Universidad del Valle

Generadores de números pseudoaleatorios

750098M Simulación computacional

Contenido

1. Introducción
2. Pruebas de bondad
3. Secuencia en otras distribuciones
4. Secuencia en otras distribuciones

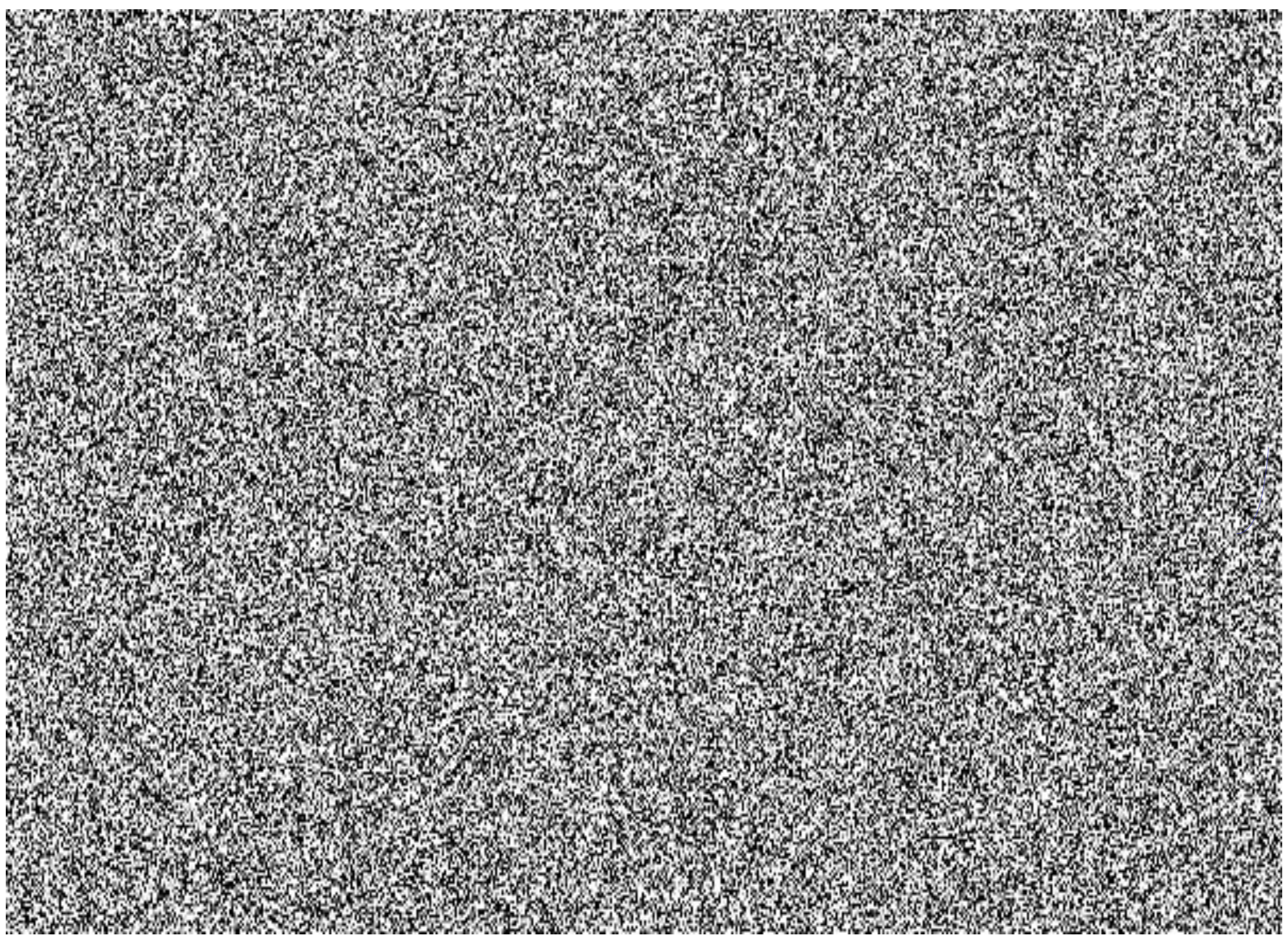
¿Qué es un número aleatorio?

Es un número generado por un proceso sistemático, cuya salida es impredecible y que no puede ser reproducido

Una secuencia es aleatoria si la cantidad de información que contiene, de acuerdo a la teoría información de Shannon, es también finita.

http://www.randomnumbers.info/content/Random.htm

Aparición en la naturaleza



1. Ruido blanco
2. Movimiento de esporas de helecho
3. **Lanzar dados**

¿Existe el azar?

¿Lanzar dados es aleatorio?

https://www.youtube.com/watch?v=tClZGWlRLoE

¿Caos?

