Estimación del tiempo de vuelo del USS Enterprise

Carlos Sánchez Romero
casanche@alu.itba.edu.ar

Germán Wachnitz
gwachnit@alu.itba.edu.ar

S.J. Migliorisi
smiglior@alu.itba.edu.ar

RESUMEN

En el presente artículo se estudia la generación de números aleatorios con distribución uniforme y triangular, y la utilización de los mismos para estimar el tiempo medio de vuelo de la nave espacial `USS Enterprise`.

Palabras clave

generador de L'Ecuyer, distribución triangular, simulación de Montecarlo.

1. INTRODUCCIÓN

2. GENERADOR DE L'ECUYER

2.1 Descripción del generador

Uno de los generadores de números aleatorios uniformes propuesto por Pierre L'Ecuyer[BANKS,1998] en 1998 consiste en la combinación de dos generadores lineales congruenciales (GLCs, o LCGs en inglés). Para el uso de este generador se necesita una semilla $X_{1,0}$ en el intervalo $[1, 2147483562]$ y otra semilla $X_{2,0}$ en el intervalo $[1, 2147483398]$. Para generar una serie de números pseudo aleatorios utilizando las semillas $X_{1,0}$ y $X_{2,0}$, el algoritmo sigue una serie de pasos:

PASO 1 Se evalua cada generador linear

$$X_{1,n+1} = 40014 X_{1,n} \ mod \ 2147483563$$
$$X_{2,n+1} = 40692 X_{2,n} \ mod \ 2147483399$$

PASO 2 Se computa:

$$X_{n+1} = (X_{1,n+1} - X_{2,n+1}) \ mod \ 2147483562$$

PASO 3 Se computa:

$$U_{n+1} = \begin{cases} \frac{Xn+1}{2147483563}, & Xn+1 > 0 \\ \frac{2147483562}{2147483563}, & Xn+1 = 0 \end{cases}$$

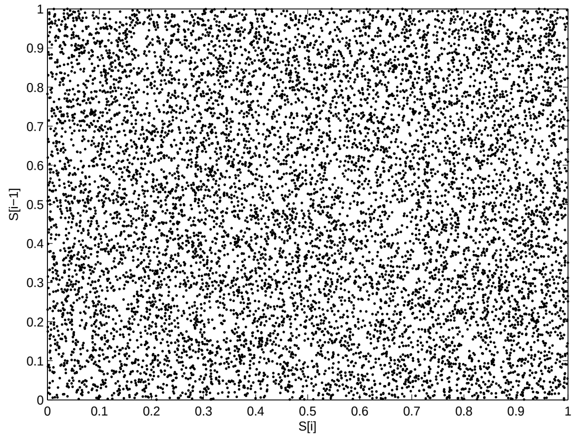

Figura 1: Nube de puntos $S[i]$ vs $S[i-1]$ vs $S[i-2]$.

PASO 4 Se modifica $n = n + 1$ y se continua con el paso 1

Este generador tiene un período $\approx$ 2x10^{18}.

2.2 Bondad del generador

En esta sección se analiza la distribución y la independencia de los números que se obtienen mediante el generador.

2.2.1 *Test gráfico en 2 dimensiones*

Se analiza la independencia de los números que se obtienen mediante el generador de l'Ecuyer. Para realizar el test se obtiene un vector de 2000 números, utilizando el generador que se describe en la sección 2.1, fijando como semillas $X_{1,0} = 1123$ y $X_{2,0} = 3542$. Así, la serie de números que se genera es: $S = (n_1, n_2, ..., n_{2000})$. Con estos números se obtiene un conjunto de puntos P_i, con coordenadas:

$$P_i = (S[i-1], S[i]) \tag{1}$$

donde $i = 2, 3, ..., 2000$. Luego se grafican estos puntos, obteniéndose la nube de puntos que se muestra en la figura **??**. En la misma, no se observa ningún tipo de patrón evidente, por lo que no se puede afirmar que el número $S[i]$ dependa del número $S[i-1]$.

