

3.- DECIDIR EN UNA SIMULACIÓN CUANDO SE ALCANZA EL ESTADO ESTACIONARIO

- **Objetivo de la simulación: estimar algún parámetro de rendimiento en estado estacionario**.
- **Periodo estacionario**: momento a partir del cual las condiciones iniciales de la simulación dejan de influir significativamente en los resultados.
- **Periodo transitorio**: desde que comienza la simulación hasta que se alcanza el estado estacionario.
- **Ejemplo**: al simular un proceso de producción, el sistema comienza vacío e interesa estudiar este sistema funcionando a pleno rendimiento.

- **Procedimiento habitual**: no tener en cuenta la parte inicial de la simulación, iniciando la contabilidad después de una cantidad de tiempo "adecuada".
- **Primera idea:** Se realiza un ejecución larga, se parte en trozos y para cada uno, se calcula alguna medida de rendimiento del sistema. Si esta medida se estabiliza a partir del instante t_E , tomar éste como inicio del periodo estacionario.
- **Mejora:** Realizar el proceso anterior varias veces. Obtener de la 1ª ejecución t_{E1} y comenzar el resto de simulaciones con las condiciones que tiene el sistema en el instante t_{E1}, obtener un instante de tiempo para cada ejecución y tomar el mayor de todos ellos como momento en que comienza el periodo estacionario.

4.- ANÁLISIS ESTADÍSTICO DE LOS RESULTADOS

Como los resultados de una simulación son aleatorios, la forma de trabajar es REALIZAR n EJECUCIONES y estimar los parámetros de rendimiento mediante técnicas estadísticas a partir de los resultados para las n ejecuciones del programa.

a) Estimación de medidas de rendimiento en estado estacionario.

b) Intervalos de confianza.

c) Número de ejecuciones (réplicas) a realizar y longitud de las mismas. Obtención de réplicas independientes.

4.-ESTIMACION DE MEDIDAS DE RENDIMIENTO EN ESTADO ESTACIONARIO

- **ESTADÍSTICA:** Sea X la característica en estudio, tal que $E[X]=\theta$, desconocido.
- **Forma de trabajo:** Tomar una m.a.s. (v.a. INDEPENDIENTES)

$$X_1, X_2, ..., X_n \quad y \quad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

- **Manera de medir el error:** intervalo de confianza para θ, con nivel $1-\alpha$.

$$\theta \in \left(\hat{\theta} \pm h\frac{S}{\sqrt{n}} \right),\ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2,\ P\left(-h < t_{n-1} < h\right) = 1-\alpha$$

ESTIMACION DE MEDIDAS DE RENDIMIENTO EN ESTADO ESTACIONARIO

- A partir del intervalo de confianza para θ, se definen:

$$\theta \in \left(\hat{\theta} \pm h \frac{S}{\sqrt{n}} \right)$$

  - **ERROR ABSOLUTO:** $EA = \left| \theta - \hat{\theta} \right| < h \frac{S}{\sqrt{n}}$

  - **ERROR RELATIVO:** $ER = \frac{\left| \theta - \hat{\theta} \right|}{\left| \hat{\theta} \right|} < \frac{h \frac{S}{\sqrt{n}}}{\left| \bar{X} \right|}$

    No depende de las unidades de medida ni de la magnitud de los datos. Se puede medir en % y se llama precisión del intervalo.

ESTIMACION DE MEDIDAS DE RENDIMIENTO EN ESTADO ESTACIONARIO

- **Aplicación en una simulación:** Sea, por ejemplo, L, la característica a estimar (se razona igual para cualquier otro parámetro, p_k o W)

- **Forma de trabajo:** Realizamos n ejecuciones INDEPENDIENTES. En cada una de ellas obtenemos una estimación para L, L_i :

Tenemos $L_1, L_2, ..., L_n$ independientes y $\hat{L} = \frac{1}{n} \sum_{i=1}^{n} L_i = \bar{L}$

- **Error absoluto en la estimación de L:**

$$EA = \left| L - \bar{L} \right| < h \frac{S}{\sqrt{n}}, S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(L_i - \bar{L} \right)^2$$

4.-RELIZACIÓN DE EJECUCIONES INDEPENDIENTES

- **MÉTODO 1:** Ejecutar n veces la simulación, cada una con números aleatorios distintos y descontar en cada una el periodo transitorio.
- **MÉTODO 2:** Realizar una sola simulación larga y partirla en n trozos. Así solo se cuenta una vez el periodo transitorio pero las ejecuciones no son independientes.
- **MÉTODO 3:** Igual que el dos pero dejar entre dos trozos un periodo sin contar, más pequeño que el transitorio. Es el más utilizado.
- **MÉTODO 4** (REGENERATIVO): Localizar instantes en el tiempo en los que el proceso se reinicie (puntos de regeneración).

4.- NÚMERO DE EJECUCIONES Y LONGITUD DE LAS MISMAS

- Se obtienen resultados equivalentes usando menos ejecuciones más largas o más ejecuciones menos largas.
- **Procedimiento:**

  1- Fijar la longitud de las ejecuciones. Fijar el error relativo γ máximo a cometer en la estimación del parámetro de interés, θ.

  2.- Realizar un número n_0 de simulaciones (8 ó 9 en la práctica). Comprobar si la estimación de θ presenta un error relativo menor que γ .

  * si : ER < γ, terminar las ejecuciones
  * sino, $n_0 = n_0+1$ y repetir el proceso.