Fractales

https://vimeo.com/219046468

https://www.windy.com/

Caos

Desorden

Orden

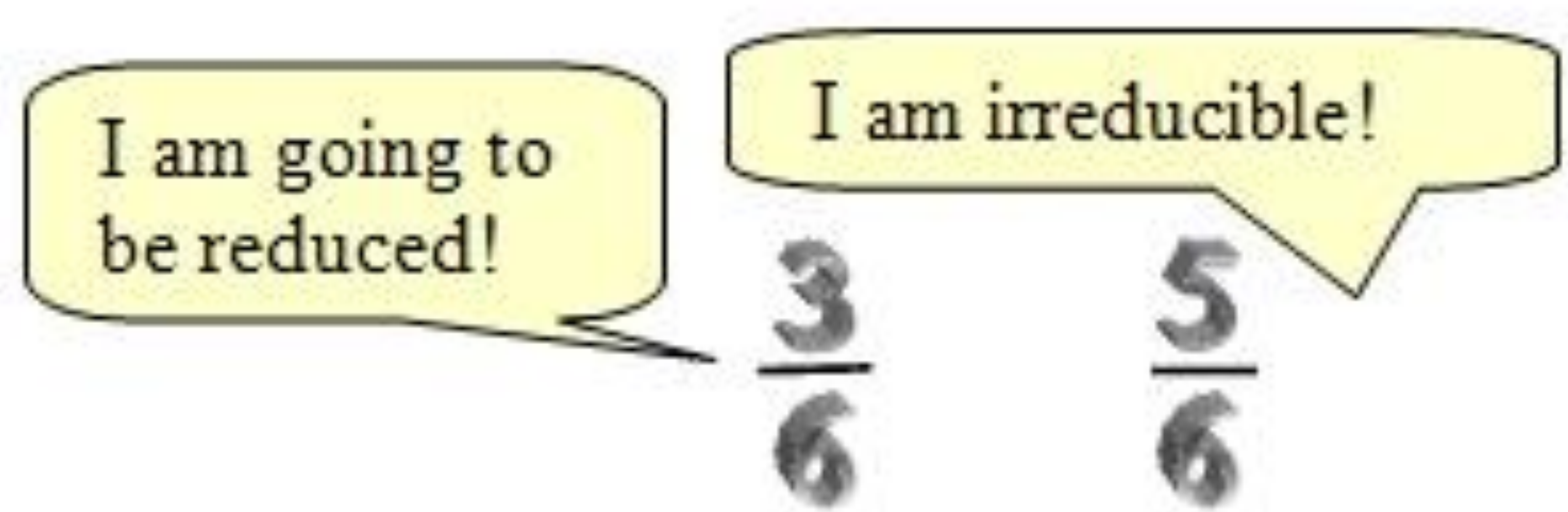
$f(x) = x^2$

Aleatorios

Una definición

Algo es aleatorio si es algorítmicamente incompresible o irreducible.

Exploring RANDOMNESS G J Chaitin, IBM Research Published by Springer-Verlag London, 2001, ISBN 1-85233-417-7



Aplicaciones

- Simulación
- Muestreo
- Análisis numérico
- Programación computacional
- Toma de decisiones

Diferencias

Números aleatorios	**Números pseudoaleatorios**
Son datos continuos	Puede resultar en datos discretos (si son generados en el computador, siempre son discretos)
Siguen distribución uniforme U(0,1) media 1/2 varianza 1/12	Si se divide el intervalo [0,1] en subintervalos iguales pueden resultar intervalos donde caen significativamente más o menos datos que el número esperado media por encima o por debajo de 1/2 varianza por encima o por debajo de 1/12
Los datos son independientes: una observación no depende de las observaciones anteriores; no hay ninguna clase de patrón	Se pueden presentar regularidades como: periodicidad autocorrelación patrones de crecimiento-decrecimiento patrones de valores encima o por de bajo de la media .....y muchos más

¿Qué es un número pseudoaleatorio?

- Es un número generado por una distribución uniforme.
- Un verdadero número aleatorio necesita una fuente impredecible y no reproducible.
- Una estrategia es usar algoritmos matemáticos para generar cadenas de números aleatorios.
- Estos algoritmos reproducen números de una forma determinística, dependiendo de la semilla

Características

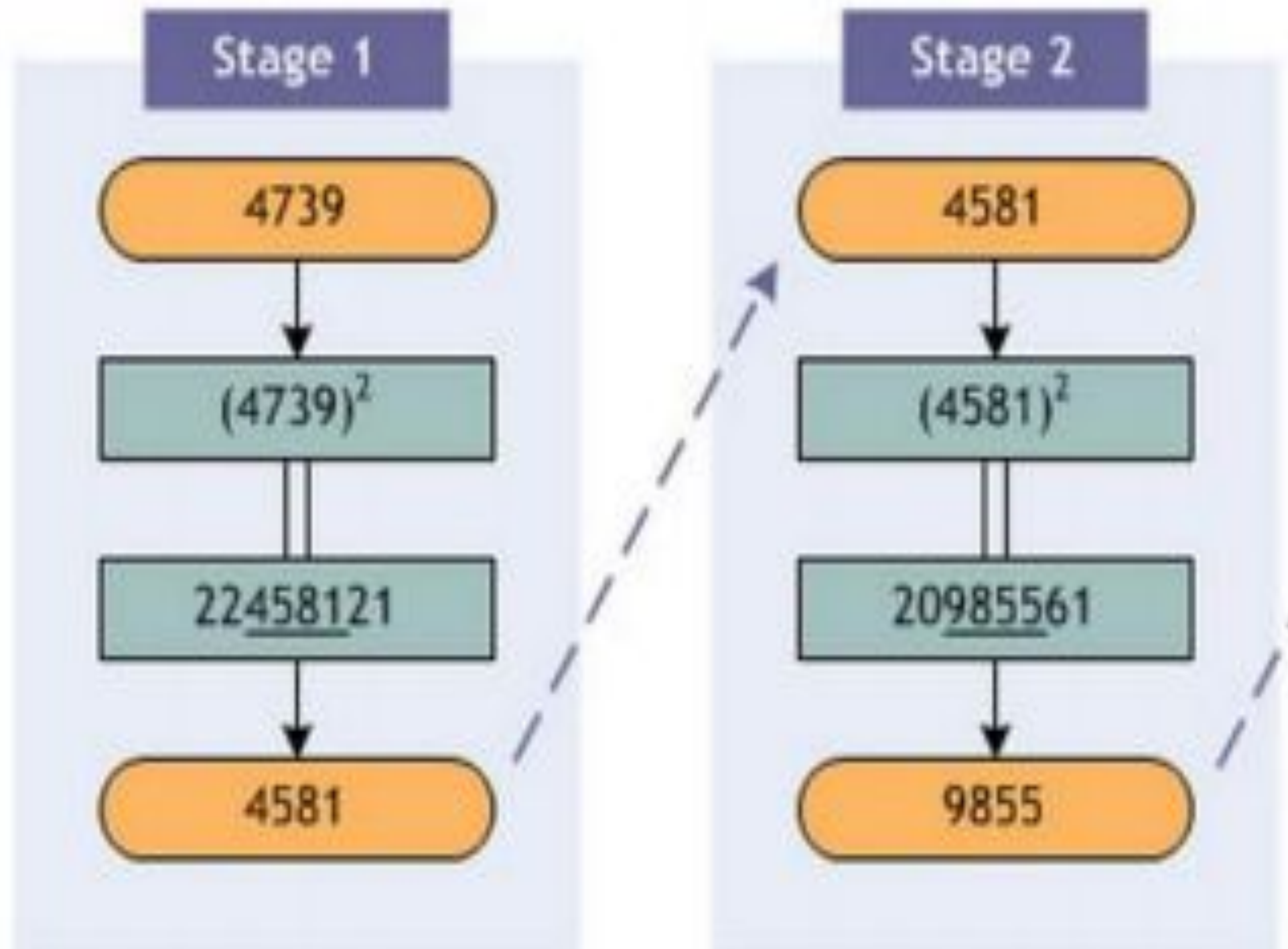
- Uniformemente distribuido
- Dependencia estadística
- Reproducible
- No se repite ningún número en una longitud dada
- Usan una semilla

