

L'Ecuyer y velocidad WARP

Gomez Vidal, Maximiliano
dgomezvi@alu.itba.edu.ar

Abramowicz, Pablo
pabramow@alu.itba.edu.ar

Sessa, Carlos
csessa@alu.itba.edu.ar

Villa Fernández, Santiago
svillafe@alu.itba.edu.ar

RESUMEN

En este artículo se analiza el generador de números pseudo-aleatorios sugerido por L'Ecuyer y se lo pone a prueba utilizando el test de Kolmogorov-Smirnov y el test χ^2 . Se realiza también una simulación del tiempo de vuelo de una nave espacial mediante el método de Montecarlo.

Palabras clave

L'Ecuyer, generador lineal congruencial, simulación de Montecarlo, test χ^2 , test Kolmogorov-Smirnov, propulsión WARP

1. INTRODUCCIÓN

El concepto de aleatoriedad está presente en diversos campos de la ciencia, tales como la criptografía y la estadística.

En determinadas ocasiones es deseable generar secuencias de números aleatorios. Resulta lógico pensar en generar dichas secuencias utilizando una computadora. Sin embargo, los números al azar surgen únicamente de procesos naturales que no pueden ser recreados en un dispositivo determinista como lo es una computadora. La solución a este problema consiste en utilizar algoritmos que permitan generar secuencias suficientemente largas de números que presenten una distribución estadística regular. Es decir, que no debe ser posible distinguir mediante pruebas estadísticas que la secuencia no ha sido construida realmente al azar.

Uno de los generadores más conocidos de números pseudo-aleatorios es el generador lineal congruencial (LCG).

En la sección 2 se analiza el generador de L'Ecuyer, que utiliza dos LCG para generar las secuencias. En la sección 3 se somete al generador de L'Ecuyer a tests estadísticos para analizar la distribución de los números generados. En la sección 4 se obtiene una distribución triangular a partir de la salida del generador. En la sección 5 se realiza una simulación del tiempo de vuelo de la nave USS Enterprise utilizando el método de Montecarlo. Finalmente, se exponen las conclusiones en la sección 6.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

2. GENERADOR DE L'ECUYER

El generador sugerido por L'Ecuyer combina dos LCG de acuerdo al siguiente algoritmo:

PASO 1 Seleccionar una semilla $X_{1,0}$ en el rango $[1, 2147483562]$ para el LCG1 y $X_{2,0}$ en el rango $[1, 2147483398]$ para el LCG2.

PASO 2 Evaluar cada generador individual

$$X_{1,n+1} = 40014 X_{1,n} \bmod 2147483563 \quad (1)$$

$$X_{2,n+1} = 40692 X_{2,n} \bmod 2147483399 \quad (2)$$

PASO 3 Computar

$$X_{n+1} = (X_{1,n+1} - X_{2,n+1}) \bmod 2147483562 \quad (3)$$

PASO 4 Computar

$$U_{n+1} = \begin{cases} \frac{X_{n+1}}{2147483563}, & X_{n+1} > 0 \\ \frac{2147483562 - X_{n+1}}{2147483563}, & X_{n+1} = 0 \end{cases} \quad (4)$$

PASO 5 Hacer $n = n + 1$ e ir al *PASO 2*.

3. PROBANDO EL GENERADOR DE L'ECUYER

Resulta de interés estudiar si la secuencia generada se encuentra distribuida de manera uniforme. A continuación se somete a prueba el generador de L'Ecuyer utilizando el test χ^2 y el test de Kolmogorov-Smirnov.

3.1 Test χ^2

Se desea establecer si es posible aceptar la hipótesis H_0 : la secuencia generada utilizando el algoritmo de la sección 2 corresponde a muestras de una variable uniforme $\mathcal{U}[0, 1)$, con un nivel de confianza del 95%. Para esto se generan 10,000 números utilizando (1, 2) como semillas. Luego se agrupan en 100 intervalos de clase.

La media esperada en cada intervalo de clase es $E_i = 100$, pues se analiza una distribución uniforme. El estadístico dado por la ecuación:

$$\chi_0^2 = \sum_{i=1}^{100} \frac{(O_i - E_i)^2}{E_i} \quad (5)$$

resulta aproximadamente 98.84. Se observa que la distribución χ^2 subyacente posee 99 grados de libertad, puesto

que se han considerado 100 categorías para agrupar los números generados.

