# Curso:
# Métodos de Monte Carlo
# Unidad 4, Sesión 10: Números aleatorios (parte 3)

Departamento de Investigación Operativa
Instituto de Computación, Facultad de Ingeniería
Universidad de la República, Montevideo, Uruguay

dictado semestre 1 - 2025

# Contenido:

1. Principales familias de generadores de números seudo-aleatorios (continuación).

2. Medidas de bondad y pruebas estadísticas.

3. Lectura adicional.

4. Ejercicio.

# Generadores basados en combinaciones de generadores lineales

Dado que los generadores congruenciales lineales tienen muy buenas propiedades respecto a su equidistribución en bajas dimensiones, su simplicidad conceptual y su facilidad de implementación, pero también algunas limitaciones respecto a su equidistribución en altas dimensiones, un enfoque posible consiste en plantearse de qué manera y bajo que condiciones es posible modificar la salida de un generador congruencial lineal para mejorar sus propiedades, sin alterar significativamente su simplicidad y eficiencia.

Hay al menos dos métodos que han sido empleados con éxito en este marco: el método de *shuffling*, y el de *merging*. Si bien el análisis teórico de ambos es parcial, la teoría existente y también los resultados empíricos muestran que ambos son métodos a considerar seriamente.

# Shuffling

El método de *shuffling* consiste en reordenar la salida de un generador cualquiera, en base a la secuencia generada por un segundo generador. Formalizando, supongamos que tenemos dos generadores congruenciales lineales de secuencias de números seudo-aleatorios, $\{X_i, i \geq 0\}$, e $\{Y_i, i \geq 0\}$, de módulos $M_1$ y $M_2$ respectivamente, y queremos generar una nueva secuencia $\{Z_i, i \geq 0\}$, con mejores propiedades. Entonces el método consiste en inicializar un vector de largo $K$, $\mathbf{T} = (T_1, \ldots, T_K)$ con los valores $X_0, \ldots, X_{K-1}$, inicializar $i = 0$, y generar cada valor $Z_i$ con el procedimiento siguiente:

Generar $Y_i$.
$j = \lceil KY_i/M_2 \rceil$.
$Z_i = T_j$.
Generar $X_{K+i}$.
$T_j = X_{K+i}$.
$i = i + 1$.

De esta forma, estamos generando una permutación (aleatoria) de la salida original del generador $X$; el generador $Y$ es utilizado simplemente como fuente de "ruido" en el orden en el que se toman los elementos de $X$.

Los resultados teóricos muestran que, siempre que $X$ e $Y$ sean independientes, la nueva secuencia $Z$ tendrá iguales o mejores propiedades que la secuencia $X$ original, aunque no resulta fácil cuantificar el grado de mejoría.

Una ventaja de este método es que es conceptualemente simple y muy rápido de implementar, requiriendo solamente un espacio de memoria relativamente pequeño, y algunas operaciones adicionales (una de las más costosa es la necesidad de generar los $Y_i$).

# Merging

Si tenemos dos generadores $\{X_i, i \geq 0\}$, e $\{Y_i, i \geq 0\}$, de módulos $M_1$ y $M_2$ respectivamente, y períodos $d_1$ y $d_2$, entonces $Z_i = (X_i + Y_i)$ mod $M_1$ es un generador de período $mcm(d_1, d_2)$, y cuya distribución es al menos tan uniforme como la de $X$ y la de $Y$.

Existen diversas variantes en torno a esta idea, en todas hay importantes ganancias en los períodos obtenidos, y las propiedades del generador pueden mejorar considerablemente.

# Generadores basados en recurrencias no lineales

Estos generadores están basados en una fórmula $Z_i = g(Z_{i-1}) \bmod M$, donde $g()$ es una función con dominio y recorrido enteros (y que en el mejor caso será periódica de período $M$).

Los dos casos más usados y de mayor interés son los generadores cuadráticos y los inversivos. Los cuadráticos tienen la forma $Z_i = (AZ_{i-1}^2 + BZ_{i-1} + C) \bmod M$, y los inversivos la forma $Z_i = (AZ_{i-1}^{-1} + C) \bmod M$, donde $x^{-1}$ es el inverso multiplicativo de $x$ módulo $M$ (es decir, tal que $xx^{-1} \bmod M = 1$).

Los generadores no lineales evitan algunas limitaciones de los lineales (cuyos valores se encuentran siempre en una grilla en el espacio), a costa de mayor costo computacional (y mayor dificultad para un análisis teórico de sus características, que hace que no siempre sea fácil lograr que su comportamiento global sea mejor que el de las mejores variantes de los generadores lineales).