Consideraciones

- Uniformidad: en cualquier punto del tiempo, la ocurrencia de cualquier número es igualmente probable
- Escalabilidad: Si una secuencia es aleatoria, cualquier subsecuencia debe ser aleatoria
- Consistencia: El comportamiento del generador debe ser bueno con varias semillas

Método Von Neumann

1. 4739
2. $(4739)^2 = 22458121$
3. $22458121 \rightarrow 4581$
4. $(4581)^2 = 20985561$
5. $20985561 \rightarrow 9855$
6. $(9855)^2 = 97121025$



Es conocido como método de los números cuadrados medios

Preocupaciones

- Velocidad de algoritmo
- Fácil implementación
- Técnicas de paralelización
- Implementación portable

Generadores de números pseudoaleatorios

- Produce números enteros X_i uniformemente en $[0,X_{MAX})$
- Se normalizan (0-1) mediante: $u_i = x_i / x_{MAX}$
- Su periodo es hasta que se repite un número (por qué?)
- Un periodo completo es igual a x_{MAX}(por qué?)

Método Congruencia Lineal

Dada una semilla dada X_0 y unos enteros a,c y m:

parametros

Xo = Semilla

1. $X_{n+1} = (a*x_n + c)$ MOD m

2. Repite $X_n = X_{n+1}$ las veces que sean necesarias

Método Congruencia Lineal

$X_{n+1} = (aX_n + c) \mod m$

Por ejemplo para $X_0=7$, $a=1$, $c=7$ y $m=10$:

$X_0 = 7$

1. $X_1 = (1*7 + 7) \mod 10 = 4$
2. $X_2 = (1*4 + 7) \mod 10 = 1$
3. $X_3 = (1*1 + 7) \mod 10 = 8$
4. $X_4 = (1*8 + 7) \mod 10 = 5$

