

Simulación con hojas de cálculo de código abierto: OpenOffice.org y Gnumeric

Sánchez García, Juan Francisco (jf.sanchez@upct.es)
Dpto. Métodos Cuantitativos e Informáticos
Universidad Politécnica de Cartagena

RESUMEN

Una de las posibles alternativas para realizar técnicas de simulación dentro del ámbito de la investigación es utilizar una aplicación de hoja de cálculo, y hacer uso de las funciones propias de las que dispone dicha aplicación. Evidentemente, existen muchas otras aplicaciones especializadas en tareas de simulación pero la hoja de cálculo presenta la característica de su fácil disposición en cualquier organización (ya sea ésta la administración pública o una empresa) unido a su bajo coste que en algunos casos concretos puede llegar a ser nulo por existir aplicaciones de tipo gratuito y de código abierto.

En este trabajo se va a realizar la validación de las funciones que las hojas de cálculo incorporan para la generación de números pseudoaleatorios, parte vital de cualquier técnica de simulación, y se hará prestando especial atención a las hojas de cálculo de “código abierto”, tomando como exponente de este tipo de software las dos más difundidas para el sistema operativo Windows: OpenOffice.org y Gnumeric.

Palabras clave: Simulación, hoja de cálculo, números pseudoaleatorios

Clasificación JEL (*Journal Economic Literature*): C15, C88

Área temática: Informática aplicada a los Métodos Cuantitativos

1. INTRODUCCIÓN

Las técnicas de simulación cobran especial importancia cuando es necesario reproducir el funcionamiento de un determinado sistema. En ocasiones, es suficiente con poder realizar cálculos finales sobre el funcionamiento del sistema, obteniéndose valores medios. Sin embargo, no siempre basta con dichos cálculos sino que es conveniente ver las posibles variaciones que pueden presentarse con respecto a los valores medios calculados. Es en este punto donde tienen una vital importancia las técnicas de simulación, las cuales se han visto potenciadas a partir de la aparición de los equipos informáticos y del software adecuado.

Dentro del software propio para efectuar la simulación existen diversos paquetes que permiten llevar a cabo estas técnicas. Sin embargo, a veces, es preferible construir los modelos de simulación a partir de aplicaciones estándar como son las hojas de cálculo de forma que se pueda decidir cómo han de obtenerse las simulaciones y además permitan ver todos los cálculos intermedios necesarios hasta llegar al resultado final. Entre todas las aplicaciones existentes de hoja de cálculo la más utilizada es Microsoft® Excel debido a su buen funcionamiento y a su disponibilidad casi universal para todo tipo de usuarios, que le ha convertido prácticamente en un estándar dentro de este tipo de aplicaciones. Este hecho, adicionalmente, hace que sea posible portar un modelo efectuado con esta aplicación a otras aplicaciones similares con bastantes garantías de éxito.

Junto al software comercial representado por Excel, y otras aplicaciones como Lotus 1-2-3 de Lotus Software (propiedad de IBM) y Quattro Pro de Corel Corporation existen soluciones de código abierto, entendido este software como aquel que es de libre distribución y además permite la modificación y redistribución del código fuente. En este último tipo de software se sitúan soluciones como OpenOffice.org (OOo) y Gnumeric, entre otras.

2. NÚMEROS PSEUDOALEATORIOS

2.1. Métodos de generación de números pseudoaleatorios

La simulación de valores aleatorios distribuidos de acuerdo con cualquier distribución estadística se basan en la generación, en primer lugar, de valores aleatorios uniformemente distribuidos entre 0 y 1. Realmente, los números generados no son números aleatorios, sino pseudoaleatorios pues no son debidos al azar, sino que proceden de cálculos matemáticos que tratan de imitar al azar.

Existen diversos métodos comúnmente utilizados en la literatura, como los propuestos por Álvarez Madrigal [1], Coss [2], Escudero [3] y Rubinstein [7], para generación de valores pseudoaleatorios:

- Método de los cuadrados medios.
- Técnica de mitad del producto.
- Método del multiplicador constante.
- Método congruencial.
- Método congruencial aditivo.
- Método congruencial lineal.

