Tarea 2 Simulación

Bernardo Mondraón Brozon

4 de septiembre de 2018

Problema 1

Probar por inducción matemática que para un GLC,

$$Z_i = a^i Z_0 + c\frac{a^i - 1}{a - 1} \mod m$$

Demostración:

Para $i = 0$ se tiene que

$$\begin{aligned} Z_0 &= Z_0 \mod m \\ &= a^0 Z_0 + c\frac{a^0 - 1}{a - 1} \mod m, \end{aligned}$$

y para $i = 1$ se tiene que

$$\begin{aligned} Z_1 &= aZ_0 + c \mod m \\ &= a^1 Z_0 + c\frac{a^1 - 1}{a - 1} \mod m. \end{aligned}$$

Procediendo por indución, se supone válido para i de manera que

$$\begin{aligned} Z_{i+1} &= aZ_i + c \mod m \\ &= a\left(a^i Z_0 + c\frac{a^i - 1}{a - 1}\right) + c \mod m \\ &= a^{i+1} Z_0 + c\frac{a^{i+1} - 1}{a - 1} \mod m. \end{aligned}$$

Por lo tanto es váido para toda $i \in \mathbb{N}$.

$\blacksquare$

Problema 2

¿Qué se puede decir del periodo de $Z_i = 630,360,016 Z_{i-1} \mod 2^{31} - 1$?

Solución:

Se puede decir que su periodo es inferor a $2^{31} - 1$.

Problema 3

Sin calcular ninguna Z_i, determinar cuál de los siguientes GLC's mixtos tienen periodo completo.

a) $Z_i = 13Z_{i-1} + 13 \mod 16$.

b) $Z_i = 12Z_{i-1} + 13 \mod 16$.

c) $Z_i = 13Z_{i-1} + 12 \mod 16$.

d) $Z_i = Z_{i-1} + 12 \mod 13$.

e) El GLC con pará metros: $a = 2,814,749,767,109$, $c = 59,482,661,568,307$, $m = 248$.

Solución:

a) $Z_i = 13Z_{i-1} + 13 \mod 16$.

   El generador lineal congruencial tiene periodo completo

b) $Z_i = 12Z_{i-1} + 13 \mod 16$.

   El generador lineal congruencial no tiene periodo completo

c) $Z_i = 13Z_{i-1} + 12 \mod 16$.

   El generador lineal congruencial no tiene periodo completo

d) $Z_i = Z_{i-1} + 12 \mod 13$.

   El generador lineal congruencial tiene periodo completo

e) El GLC con pará metros: $a = 2,814,749,767,109$, $c = 59,482,661,568,307$, $m = 248$.

   El generador lineal congruencial tiene periodo completo

Problema 4

Mostrar que el promedio de las U_i's tomadas de un ciclo completo de un GLC de periodo completo es $\frac{1}{2} - \frac{1}{2m}$.

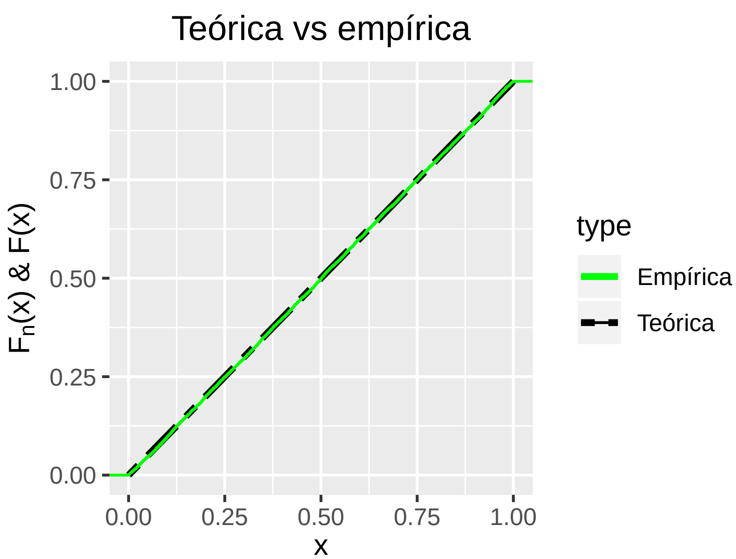
Problema 5

Generar 10,000 números uniformemente distribuidos entre 0 y 1 en Excel. Hacer un breve estudio para probar la calidad de los generadores; aplicar las pruebas de uniformidad e independencia a cada conjunto de datos. Resumir resultados en NO MAS de 2 cuartillas, incluyendo gráficas. De acuerdo a tus resultados, ¿cómo calificarías al generador de Excel?