$X_5 = (5 + 7) \mod 10 = 2$

$X_6 = (2 + 7) \mod 10 = 9$

$X_7 = (9 + 7) \mod 10 = 6$

$X_8 = (6 + 7) \mod 10 = 3$

$X_9 = (3 + 7) \mod 10 = 0$

Ahora para $x_0=4$, $a=1$, $c=3$ y $m=5$ haga los 6 primeros pasos

$X_0 = 4$

$X_1 = (4+3) \bmod 5 = 2$

$X_2 = (2+3) \bmod 5 = 0$

$X_3 = (0+3) \bmod 5 = 3$

$X_4 = (3+3) \bmod 5 = 1$

$X_5 = (1+3) \bmod 5 = 4$

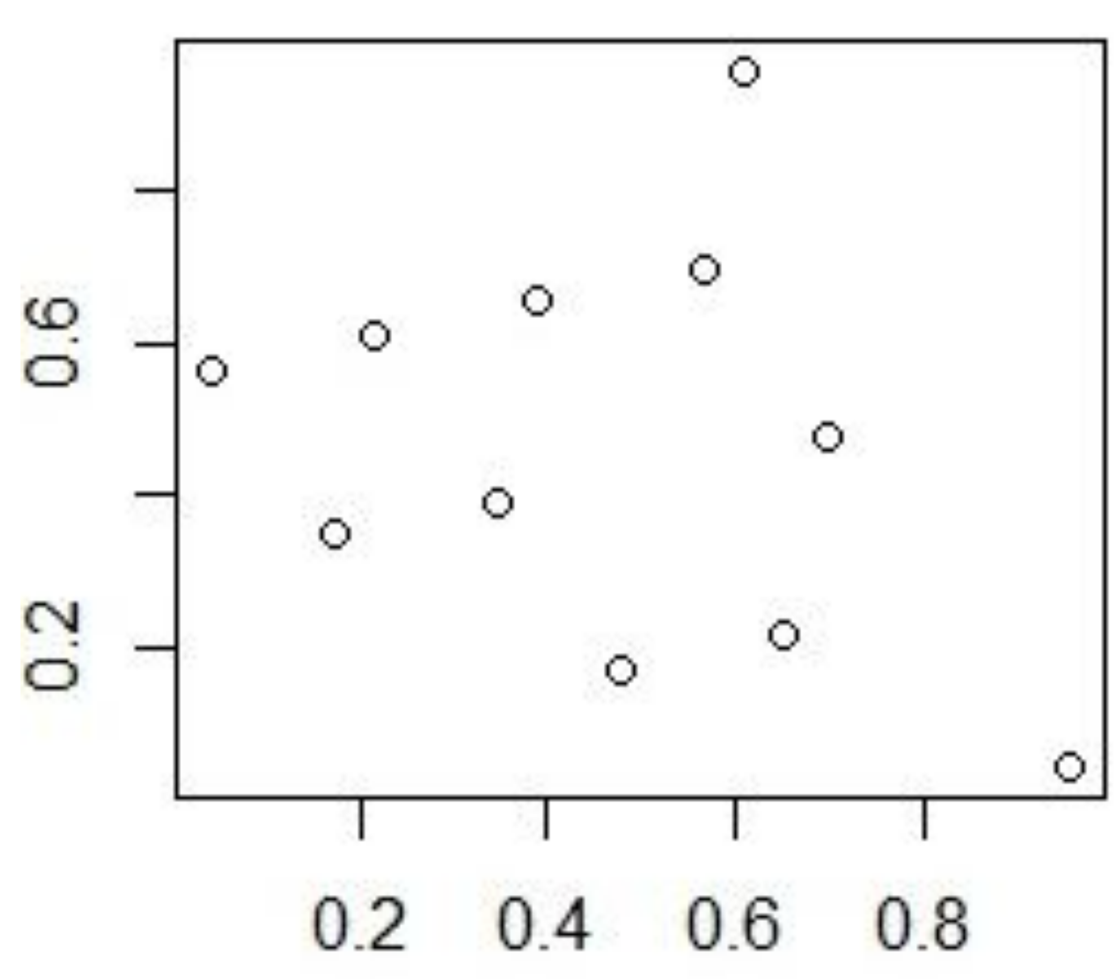
$X_6 = 2$

Método general de congruencia

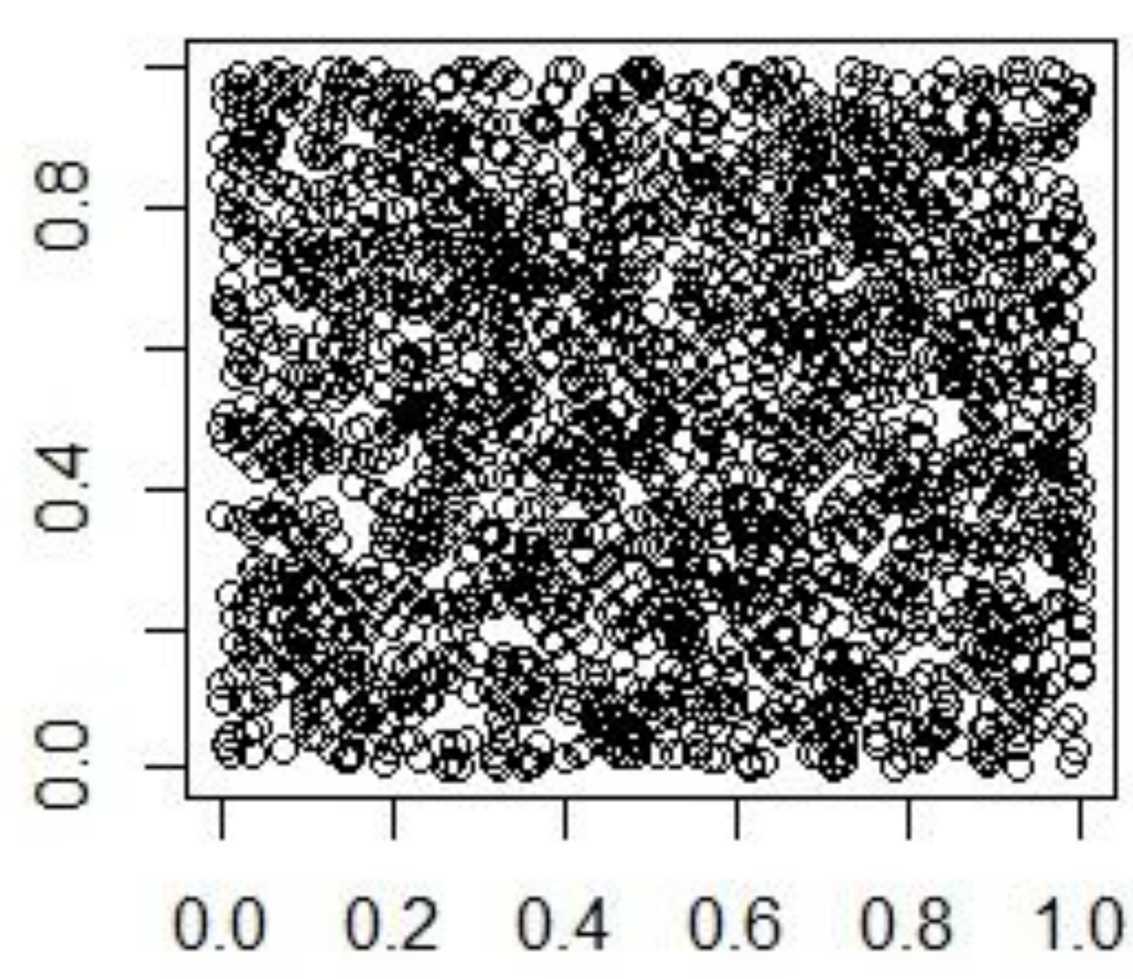
La expresión general es: $x_{i+1} = f(x_i, x_{i-1}) \pmod{m}$

Donde f() es una función de los números previamente generados

MALO



BUENO



Seleccionar el m

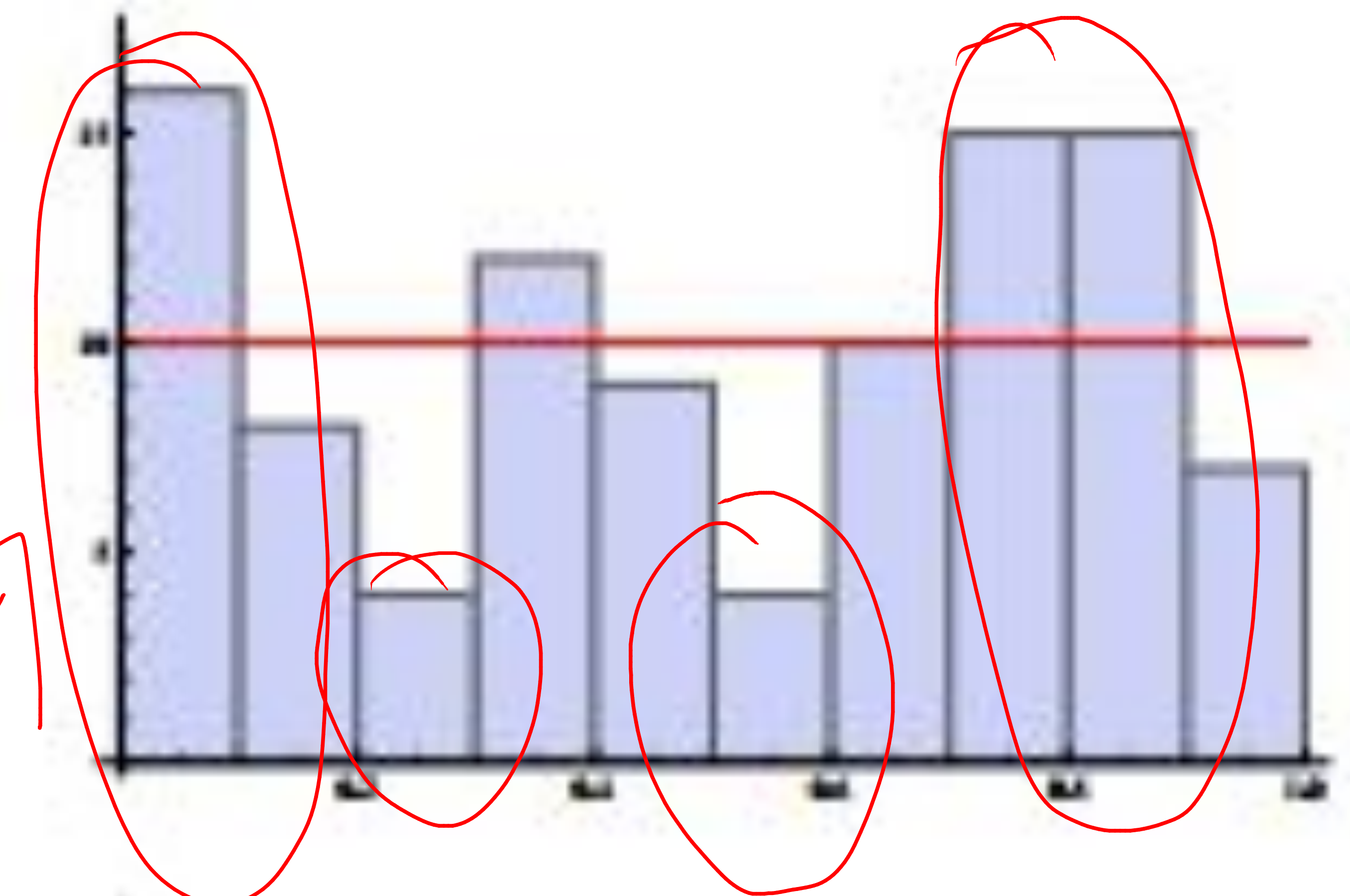
- La secuencia de números es finita
- La secuencia es máximo m -> m debe ser grande
- Se recomienda que m debe ser un número primo
- Se recomienda que m sea una potencia de 2