2.2. Validación de los números pseudoaleatorios generados

Una vez que se han generado los valores pseudoaleatorios según la distribución uniforme se debe comprobar que efectivamente están uniformemente distribuidos, lo que significa que son uniformes e independientes. En este sentido, existen pruebas específicas para comprobar si una secuencia de números presenta las propiedades estadísticas que se podría esperar si fuesen debidos al azar. Una de estas pruebas es

la “Diehard Battery of Tests of Randomness”, desarrollada por el profesor George Marsaglia [5] en el año 1995.

Para probar la uniformidad se aplica la prueba de Kolmogorov-Smirnov, la prueba de la χ^2 y la prueba de los promedios.

- Prueba de Kolmogorov-Smirnov. En esta prueba se estima la probabilidad de que una distribución de datos sea igual a una distribución conocida mediante la comparación de las funciones de probabilidad acumulada:
 1. Se determina la función de distribución acumulada que se supone que siguen los valores generados.
 2. Se calcula la función de distribución acumulada de los valores generados.
 3. Se obtiene la máxima diferencia entre ambas funciones de distribución.
 4. Se determina el valor máximo permitido para la diferencia a partir de la tabla de la prueba de Kolmogorov-Smirnov.
 5. Se determina si existe o no, para un determinado nivel de significación α , diferencia entre ambas funciones de probabilidad.

- Prueba de la χ^2 . Para esta prueba se necesita un mínimo de 50 valores generados.
 1. Se organizan los valores obtenidos en un histograma (con n clases, donde el número de elementos esperados para cada clase debe ser al menos de 5).
 2. Se calcula el número de datos que debería haber en cada clase suponiendo una distribución uniforme teórica.
 3. Se calcula el estadístico χ_0^2 con las diferencias existentes entre las cantidades de valores generados (O_i) y de valores ideales (E_i) para cada clase

n del histograma de acuerdo con la siguiente expresión.

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

4. Se establece el nivel máximo de variación del estadístico χ_0^2 para $n - 1$ grados de libertad con nivel de significación α .
5. Se compara el estadístico calculado con el máximo permitido y si el primero es menor, no hay evidencia para afirmar que los valores generados no están uniformemente distribuidos.

- Prueba de los promedios. La función de densidad uniforme se define como:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & 0 > x > 1 \end{cases}$$

En esta expresión, x es una variable aleatoria uniforme definida en el intervalo $[0, 1]$. La distribución acumulada $F(x)$ de una variable x aleatoria uniformemente distribuida se puede obtener como:

$$F(x) = \int_0^x 1 dx = x$$

El valor esperado y la varianza de dicha variable están definidos por:

$$E(x) = \int_0^x x dx = \frac{1}{2}$$

$$var(x) = \int_0^1 \left(x - \frac{1}{2}\right)^2 dx = \frac{1}{12}$$

Con estos parámetros se puede plantear un contraste cuya hipótesis nula sería que la media debe ser $1/2$, siendo el estadístico Z el siguiente:

$$Z_0 = \frac{(\bar{x} - 1/2)\sqrt{N}}{\sqrt{1/12}}$$

Si $|Z_0| < Z_{\alpha/2}$ no se puede rechazar la hipótesis de que los números pseudo-aleatorios generados provienen de una distribución uniforme con media $1/2$.

Para probar la independencia se utiliza el test de rachas.

- Test de rachas (The-Up-and-Down-Test) [6].
 1. La magnitud de cada elemento de la lista es comparado con su predecesor y se le asigna un valor + o un valor – según que éste sea inferior o superior al elemento actual.
 2. Se cuenta el número de veces que se produce un cambio de valor de + a – ó viceversa (U).
 3. El número obtenido anteriormente se distribuye de acuerdo a una normal con media y desviación típica conocidas:

$$\frac{2N - 1}{3} \quad \text{y} \quad \sqrt{\frac{16N - 29}{90}}$$

4. Levene [4] demostró que el test de hipótesis a aplicar es:

$$-Z_{\alpha/2} \leq \frac{U - (2N - 1)/3}{\sqrt{(16N - 29)/90}} \leq Z_{\alpha/2}.$$

2.3. Hojas de cálculo

Para la generación de valores pseudoaleatorios uniformemente distribuidos, las hojas de cálculo disponen de funciones como `ALEATORIO()` en Excel y OpenOffice.org, y `RAND()` en Gnumeric.