Para el nivel de significación deseado, $\alpha = 0.05$, se obtiene el valor crítico $\chi_{99,0.05}^2 = 124.3421$. Como $\chi_{99,0.05}^2 > \chi_0^2$ se acepta H_0 y se determina que los números están distribuidos de manera uniforme. En la figura 1 se observa la distribución de los 10,000 números para cada clase. En las figuras 2 y 3 se observa que no es posible detectar patrones en la distribución de la secuencia generada, puesto que la misma presenta una densidad uniforme.

3.2 Test Kolmogorov-Smirnov

Este test se utiliza para contrastar uniformidad. Compara la distribución empírica obtenida a partir de una muestra con la distribución teórica correspondiente a la variable aleatoria en estudio.

Se mantienen los mismos 10,000 números generados en la sección 3.1 y la misma hipótesis H_0 . Se agrupan los números nuevamente en 100 intervalos. Se realiza el test correspondiente comparando la secuencia arrojada por el generador de L'Ecuyer con otra muestra que presenta una distribución uniforme.

El resultado arrojado es una aceptación de H_0 con aproximadamente un 96.45% de confianza, por lo que no se rechaza la hipótesis nula a un nivel de significación de 0.05. Es decir, la secuencia generada presenta una distribución uniforme.

4. DENSIDAD TRIANGULAR

En los experimentos de simulación es necesario generar secuencias de números pseudo-aleatorios distribuidos de acuerdo a una función $F(x)$ arbitraria.

A continuación se utiliza la técnica de la transformada inversa aplicada a la función de densidad triangular dada por:

$$f_X(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)}, & a \leq x \leq b \\ \frac{2(c-x)}{(c-b)(c-a)}, & b < x \leq c \\ 0, & \text{en otro lado} \end{cases} \quad (6)$$

Esta técnica establece que dado un número u que es una realización de $\mathcal{U} \sim \mathcal{U}[0,1]$, entonces $F^{-1}(u)$ es una realización de una variable aleatoria X con función de distribución $F(x)$.

Integrando la ecuación 6 se obtiene la función de distribución $F_X(x)$:

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x^2 - 2ax + a^2}{(b-a)(c-a)}, & a \leq x \leq b \\ \frac{b-a}{c-a} + \frac{2cx - x^2 - b(2c-b)}{(c-b)(c-a)}, & b < x \leq c \\ 1, & x > c \end{cases} \quad (7)$$

Por último, la VA X resulta:

$$X = \begin{cases} a + \sqrt{U(c-a)(b-a)}, & 0 \leq U \leq \frac{b-a}{c-a} \\ c - \sqrt{(c-b)^2 - U(c-b)(c-a) + (b-a)(c-b)}, & \frac{b-a}{c-a} < U \leq 1 \end{cases} \quad (8)$$

En la figura 4 se observa la distribución para los valores $a = 0$, $b = 1$ y $c = 3$ al generar las variables $X_1, X_2, \dots, X_{10000}$.

5. SISTEMA DE PROPULSIÓN WARP

El sistema de propulsión WARP de la nave espacial USS Enterprise consiste en un núcleo WARP en el interior de un reactor donde se llevan a cabo reacciones de aniquilación materia-antimateria, moderadas por cristales de Dilitio. Dichos cristales se procesan en una cámara controlada llamada Matriz de Dilitio, el cual tiene un subsistema de redundancia. Este proceso se realiza para alimentar a los propulsores en las nescellas laterales de la nave. El tiempo de operación de cada propulsor es una variable aleatoria exponencialmente distribuida X_1, X_2 con tiempo medio $T_m = 240$ horas. El tiempo de operación entre fallos del núcleo WARP es una variable uniforme $X_3 \sim \mathcal{U}[72, 12]$, el tiempo de operación de la cámara de Dilitio es una variable uniforme $X_4 \sim \mathcal{U}[20, 30]$ y, por último, el tiempo entre fallos de la cámara de redundancia es una variable uniforme $X_5 \sim \mathcal{U}[5, 7]$, todas medidas en horas.