2.2.2 *Test gráfico en 3 dimensiones*

Siguiendo la idea de la sección 2.2.1 y utilizando las mismas condiciones iniciales, se obtiene un conjunto de puntos con coordenadas:

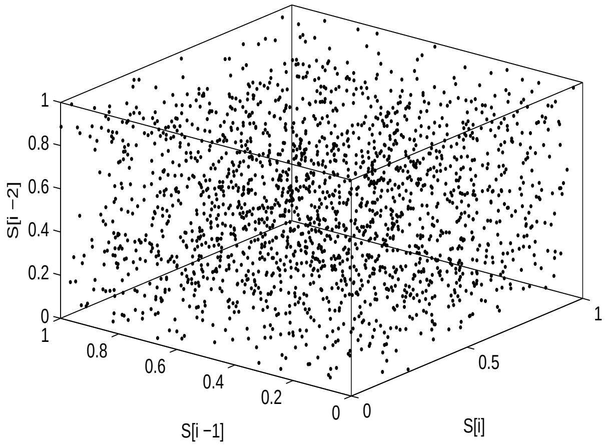

Figura 2: Nube de puntos $S[i]$ vs $S[i-1]$ vs $S[i-2]$.

$$P_i = (S[i-2], S[i-1], S[i]) \quad (2)$$

donde $i = 3, 4, ..., 2000$. Luego se grafican estos puntos, obteniéndose la nube de puntos que se muestra en la figura **??**. En la misma, no se observa ningún tipo de patrón evidente, por lo que no se puede afirmar que el número $S[i]$ dependa del número $S[i-1]$ y/o del número $S[i-2]$.

2.2.3 Test χ^2

Las hipótesis del test son:

H_0: $\chi_0^2 < \chi_{n-1,\alpha}$(Los datos que se generan mediante el algoritmo de L'Ecuyer están uniformemente distribuidos)

H_1: $\chi_0^2 \geq \chi_{n-1,\alpha}$(Los datos que se generan mediante el algoritmo de L'Ecuyer no están uniformemente distribuidos)

siendo H_0 la hipótesis nula y H_1 la hipótesis alternativa. Para realizar el test se utiliza el vector de números que se mencionan en la sección 2.2.1. Los números que se generan se agrupan en 10 intervalos de clase: $C_1, C_2, ..., C_{10}$. Luego la cantidad de números que se observa en cada categoría resulta:

$O = \{$

$O(C_1) = 197, O(C_2) = 201, O(C_3) = 228, O(C_4) = 198, O(C_5) = 196, O(C_6) = 173, O(C_7) = 216, O(C_8) = 224, O(C_9) = 188, O(C_10) = 179$

$\}$

Como se quiere contrastar los números que se generan con una distribución uniforme, el valor esperado teórico es el mismo para cada intervalo de clase: $E = \frac{2000}{10} = 200$.

Entonces, a partir de O y E, se computa el estadístico: $\chi^2 = 14.80$. El valor crítico, para una significación $\alpha = 0.05$, es $\chi_{9,0.05} = 16.92$. Por lo tanto, se rechaza H_1 en favor de H_0.

2.2.4 Test Kolmogorov-Smirnov

En vista a que el test χ^2 no rechaza la posibilidad de que los datos que se generan tengan una distribución uniforme, se realiza un test de Kolmogorov-Smirnov, ahora en adelante test KS. Las hipotesis del test KS son:

H_0: $D < D_\alpha$(Los datos que se generan mediante el algoritmo de L'Ecuyer están uniformemente distribuidos)

H_1: $D \geq D_\alpha$(Los datos que se generan mediante el algoritmo de L'Ecuyer no están uniformemente distribuidos)

donde D es el resultado del test y D_α es el correspondiente valor crítico.

El test KS se realiza utilizando las mismas condiciones iniciales que en la sección 2.2.1.

Realizando los cómputos pertinentes, resulta: $D = 0.1165$. El valor crítico, para una significación de $\alpha = 0.05$ y para 10 intervalos de clase, es $D_{0.05} = 0.4092$ [NEAVE,1981]. Luego como $D < D_{0.05}$, no se puede rechazar H_0.