Otras selecciones

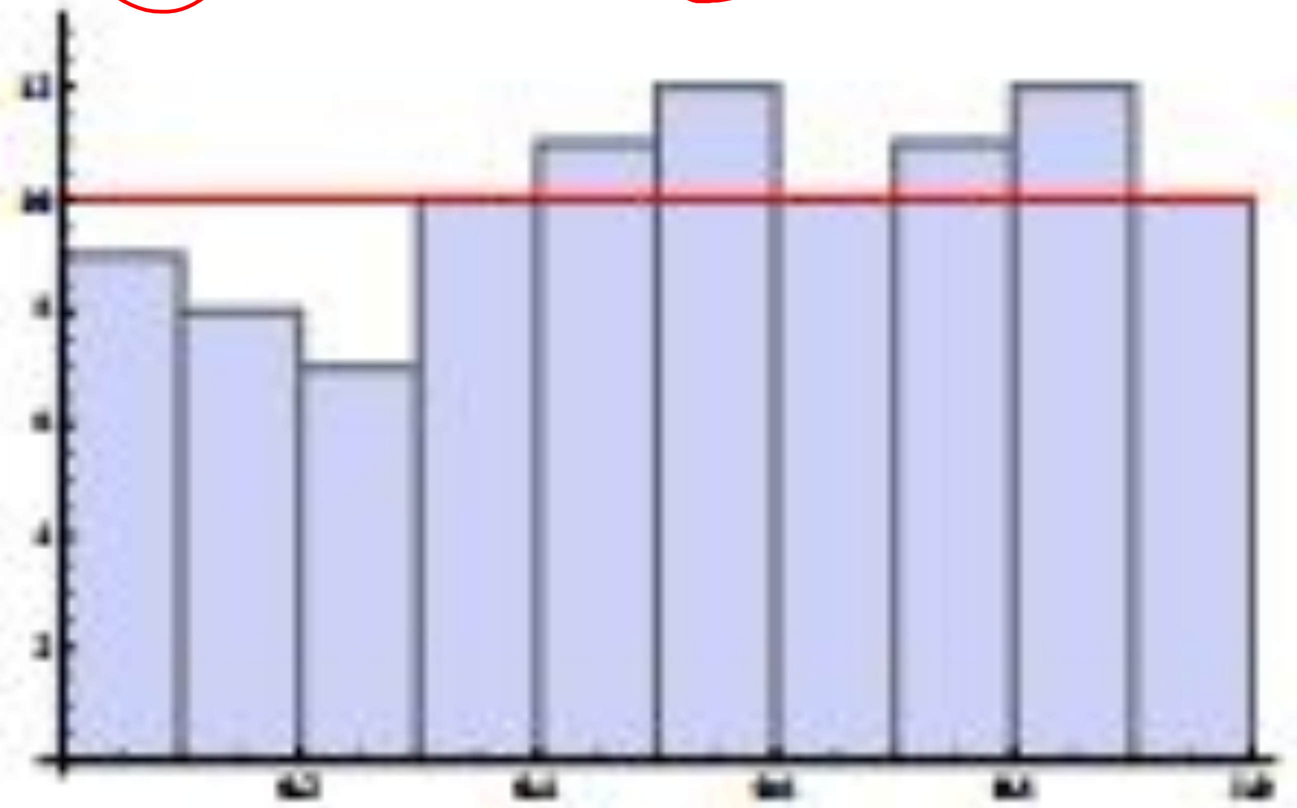
- $0.01m < a < 0.99m$
- Que pasa cuando c=0? (más rápido, periodo corto)
- ¿Cómo escoger el x_0? reloj, ultimo valor, etc
- ¿Que pasa si selecciono un mismo valor de x_0?

Ejemplos, Cuál es mejor?