Las funciones de distribución se describen como:

Siendo x_1 y x_2 realizaciones de X_1 y X_2 :

$$f_{x_1, x_2}(t) = \begin{cases} \frac{1}{T_m} e^{-\frac{t}{T_m}} & t > 0 \\ 0 & t < 0 \end{cases} \quad (9)$$

Siendo x_3 una realización de X_3 :

$$f_{x_3}(x) = \begin{cases} \frac{1}{12} & x \in (72, 84) \\ 0 & x \notin (72, 84) \end{cases} \quad (10)$$

Siendo x_4 una realización de X_4 :

$$f_{x_4}(x) = \begin{cases} \frac{1}{30} & x \in (20, 50) \\ 0 & x \notin (20, 50) \end{cases} \quad (11)$$

Siendo x_5 una realización de X_5 :

$$f_{x_5}(x) = \begin{cases} \frac{1}{7} & x \in (5, 12) \\ 0 & x \notin (5, 12) \end{cases} \quad (12)$$

El tiempo de operación del sistema integrado es una variable aleatoria T definida como:

$$T = \min\{\max\{X_4, X_5\}, X_3, \max\{X_2, X_1\}\} \quad (13)$$

Cuyo tiempo medio entre fallos del sistema se describe de la forma:

$$\mathcal{E}\{T\} = \int_{\mathcal{D}} \min\{\max\{x_4, x_5\}, x_3, \max\{x_2, x_1\}\} f_{x_1}(x_1) f_{x_2}(x_2) f_{x_3}(x_3) f_{x_4}(x_4) f_{x_5}(x_5) d^5x \quad (14)$$

donde \mathcal{D} es el dominio $[0, \infty] \times [0, \infty] \times [72, 84] \times [20, 50] \times [5, 12]$.

Se transforma el dominio de integración $\mathcal{D} \rightarrow [0, 1]^5$ de manera de poder realizar la simulación de Montecarlo para este sistema. Las transformaciones se definen como:

$$x_1 = -T_m \ln(1 - u_1) \quad (15)$$

$$x_2 = -T_m \ln(1 - u_2) \quad (16)$$

$$x_3 = 12u_3 + 72 \quad (17)$$

$$x_4 = 30u_4 + 20 \quad (18)$$

$$x_5 = 5u_5 + 7 \quad (19)$$

siendo u_i realizaciones de variables aleatorias $\mathcal{U}[0, 1]$.

Se obtiene el determinante jacobiano de la transformación:

$$J = \begin{vmatrix} \frac{T_m}{1-u_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{T_m}{1-u_2} & 0 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{vmatrix} = \frac{180T_m^2}{(1-u_1)(1-u_2)} \quad (20)$$

Por último se reemplaza en (14) con (15), (16), (17), (18), (19), se multiplica por (20), y se obtiene la integral transformada:

$$\mathcal{E}\{T\} = \int_{[0,1]^6} \min\{\max\{30u_4 + 20, 5u_5 + 7\}, 12u_3 + 72, \max\{-T_m \ln(1 - u_2), -T_m \ln(1 - u_1)\}\} d^5 u \quad (21)$$

Para la realización de la simulación del sistema mediante el método de Montecarlo se utiliza el generador de números pseudo-aleatorios de L'Ecuyer obtenido en la sección 3, calculando 5 secuencias correspondientes a $u_i \in [0, 1]$. Si consideramos la variable aleatoria $Z = T_1 + T_2 + \dots + T_N$, para N suficientemente grande Z se aproxima a una VA con distribución normal utilizando el teorema del límite central.

Aplicando la regla de las 3S se determina que

$$P(a - 3S \leq Z \leq a + 3S) \approx 0.997 \quad (22)$$

donde se considera que a es la media muestral de Z . Operando algebraicamente se obtiene la expresión

$$P(|\sum \frac{T_i}{N-m}| \leq \frac{3S}{\sqrt{N}}) \approx 0.997$$

donde m representa el valor esperado de T_i . Esto significa que podemos estimar el valor medio de la VA T calculando el promedio de las distintas muestras obtenidas, sabiendo con una probabilidad muy cercana a 1 que el error de dicho promedio está acotado por $\frac{3S}{\sqrt{N}}$.