3. ALGORITMO PARA OBTENER NÚMEROS ALEATORIOS CON DISTRIBUCIÓN TRIANGULAR

Para obtener números pseudo aleatorios según la función de densidad triangular, se define:

$$f(x) = \begin{cases} \frac{2(x-a)}{(c-a)(b-a)} & a \leq x \leq b \\ \frac{2(c-x)}{(c-a)(c-b)} & b \leq x \leq c \\ 0 & \text{en otro caso} \end{cases} \quad (3)$$

donde $a < b < c$. Luego se obtiene la función de distribución $F(x)$ integrando la ecuación partida **??**:

$$F(x) = \int_{-\infty}^{x} f(x)dx \quad (4)$$

resultando:

$$F(x) = \begin{cases} \frac{(x-a)^2}{(c-a)(b-a)} & a \leq x \leq b \\ 1 - \frac{(c-x)^2}{(c-a)(c-b)} & b < x \leq c \end{cases} \quad (5)$$

Si X es una variable aleatoria con función distribución $F(x)$ y función de densidad triangular $f(x)$, se define $U = F(X)$, donde U es una variables aleatoria con distribución uniforme en el intervalo $[0, 1]$, entonces $X = F^{-1}(U)$ (la demostración se puede ver en [DIAZ,2007]). De esta manera se pueden utilizar números aleatorios uniformemente distribuidos en el intervalo $(0, 1)$ para generar números con una distribución triangular. Calculando $F^{-1}(u)$ se obtiene:

$$F^{-1}(u) = \begin{cases} a + \sqrt{u(c-a)(b-a)} & 0 \leq u \leq \frac{b-a}{c-a} \\ c - \sqrt{(c-a)(c-b)(1-u)} & \frac{b-a}{(c-a)} < u \leq 1 \end{cases} \quad (6)$$

de esta manera con una realización u de la variable aleatoria U podemos generar valores x que tienen una distribución triangular con parámetros a,b y c.

3.1 Estimación del valor esperado de una variable aleatoria con distribución triangular

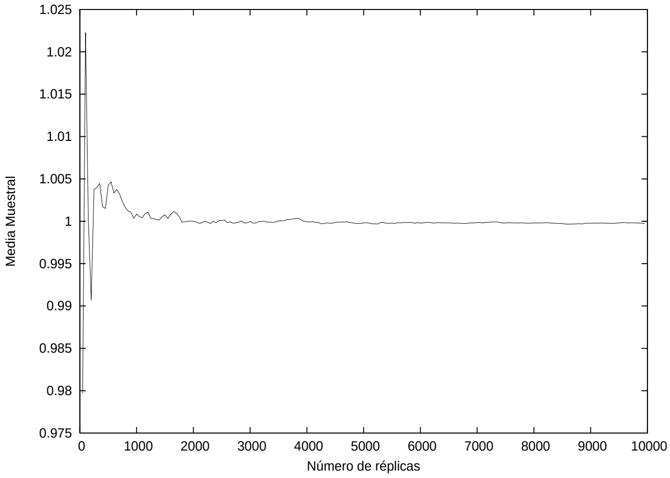

Figura 3: Valores de $\overline{X}$ para muestras de tamaño $n = 50, 100, ..., 10000$.

Según la *Ley débil de los grandes números*, si n es el tamaño de la muestra y $n \to \infty$, la media muestral $\overline{X} \xrightarrow{P} E(X)$.

Para verificar empíricamente esta ley, se define la variable aleatoria $X : Triangular(a, b, c)$, con $a = 0$, $b = 1$ y $c = 2$, y se simulan n valores de X, con $n = 50, 100, 150, ..., 10000$, utilizando la función $F^{-1}(u)$ que se obtiene en la sección 3 y el generador de números pseudo aleatorios con distribución uniforme que se describe en la sección 2.1.

Considerando que estos n valores son una muestra aleatoria de tamaño n, se calcula la media muestral $\overline{X}$.

Teniendo en cuenta que $E(X)$ se computa como:

$$E(X) = \int_a^c xf(x)dx \tag{7}$$

se obtiene que $E(X) = \frac{a+b+c}{3}$. Para el caso en que $a = 0$, $b = 1$ y $c = 2$ resulta $E(X) = 1$. En la figura **??** se puede ver como los valores de la media muestral $\overline{X}$ convergen a $E(X)$ en probabilidad.

4. ESTIMACIÓN DEL TIEMPO DE VUELO DEL ENTERPRISE

4.1 Cálculo analítico

El sistema de propulsión `WARP` de la nave espacial `USS Enterprise` se compone de dos propulsores que se alimentan mediante un núcleo `WARP`. El núcleo se encuentra en el interior de un reactor donde se llevan a cabo reacciones de aniquilación materia-antimateria, que se moderan con el uso de cristales de dilitio. Para modelar este sistema se definen tres componentes: $\texttt{PROPULSOR}_1$, $\texttt{PROPULSOR}_2$ y `NÚCLEO WARP`.