# Medidas de bondad y pruebas estadísticas

Para estudiar el comportamiento de los generadores de números seudo-aleatorios, hay dos enfoques que se complementan. Por un lado, el estudio teórico permite calcular ciertas medidas que reflejan que tan "bueno" es un generador (en el sentido de comportarse de manera similar a una secuencia aleatoria). Por otro, es posible realizar tests empíricos, bajo la forma de prueba de hipótesis, que permiten si el generador no es adecuado, refutar la hipótesis nula "el generador se comporta como una secuencia de valores uniformes e idénticamente distribuidos".

En relación a los llamados tests teóricos, los dos más importantes se basan en los conceptos de discrepancia (una medida de $k$-equidistribución, relacionada a la diferencia entre la distribución esperada en un hipercubo de dimensión $k$ y y la distribución empírica), y de test espectral (relacionado con la inversa de la mínima distancia entre hiperplanos que contienen soluciones). También se estudian otra medidas tales como el mínimo número de hiperplanos necesarios para cubrir todos los puntos de la secuencia, y la mínima distancia entre puntos en el hiper-espacio de

dimensión $k$.

Respecto a los tests empíricos, es posible pensar y aplicar una cantidad muy grande de pruebas posibles. Las dos más clásicas son quizá el uso de tests de Kolmogorov Smirnov (junto con las estadísticas de $\chi^2$) para verificar si un generador cumple con la uniformidad en su salida, así como la ausencia de correlación entre pares de valores sucesivos.

Una discusión detallada de los mismos escapa a los objetivos del curso.

Existen diversos paquetes y blbliotecas para realizar pruebas a generadores de números aleatorios, damos una lista a continuación (no es obligatorio entrar a cada uno):

- Set de tests Diehard, elaborado por George Marsaglia, `http://www.staff.science.uu.nl/~sleij101/Opgaven/LabClass/site/asm_diehard.php`. (último acceso: 2025-02-22)
  `https://webhome.phy.duke.edu/~rgb/General/rand_rate/rand_rate.abs`. (último acceso: 2025-02-22)

- TestU01 - Empirical Testing of Random Number Generators, elaborado por Pierre L'Ecuyer, `http://simul.iro.umontreal.ca/testu01/tu01.html`. (último acceso:2025-02-22)

- The ENT test program, `http://www.fourmilab.ch/random/`. (último acceso:2025-02-22)

- Random Number Generation and Testing - NIST (con énfasis en aplicaciones criptográficas): `https://csrc.nist.gov/projects/random-bit-generation`(último acceso:2025-02-22)

La biblioteca científica de GNU, GSL, provee implementaciones de varios buenos generadores (así como de otros clásicos, aunque con defectos conocidos):
`https://www.gnu.org/software/gsl/doc/html/rng.html`.(último acceso:2025-02-22)

En la Wikipedia hay descripciones de varios de los métodos actualmente en uso: `http://en.wikipedia.org/wiki/List_of_pseudorandom_number_generators`.(último acceso:2025-02-22)

# Lectura adicional obligatoria

Artículo científico sobre propiedades deseables y resultados de tests sobre generadores de números seudo-aleatorios en bibliotecas de lenguajes usualmente usados: P. L'Ecuyer, "Software for Uniform Random Number Generation: Distinguishing the Good and the Bad", Proceedings of the 2001 Winter Simulation Conference, IEEE Press, Dec. 2001, 95-105. `http://www.informs-sim.org/wsc01papers/012.PDF`(último acceso:2025-02-22)

# Preguntas para auto-estudio

- ¿Qué métodos basados en combinaciones de generadores lineales conoce?
- ¿Cómo funcionan los métodos basados en recurrencias no lineales?
- ¿Qué medidas permiten evaluar la bondad de un generador de números seudo-aleatorios?
- Ver este video de divulgación, `https://www.youtube.com/watch?v=RzEjqJHW-NU` y ponerlo en relación con los conceptos discutidos en esta unidad y en las precedentes.

# Entrega 6

Ejercicio 10.1: (individual)

Resumir en un texto (de entre media y una carilla) los principales contenidos del paper de P. L'Ecuyer, "Software for Uniform Random Number Generation: Distinguishing the Good and the Bad", Proceedings of the 2001 Winter Simulation Conference, IEEE Press, Dec. 2001, 95-105. ¿Cuál es el objetivo del trabajo? ¿Qué generadores de números seudo-aleatorios discute? ¿Cuáles son los hallazgos de esta investigación, y cuáles las conclusiones y recomendaciones presentadas por el autor?

Fecha entrega: Ver Avance del curso (cronograna)