$X_0 = 5$
$a = 255$
$c = 100$
$m = 1032$



$X_0 = 5$
$a = 255$
$c = 100$
$m = 1031$



100 Datos

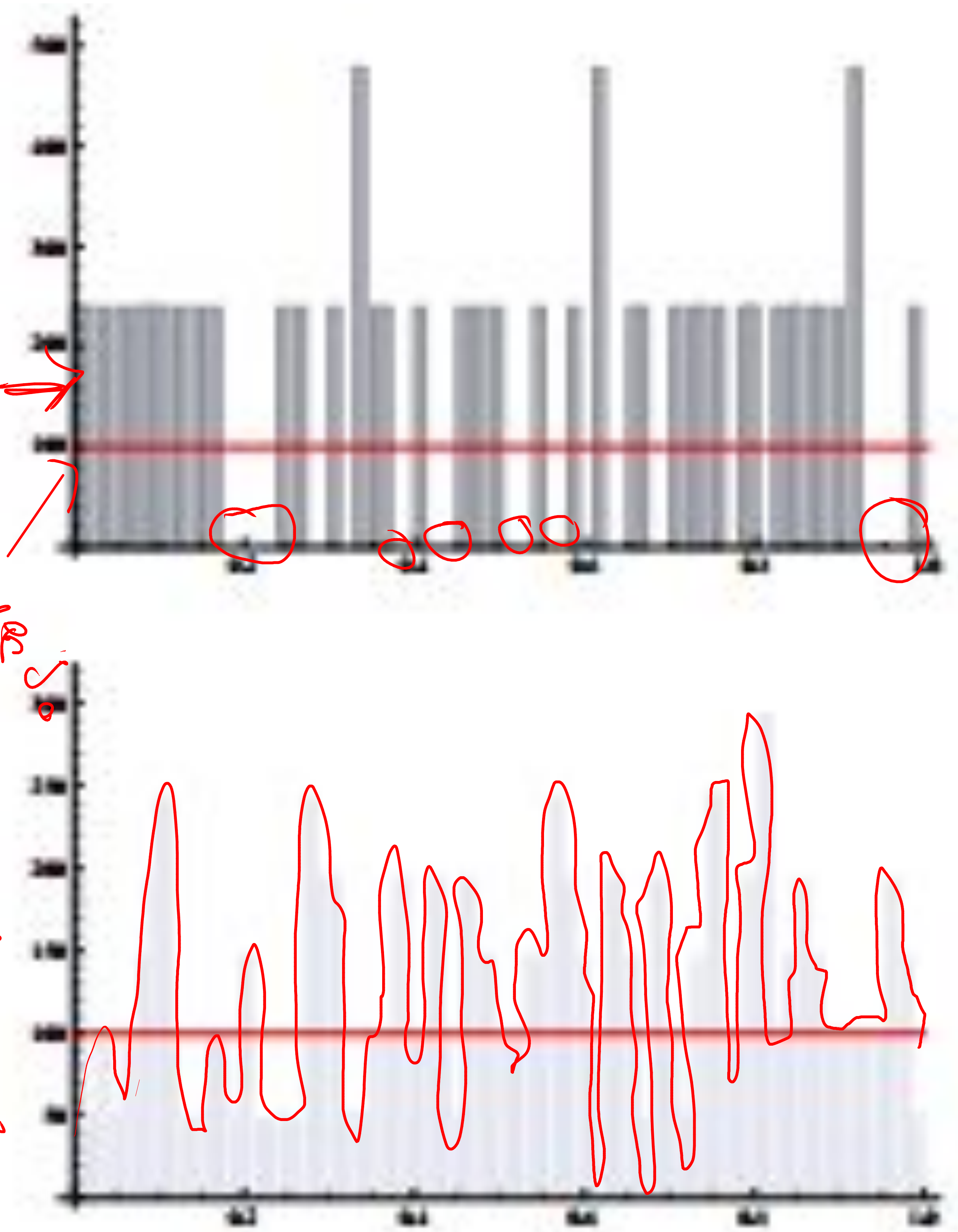
Ejemplos, Cuál es mejor?