Se realizan 20 simulaciones de la integral (21), resultando el estimador de tiempo medio de vuelo $\langle T \rangle = 34.833650$ horas y el desvío muestral $S = 0.968095$ horas. Se determina que para un nivel de significación del 5%, el tiempo medio de vuelo del sistema de propulsión WARP pertenece al rango 34.83 ± 0.453 horas.

u_1	4	19
u_2	27	54
u_3	73	98
u_4	101	135
u_5	141	175

Table 1: Semillas para los generadores de L'Ecuyer de u_i

6. CONCLUSIONES

A partir de las simulaciones y los tests que se realizan al generador de números pseudo-aleatorios de L'Ecuyer se desprende que es un buen generador, ya que no solo no se observan hiperplanos en los gráficos sino que también fue sometido a los tests χ^2 y Kolmogorov-Smirnov con resultado exitoso. Este generador permite realizar diversas simulaciones con la seguridad de que los datos de entrada presentan una distribución uniforme. En lo que respecta al USS Enterprise la principal limitación del sistema es la Matriz de Dilitio ya que el tiempo de operación es el menor comparado a los demás componentes la nave. La cámara redundante no es de suma ayuda ya que el tiempo entre fallos es mucho menor que el de operación del sistema principal. Además del hecho que el componente funciona siempre a la par del principal y no cuando este falla, logra que el sistema integro base su tiempo de vuelo únicamente en la Matriz de Dilitio.

