Generación de secuencias pseudoaleatorias

Andrés Pomi

Taller MMC-Bio 2021

Generación de aleatoriedad en la computadora

- Características de la aleatoriedad: impredictibilidad, irregularidad, aperiodicidad.

- No se pueden generar secuencias verdaderamente aleatorias,

  por lo que hablaremos de pseudoaleatoriedad.

- En efecto, una medida de la bondad de una secuencia pseudo-aleatoria es su aperiodicidad, sin embargo cualquier algoritmo computacional finito sólo puede generar secuencias periódicas.

- Entonces el objetivo será obtener secuencias irregulares de período muy largo.

Objetivo de la simulación del azar

- Generar valores aleatorios de variables con cierta distribución aleatoria: normal, exponencial, etc

- Se logra en dos pasos:

1. Generación de números aleatorios en [0,1] uniformemente distribuidos.

1. Transformar la secuencia aleatoria de distribución uniforme a una secuencia aleatoria con la distribución deseada.

Propiedades de los generadores de números pseudo-aleatorios

- Dada una semilla se determina toda la secuencia
- Pasan los tests de aleatoriedad
- Reproducibilidad

D. H. Lehmer (1951)

- Propuso un algoritmo que se convirtió de hecho en la forma estándar de generar números pseudo-aleatorios.

- La forma del generador de Lehmer es

$$x_{n+1} = a\, x_n \pmod{m}$$

Donde mod significa módulo y caracteriza a objetos matemáticos llamados ‘congruencias’

¿Qué son las congruencias?

- Recorramos todos los naturales x (enteros positivos) y los vamos dividiendo por el mismo entero positivo m al que llamaremos módulo.
- A cada entero le haremos corresponder el Resto (r) de su división por m
- Notemos que todo entero x se puede expresar de modo único mediante un entero positivo m en la forma:

$x = q.m + r \qquad 0 \leq r < m$

- Se toma q.m como el máximo múltiplo de m que no supere x y se observa el resto r
- Entonces: $x \longrightarrow r$

- Si a dos enteros x_i y x_j les corresponde igual resto r al dividir por m, estos números se llaman 'congruentes' y se expresa

$$x_i \equiv x_j \pmod{m}$$

Ejemplo: $4 = 1.3 + 1$

$7 = 2.3 + 1$ (es decir 7 dividido 3 da 2 y resta 1)

Es decir que tanto 7 como 4 son congruentes con 1, módulo 3

Ej: $7 \equiv 1 \pmod{3}$

- Notar que si $a \equiv b \pmod m$, entonces (a-b) es divisible por m,

lo cual se expresa $(a-b) \equiv 0 \pmod m$

- Si volvemos al ejemplo anterior, vimos que 4 y 7 son ambos congruentes con 1 módulo 3. A esta misma 'clase' pertenecen todos los números naturales que si les restan 1 son múltiplos de 3.

$$n = q.3 + 1$$

$$q=1, 2, 3, \ldots\ldots$$

- Las conguencias tienen muchas propiedades interesantes como que se pueden sumar y multiplicar término a término, o que un sumando en cualquier término se puede pasar al otro cambiándole de signo.

Ej: $7 \equiv 1 \pmod 3$ y $4 \equiv 1 \pmod 3$

entonces $7+4 \equiv 1+1 \pmod 3$ es decir $11 \equiv 2 \pmod 3$

ya que 11 div3 da 3 (3x3=9) y restan 2.

Asimismo $7x4 = 28 \equiv 1 \pmod 3$ ya que 9x3=27, que +1 da 28.

Volvamos al generador de Lehmer

$$x_{n+1} = a\, x_n \pmod{m}$$

- Inyectando una ‘semilla’ x_0, dados a y m se van generando sucesivos x_1, x_2, …

- Puede teóricamente obtenerse un período de longitud (m-1) si m es un número primo, y el parámetro a cumple ciertas propiedades proporcionadas por la *Teoría de Números*.

Probemos $x_{n+1} = a\, x_n \pmod{13}$ para distintos a

- Caso 1 – Elegir a=6
- Empezar x0=1
- Entonces la secuencia empieza $1 \rightarrow 6 \rightarrow 10 \rightarrow \ldots..$ Continuarla
- Obtener el período y observar si aparenta irregular, aleatorio.

- Caso 2 – Elegir a=7
- Empezar con x0=1
- Evaluar.

- Caso 3 – Probar a=5
- Iniciar la secuencia con semillas x0={1,2,4}
- Evaluar

- Caso 1 – $x_{n+1} = 6\ x_n \pmod{13}$
  - 1, 6, 10, 8, 9, 2, 12, 7, 3, 5, 4, 11, 1, …

    período completo, desordenado, de 1….m-1

- Caso 2 – $x_{n+1} = 7\ x_n \pmod{13}$
  - 1, 7, 10, 5, 9, 11, 12, 6, 3, 8, 4, 2, 1, ….

    también período completo pero luce algo menos aleatorio

- Caso 3 – $x_{n+1} = 5\ x_n \pmod{13}$
  - Produce una de las secuencias:

    1, 5, 12, 8, 1, …

    2, 10, 11, 3, 2 …

    4, 7, 9, 6, 4, …

  entonces, si m=13, tomar a=5 da una producción muy pobre.

Propuesta de Park & Miller (1988)

para un generador de Lehmer mínimo estándar:

módulo primo $m = 2^{31} - 1$

Multiplicador $a = 7^5$

$$x_{n+1} = 16.807\ x_n \pmod{2.147:483.647}$$

Es un generador de período completo; puede elegirse cualquier semilla x0 entre 1 y (m-1).

Primos de Mersenne

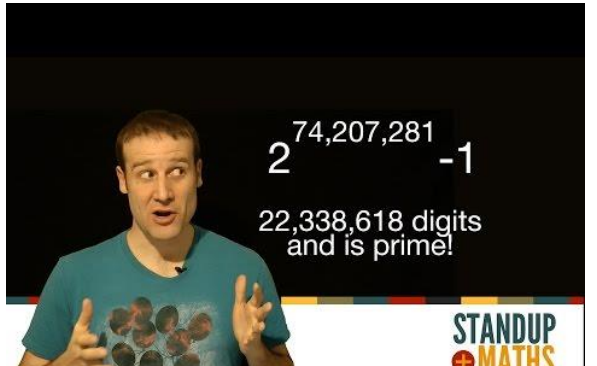
- Números primos de la forma $2^P - 1$, con p primo:

$M_2 = 2^2 - 1 = 3$
$M_3 = 2^3 - 1 = 7$
$M_5 = 2^5 - 1 = 31$
$M_7 = 2^7 - 1 = 127$ todos primos,

… pero $2^{11} - 1 = 2047 = 23 \times 89$ es compuesto.

Hasta 2015 se conocían 48 números primos de Mersenne. El 7 de enero de 2016 se encontró el primo de Mersenne número 49:

$2^{74:207.281} - 1$ que tiene más de 22 millones de dígitos



- El generador de Lehmer estándar fue utiliado por Matlab hasta hace no mucho.

- Actualmente el generador estándar es el llamado “Mersenne twister” cuyo período, que es muy largo, tiene la longitud del primo de Mersenne: $(2^{19.937} – 1)$.