$X_0 = 5$
$a = 255$
$c = 100$
$m = 1032$

0 - 1031

$X_0 = 5$
$a = 255$
$c = 100$
$m = 1031$

0 - 1030

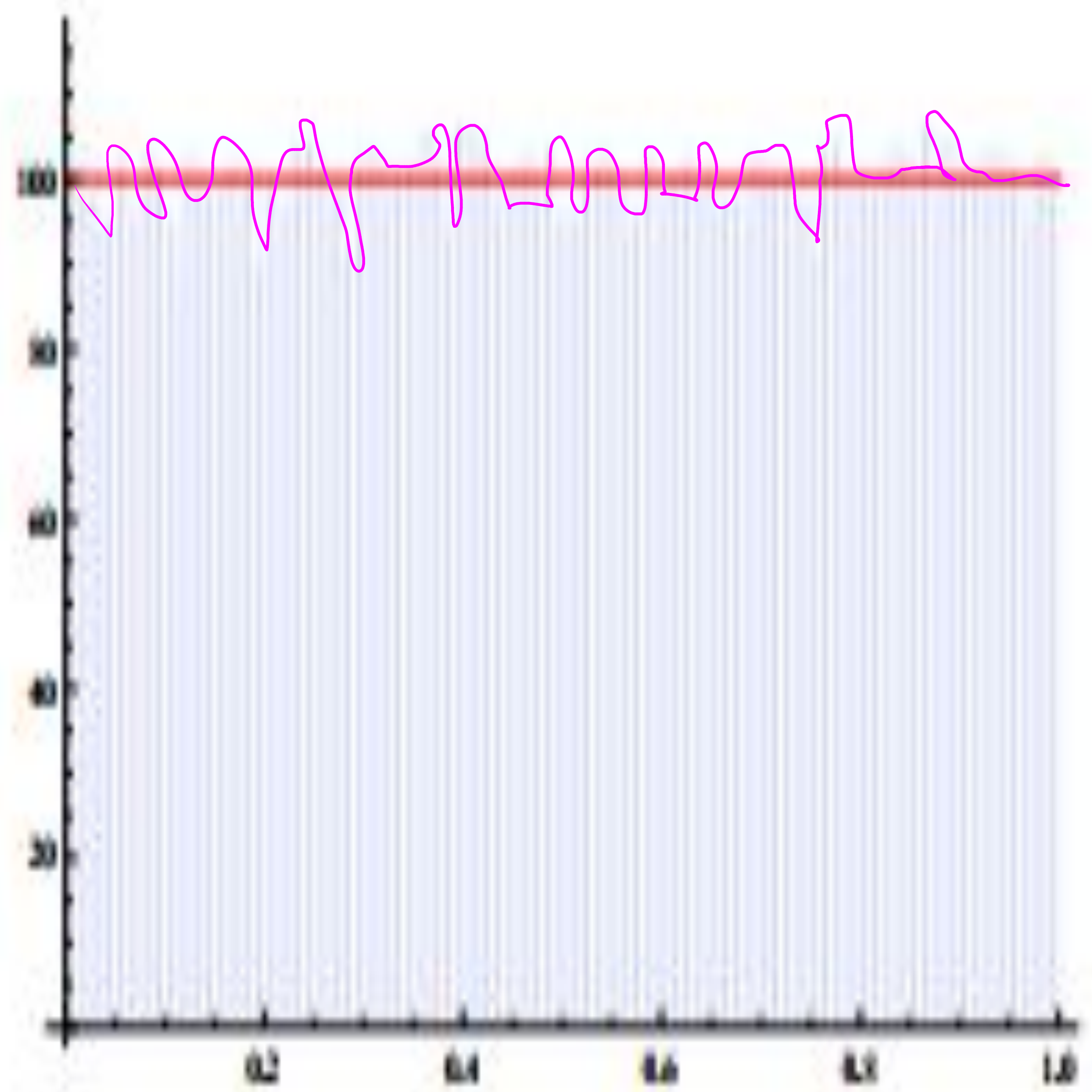


Fix

10000 Datos

Un mejor generador

$$a = 106$$
$$c = 1283 \qquad X_0 = 5$$
$$m = 6075$$



Generador estándar mínimo (GEM)

$\boxed{a X_n \bmod m}$

$$a = 7^5 = 16807 \qquad c = 0 \qquad \boxed{m = 2^{31} - 1}$$

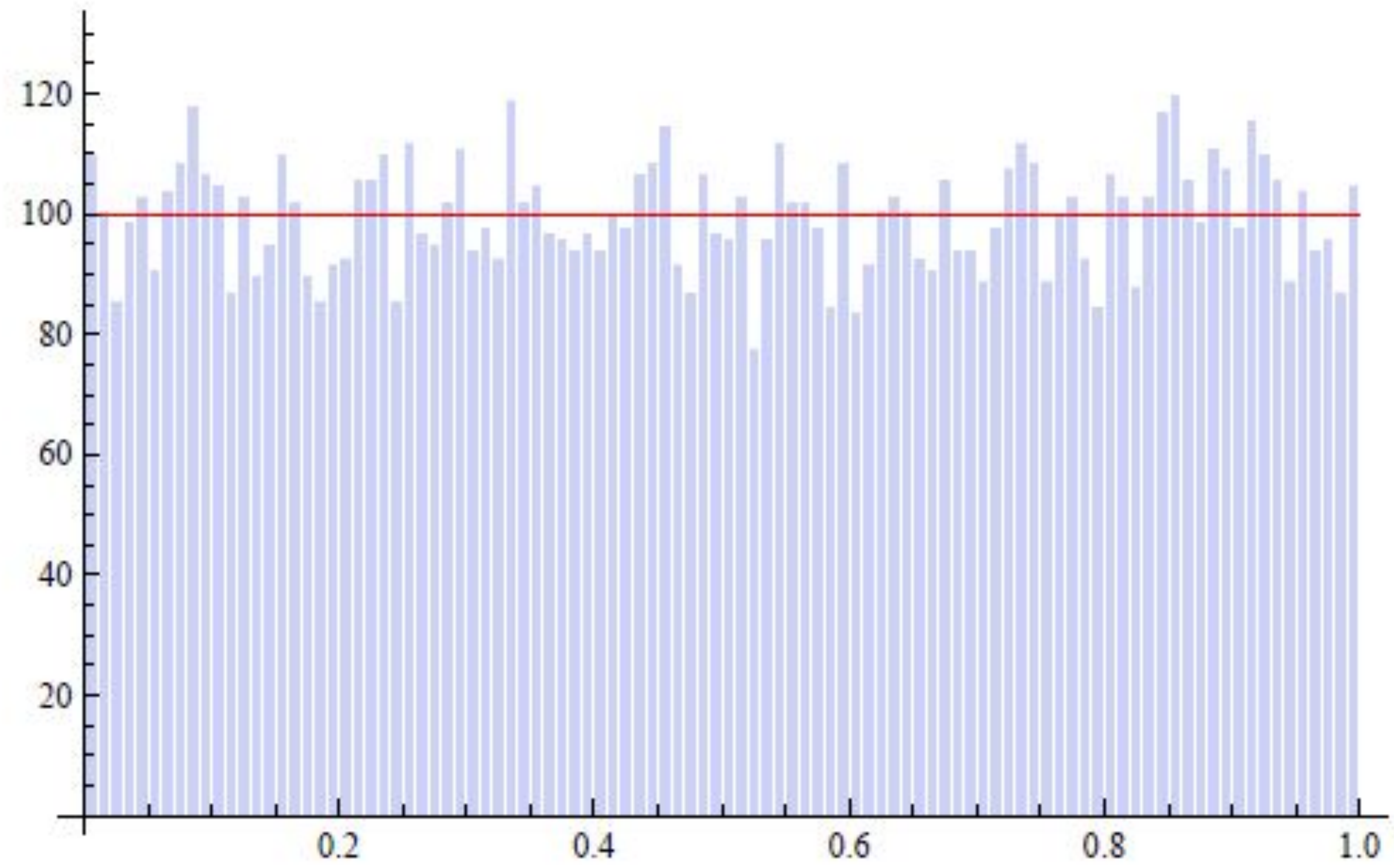
int

- Cumple con las exigencias para ser un buen generador
- Se debe garantizar que no se use 0 como semilla
- Su periodo es m -1

$0 \bmod 5 \rightarrow 0$

Generador estándar mínimo (GEM)

Histograma con 10000 puntos del generador de estándar mínimo



Generador Fibonacci (LFG)

La secuencia Fibonacci es 0,1,1,2,3,5,8,13,21,...