Solución:

Pruebas de bondad de ajuste

A continuación se muestra la gráfica de la distribución empírica de los valores generados con Excel y la gráfica de la distribución uniforme en el intervalo $[0, 1]$:



Se realizará la prueba de Kolmogorov-Smirnov (prueba de bondad de ajuste) a los valores simulados para determinar si efectivamente provienen de la distribucion uniforme en el intervalo $[0, 1]$. Para realizar la prueba se contrastaran las siguientes hipótesis:

$$H_0 : F(x) = x \quad \forall x \quad \text{vs.} \quad H_1 : F(x) \neq x \quad \text{para alguna } x$$

Entonces la estadística de prueba está dada por

$$D_n = \max_x |F_n(x) - x|.$$

Se rechazará la hipótesis H_0 si D_n es muy grande. En este caso, la estadística de prueba toma el siguiente valor

$$D_n = 0.0055188$$

El valor p correspondiente al valor de la estadística de prueba está dado por

$$\text{p-value} = 1.0876343$$

En este caso el valor p es mayor a 1 porque se calculó mediante una aproximación, sin embargo el verdadero valor es muy cercano a 1, de manera que no se rechaza la hipótesis de que los números aleatorios provienen de la distribución uniforme.

Aplicando la prueba de Crámer-von Misses se obtiene el siguiente valor de la estadística de prueba:

$$W_n = n \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 \, \mathrm{d}F(x) = 10000 \int_{-\infty}^{\infty} (F_n(x) - x)^2 \, \mathrm{d}x = 0.0459533.$$

De manera que el valor p está dado por

$$\text{p-value} = 0.9003661$$

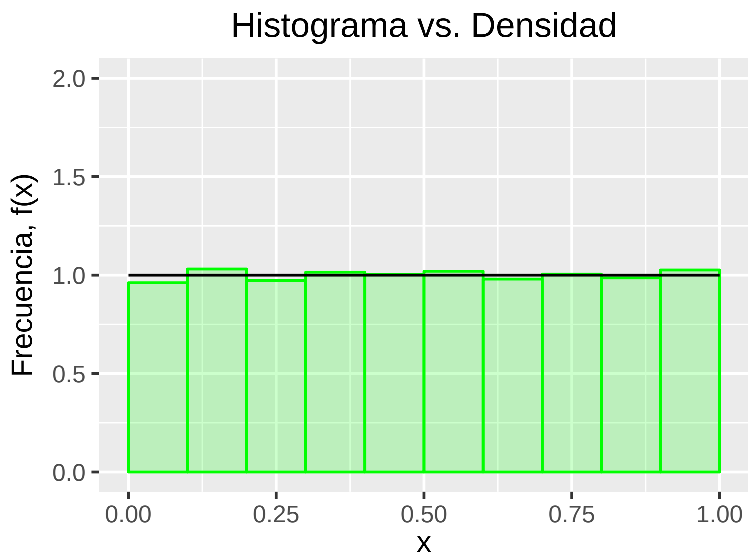
Aplicando la prueba de Anderson-Darling se obtiene el siguiente valor de la estadística de prueba:

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)\,(1 - F(x))} \mathrm{d}F(x) = 10000 \int_{-\infty}^{\infty} \frac{(F_n(x) - x)^2}{x\,(1 - x)} \mathrm{d}x = \infty$$

De manera que el valor p está dado por

$$\text{p-value} = 6 \times 10^{-8}$$

A continuación se muestra el histograma de los datos generados en Excel y la función de densidad uniforme en el intervalo $[0, 1]$:



Para la prueba de bondad de ajuste χ^2 de Pearson, las hipótesis a contrastar son las mismas que las de la prueba de Kolmogorov-Smirnov En este caso la estadistica de prueba toma el siguinete valor:

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_j - np_j)^2}{np_j} = \sum_{i=1}^{1} \frac{(N_j - 10000p_j)^2}{10000p_j} = 2026.8$$

De manera que el valor p está dado por

$$\text{p-value} = 0.3270129$$

Por lo tanto no se rechaza la hipótesis nula de que los datos generados por Excel provienen de la distribución uniforme en el intervalo $[0, 1]$.