Se definen X_1 y X_2 como variables aleatorias exponencialmente distribuidas que determinan el tiempo de operación del $\texttt{PROPULSOR}_1$ y $\texttt{PROPULSOR}_2$ respectivamente. Ambas variables tienen un tiempo medio de $T_1 = T_2 = 10$ días, o lo que es lo mismo, 240 horas. Además se define una variable aleatoria X_3 que representa el tiempo de operación entre fallos del `NÚCLEO WARP`, el cual es de 72 horas pudiendo variar linealmente hasta 12 horas. Por lo tanto, X_3 es una variable aleatoria con distribución triangular de parámetros $a = 60$ horas, $b = 72$ horas y $c = 84$ horas.

El tiempo de vuelo del `USS Enterprise` es una variable aleatoria T, que se define como:

$$T = \min\{\max\{X_1, X_2\}, X_3\} \tag{8}$$

Conociendo el modelo de las variables aleatorias X_1, X_2 y X_3, todas ellas independientes, resulta que el tiempo medio de vuelo de la nave es:

$$\begin{aligned} E\{T\} &= \iiint_D \min\Big\{\max\{x_1, x_2\}, x_3\Big\}\phi_{X_3}(x_3) \\ &\quad \frac{1}{T_1 T_2} e^{-(\frac{x_1}{T1}+\frac{x_2}{T2})} d^3x \end{aligned} \tag{9}$$

donde D es el dominio de x_1, x_2 y x_3 y $\phi_{X_3}(x_3)$ es la función densidad de probabilidad triangular $f(x)$ definida en la ecuación **??**.

Se transforma el dominio de integración: $D \to [0, 1]^3$ para simplificar los cálculos. Por lo tanto, las transformaciones de x_1,x_2 y x_3 son:

$$x_1 = -T_1 \ln(1 - u_1) \tag{10}$$

$$x_2 = -T_1 \ln(1 - u_2) \tag{11}$$

$$x_3 = \begin{cases} \frac{(x-a)^2}{(c-a)(b-a)} & 0 \le u_3 \le \frac{(b-a)}{(c-a)} \\ 1 - \frac{(c-x)^2}{(c-a)(c-b)} & \frac{(b-a)}{(c-a)} < u_3 \le 1 \\ 0 & \text{en otro caso} \end{cases} \tag{12}$$

siendo $u_i, i = 1, \ldots, 3$ uniformes en el $(0, 1)$.

El determinante jacobiano de la transformación para $0 \le u_3 \le \frac{(b-a)}{(c-a)}$ es:

$$J_1 = \begin{vmatrix} \frac{T_1}{1-u_1} & 0 & 0 \\ 0 & \frac{T_2}{1-u_2} & 0 \\ 0 & 0 & \frac{(c-a)(b-a)}{2\sqrt{u_3(c-a)(b-a)}} \end{vmatrix} \tag{13}$$

, resultando:

$$J_1 = \frac{T_1 T_2 (c-a)(b-a)}{(1-u_1)(1-u_2)2\sqrt{u_3(c-a)(b-a)}} \tag{14}$$

Luego el determinante jacobiano de la transformación para $\frac{(b-a)}{(c-a)} < u_3 \le 1$ es:

$$J_2 = \begin{vmatrix} \frac{T_1}{1-u_1} & 0 & 0 \\ 0 & \frac{T_2}{1-u_2} & 0 \\ 0 & 0 & \frac{(c-a)(c-b)}{2\sqrt{(c-a)(c-b)(1-u_3)}} \end{vmatrix} \tag{15}$$

, resultando:

$$\frac{T_1 T_2 (c-a)(c-b)}{(1-u_1)(1-u_2)2\sqrt{(c-a)(c-b)(1-u_3)}} \tag{16}$$

Reemplazando el cambio de variables en la integral **??**, se obtiene:

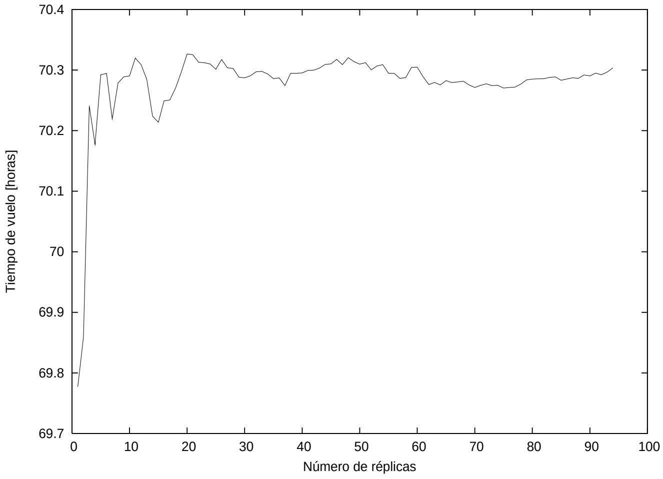

Figura 4: Tiempo medio de vuelo del USS Enterprise.

$$
\begin{aligned}
E\{T\} \;=\; & \int_0^1 \int_0^1 \int_0^{\frac{(b-a)}{c-a}} \mathbf{min}\Big\{\mathbf{max}\{-T_1 \ln(1-u_1), \\
& - T_2 \ln(1-u_2)\}, \sqrt{u_3(c-a)(b-a)} + a\Big\} d^3u_i + \\
& \int_0^1 \int_0^1 \int_{\frac{(b-a)}{c-a}}^1 \mathbf{min}\Big\{\mathbf{max}\{-T_1 \ln(1-u_1), \qquad (17) \\
& - T_2 \ln(1-u_2)\}, c - \sqrt{(1-u_3)(c-a)(c-b)}\Big\} d^3u_i
\end{aligned}
$$

4.2 Simulación de Montecarlo

Para aproximar numéricamente la integral 17 se utiliza el método de Montecarlo. Mediante simulaciones en *GNU Octave*, se obtienen $n = 1, 2, ...$ realizaciones T_i de la integral 17 utilizando 500 puntos comprendidos en $[0, 1]^3$. Los puntos se computan usando el generador que se describe en la sección 2.1 con semillas 1123 y 3542.

Para cada realización n se calcula el tiempo medio de vuelo:

$$\langle T \rangle = \frac{1}{n} \sum_{i=1}^{n} T_i \qquad (18)$$

y la varianza muestral:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (T_i - \langle T \rangle)^2 \qquad (19)$$

de tal manera que la simulación se detiene cuando

$$\|S_n^2 - S_{n-1}^2\| < \epsilon \qquad (20)$$

S_n^2 y S_{n-1}^2 son las varianzas de la media muestral para las realizaciones n y $n-1$ respectivamente mientras que ϵ es la cota de error, la cual es fijada en 5 x 10^{-5} $horas^2$.

Se obtiene para $n = 94$ realizaciones, una media muestral $\langle T \rangle = 70.30$ horas y una varianza de la media muestral $S^2 = 0.005$. En la figura **??** y **??** se muestran las gráficas de $\langle T \rangle$ y S^2 en función de la cantidad de realizaciones.

5. CONCLUSIONES

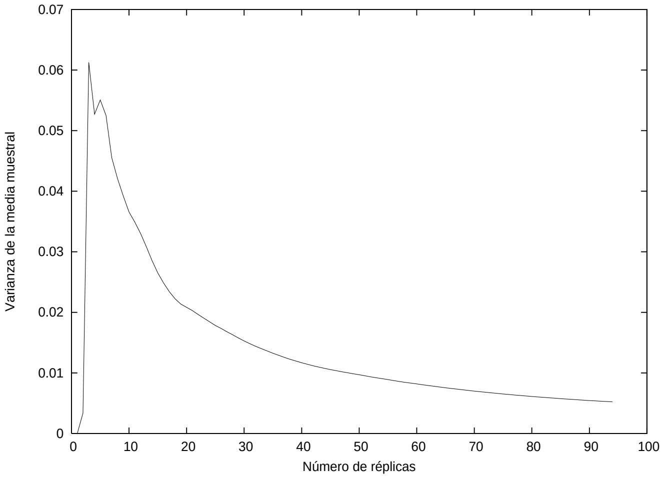

Figura 5: Error de estimación del tiempo medio de vuelo.

6. REFERENCIAS

[BANKS,1998] Jerry Banks. *Handbook of Simulation.* Wiley-IEEE, 1998.

[NEAVE] Henry R. Neave. *Elementary Statistics Tables.* Routledge, 1981.

[DIAZ,2007] Alejandro R. Díaz. *Números Pseudo aleatorios.* ITBA, 2007.