$X_n = x_{n-1} + x_{n-2}$

El generador se puede expresar entonces en:

$$\boxed{X_n = X_{n-j} \text{ Op } X_{n-k} \bmod m}$$

Donde $0 < j < k$ y OP es sumar o multiplicar

LFG Aditivo vs Multiplicativo

Aditivo

$X_n = X_{n-j} + X_{n-k} \mod m$

Multiplicativo

$X_n = X_{n-j} * X_{n-k} \mod m$

Un periodo $m^k - 1$ si m es primo

m es por lo general 2^{32} o 2^{64}

$X_0 = 4 \quad X_1 = 2$

$m = 10$

$6 \bmod 10 = 6$

$6 + 2 \bmod 10 = 8$

$8 + 6 \bmod 10 = 4$

$12 \bmod 10 = 2$

Mersenne Twister

En un generador desarrollado por Matsumoto y Nishimura. En este se utilizan números tipo Mersenne.

Se dice que un número M es de tipo Mersenne si es una unidad inferior a una potencia de 2, es decir $M_n = 2^n - 1$

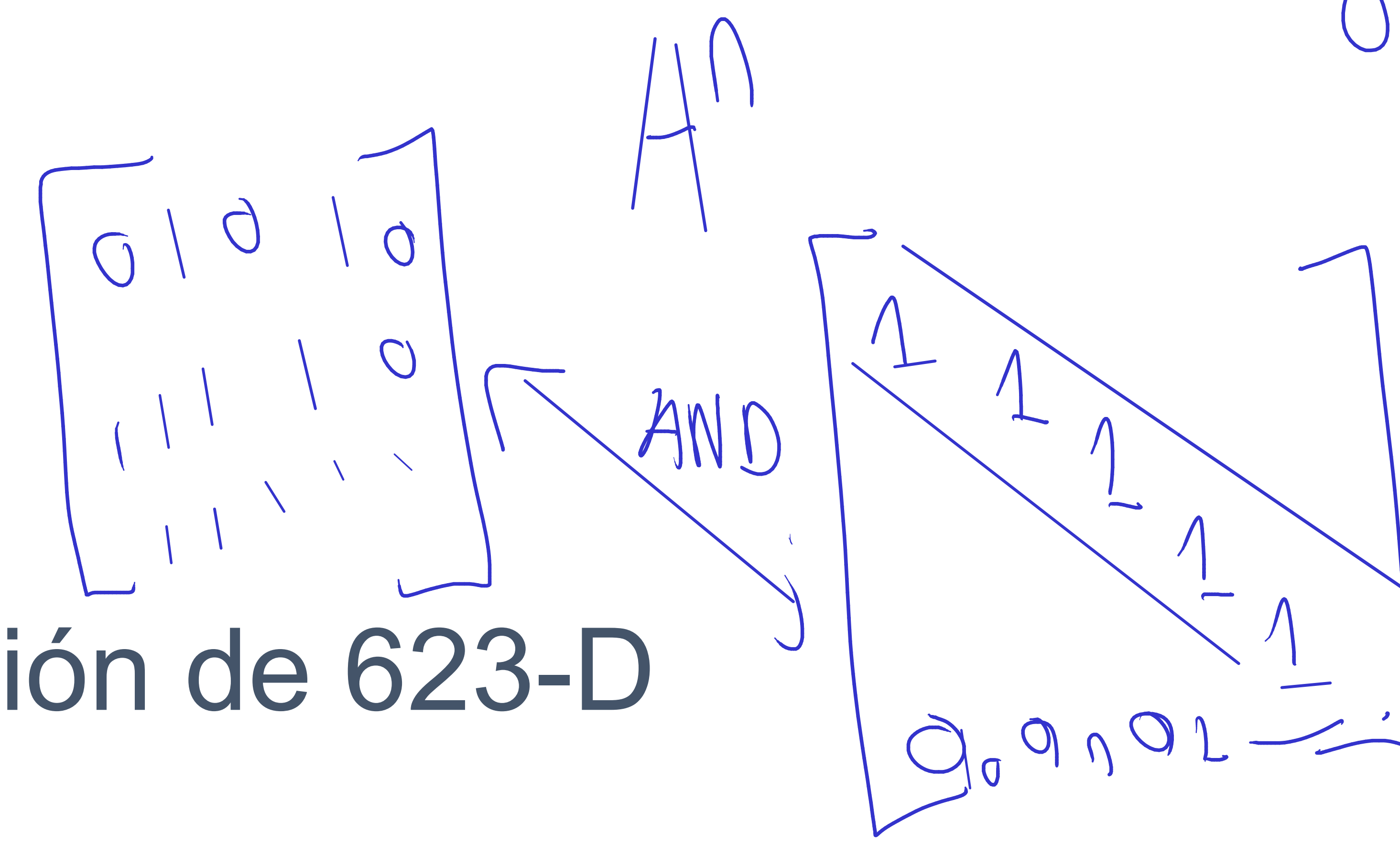
$2^n - 1 \longrightarrow 3 \quad 7 \quad 15 \quad 31$

63

Matsumoto, Makoto & Nishimura, Takuji. (1998). Matsumoto M, Nishimura T.. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comp Simul (TOMACS) 8: 3-30. ACM Trans. Model. Comput. Simul.. 8. 3-30. 10.1145/272991.272995.

Mersenne Twister

623



Este algoritmo:

- Ofrece un periodo largo 2^{19937}
- Esta garantizada una equidistribución de 623-D
- Es de generación rápida
- Uso eficiente de memoria
- Sus operaciones consisten en realizar corrimientos de bytes

Matsumoto, Makoto & Nishimura, Takuji. (1998). Matsumoto M, Nishimura T.. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comp Simul (TOMACS) 8: 3-30. ACM Trans. Model. Comput. Simul.. 8. 3-30. 10.1145/272991.272995.

Resumen

1) Aleatorio vs predicible
2) No se puede computar la aleatoridad
 -- Aleatorio: No es reproducible,
 Dominio Continuo
 Distribución $U[0,1)$
 -- Pseudoaleatorio:
 Reproducible
 Determinista
 Dominino discreto
 Pueden aparecer periodicida-
 des y patrones.

Generadores de números pseudoaleatorios.

Generador del señor Von Neuman. Es bueno, pero costoso computacionalmente (lento)

Generador lineal congruente

$$x_{n+1} = (a x_n + c) \bmod m$$

Depende del m (grande y preferiblemente prim

$0 < a < m$

Depende de la SEMILLA

$m = 2^{31} - 1,$

Generador GEM. a = 7^5 c = 0

$x_0 \neq 0$

Generador de Fibunnacci

$$x_n = (x_{n-i} ** x_{n-j}) mod m$$