A continuación se presenta un qq-plot de lo valores generados por Excel. Si el qq'plot sigue la recta identidad cuando se grafica contra la distribución teórica, entonces se puede decir que los datos siguien adecuadamente la distribución uniforme en el intervalo $[0, 1]$.

Al parecer se tiene un buen ajuste, sin embargo, esto no es una prueba en sentido estricto; es simplemente una guía visual para determinar si se tiene un buen ajuste.

Pruebas de independencia

Aplicando la prueba de independencia que utiliza a las rachas en la sucesión de números aleatorios se obtiene el siguiente número de rechas:

$$R = 6619$$

Mediante la aproximación normal se tiene el siguiente valor del estadístico de prueba que sigue la distribucion normal estandar:

$$Z = \frac{R - (2n-1)/3}{\sqrt{(}\,(16R - 29)/90)} = -1.3800399.$$

El correspondiente valor p para el valor dado de la estadística de prueba está dado por

$$\text{p-value} = 0.0837872$$

Aplicando la prueba de independencia de Levene-Wolfowitz se obtiene el siguiente valor para la estadística de prueba:

$$R_n = 9.8723467.$$

El correspondiente valor p para el valor dado de la estadística de prueba está dado por

$$\text{p-value} = 0.1301317$$

Aplicando la prueba de gaps se obtiene el siguiente valor para la estadística de prueba:

$$R_n = 8.99.$$

El correspondiente valor p para el valor dado de la estadística de prueba está dado por

$$\text{p-value} = 0.77$$

Aplicando la prueba del poker se obtiene el siguiente valor para la estadística de prueba:

$$R_n = 7.07.$$

El correspondiente valor p para el valor dado de la estadística de prueba está dado por

$$\text{p-value} = 0.13$$

Aplicando pruebas de autocorrelacion, se obitnene los siguiente:

```
## [1] 0.77
```

```
##
##  Box-Pierce test
##
## data:  unifs
## X-squared = 1, df = 3, p-value = 0.8
```

```
##
##  Box-Ljung test
##
## data:  unifs
## X-squared = 1, df = 3, p-value = 0.8
```

De acuerdo a las pruebas anteriores, se puede concluir que el generador de números aleatorios de excel es un buen generador de números aleatorios. De manera que no se rechaza la hipótesis de que los números son independientes.

Problema 6

Probar que la parte fraccional de la suma de uniformes en $[0, 1]$: $U1 + \cdots + U_k$ es también uniforme en el intervalo $[0, 1]$.

Solución:

Sea $\{x\} = x - \lfloor x \rfloor$ la parte fraccional del número x, y $S_k = \sum_{i=1}^{k} U_i$ la suma de k variables aleatorias independientes e uniformemente distribuidas en el intervalo $[0, 1]$. Para $k = 1$ se tiene que $\{S_1\} = \{U_1\} = U_1 \sim U[0, 1]$. Para $k = 2$ se tiene que $\{S_2\} = \{U_1 + U_2\}$ con

$$f_{S_2}(s) = \left\{ \begin{array}{ll} s & 0 \leq s \leq 1 \\ 2 - s & 1 < s \leq 2 \end{array} \right. .$$

de manera que

$$\begin{aligned} F_{\{S_2\}}(x) &= Pr\left[\{S_2\} \leq x\right] \\ &= \int_{u=0}^{s} f_{S_2}(u)\, du + \int_{u=1}^{1+s} f_{S_2}(u)\, du = s. \end{aligned}$$

Entonces $\{S_2\}$ sigue la distribución uniforme en el intervalo $[0, 1]$. Procediendo mediante inducción se supone válido para k, de manera que

$$\{S_k\} \sim U[0, 1].$$

Por lo tanto, para $k + 1$ se tiene que

$$\{S_{k+1}\} = \{U_1 + \cdots + U_{k+1}\} = \{\{U_1 + \cdots + U_k\} + U_{k+1}\} = \{\{S_k\} + U_{k+1}\}.$$

Pero $\{S_k\}$ y U_{k+1} tienen la distribución uniforme y son independientes, entonces

$$\{S_{k+1}\} \sim U[0, 1].$$

■