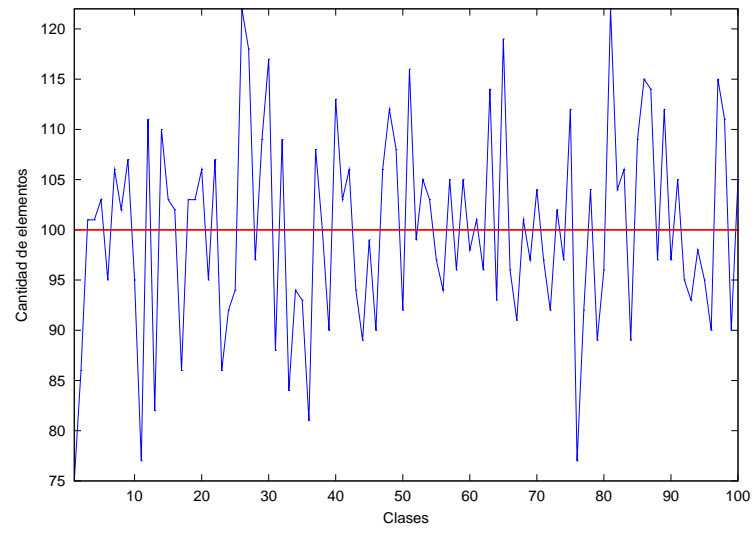


Figura 1: Cantidad de resultados por categoría para el test χ^2

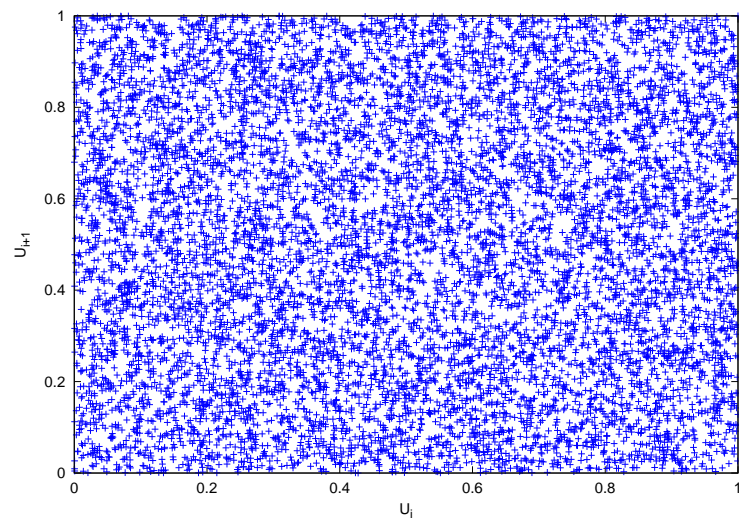


Figura 2: Duplas (U_i, U_{i+1}) de números pseudo-aleatorios generados

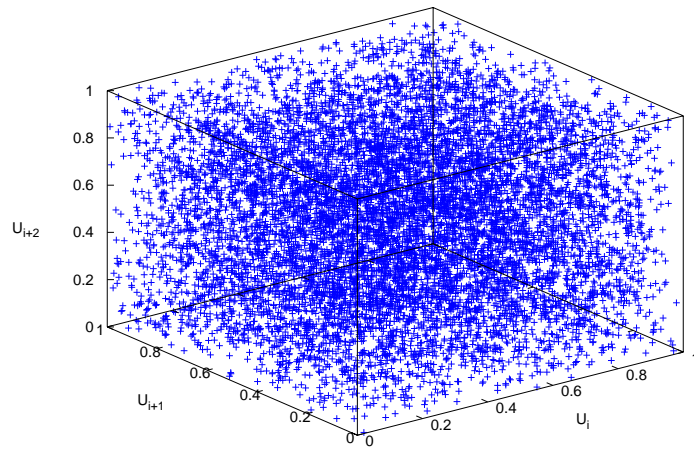


Figura 3: Ternas (U_i, U_{i+1}, U_{i+2}) de números pseudo-aleatorios generados

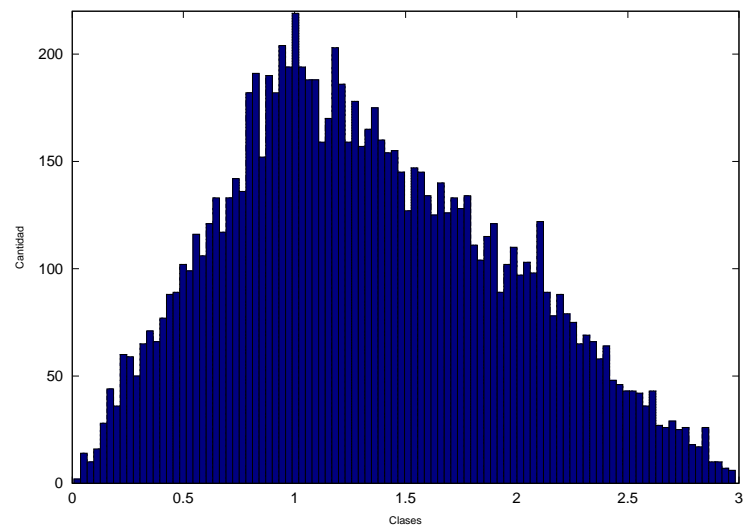


Figura 4: $X_1, X_2, \dots, X_{10000}$ con distribución triangular

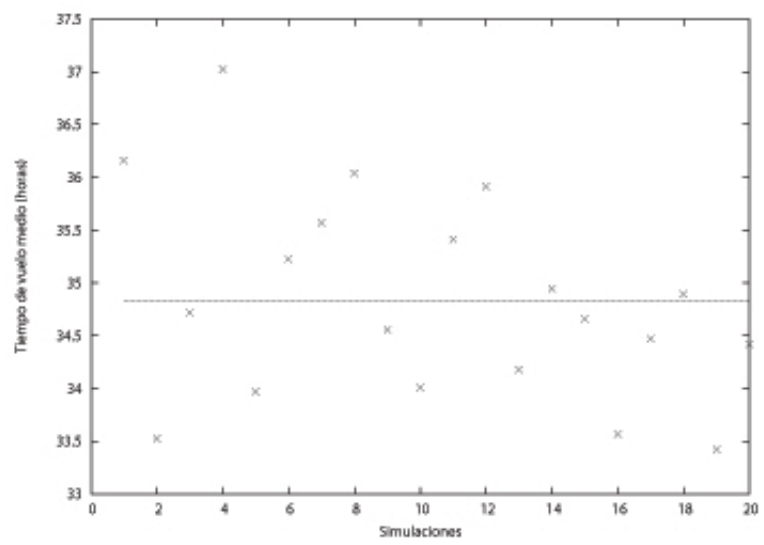


Figura 5: Tiempos medios de vuelo vs. Simulaciones