Antes de poder utilizar estos números generados en técnicas de simulación es preciso comprobar, previamente, si los valores obtenidos cumplen los requisitos exigibles de uniformidad e independencia. Para poder comprobar estos extremos se ha procedido a la generación simultánea de 50.000 valores y a continuación se han efectuado las correspondientes pruebas estadísticas, obteniéndose los resultados que se muestran en las figuras 1, 2 y 3. En todos los casos los resultados han sido satisfactorios.

2.3.1. Microsoft Excel

La experiencia de Microsoft en software de hoja de cálculo data del año 1982, cuando lanzó su primera aplicación: *Multiplan*, que no llegó a tener demasiado éxito. La primera aparición de Excel en el mercado se produjo en el año 1985, con Macintosh, y la primera aparición en el mercado para el sistema Windows se produjo en el año 1987, con Excel 2.0. Excel 5.0 salió en el año 1994 y fue la primera versión en incorporar VBA (Visual Basic para Aplicaciones). Excel 2003 (Excel 11) está en el mercado desde 2003.

2.3.2. OpenOffice.org Calc

OpenOffice.org (OOo) es una suite informática desarrollada a partir del código fuente de StarOffice liberado por Sun Microsystems en julio de 2000. A su vez Sun había adquirido StarOffice tras su compra a StarDivision en el año 1999. Actualmente, la última versión final disponible es la 2.4.0 y está en fase de desarrollo la versión 3.0. Algunas de las principales características de esta aplicación son la disponibilidad de su código fuente, la lectura y escritura de ficheros de Microsoft Office y la interfaz similar con la versión 2003 de dicha aplicación. Por último, cabe destacar que maneja el formato abierto de ficheros ODF (Open Document Format). La hoja de cálculo que incorpora la suite OpenOffice.org recibe el nombre de OpenOffice.org Calc.

2.3.3. Gnumeric

Gnumeric es una hoja de cálculo libre que forma parte del entorno de escritorio GNOME, propio de sistemas operativos de tipo Unix basados en la tecnología X Window. Al igual que OpenOffice.org Calc es capaz de leer y escribir ficheros XLS de Excel.

| | B | C | D | E | F |
|----|---|-------------------|-------------------|---------------------|-------------------|
| 2 | CLASE | FRECUENCIA | F.R.ACUM. | TEÓRICA | DIFERENCIA |
| 3 | 0,1 | 4997 | 0,09994 | 0,1 | 0,000 |
| 4 | 0,2 | 5065 | 0,20124 | 0,2 | 0,001 |
| 5 | 0,3 | 4919 | 0,29962 | 0,3 | 0,000 |
| 6 | 0,4 | 5051 | 0,40064 | 0,4 | 0,001 |
| 7 | 0,5 | 5026 | 0,50116 | 0,5 | 0,001 |
| 8 | 0,6 | 4971 | 0,60058 | 0,6 | 0,001 |
| 9 | 0,7 | 5011 | 0,7008 | 0,7 | 0,001 |
| 10 | 0,8 | 5028 | 0,80136 | 0,8 | 0,001 |
| 11 | 0,9 | 5018 | 0,90172 | 0,9 | 0,002 |
| 12 | 1 | 4914 | 1 | 1 | 0,000 |
| 13 | TOTAL | | 50000 | | |
| 14 | | | | | |
| 15 | TEST KOLMOGOROV-SMIRNOV | | | | |
| 16 | Diferencia máxima | | 0,001720 | | |
| 17 | Estimador Kolmogorov-Smirnov | | | | |
| 18 | $D_{0,05;1000}$ | | 0,006082 | | |
| 19 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |
| 20 | | | | | |
| 21 | CLASE | FRECUENCIA | F. TEÓRICA | CHI CUADRADO | |
| 22 | 0,1 | 4997 | 5000 | 0,001800 | |
| 23 | 0,2 | 5065 | 5000 | 0,845000 | |
| 24 | 0,3 | 4919 | 5000 | 1,312200 | |
| 25 | 0,4 | 5051 | 5000 | 0,520200 | |
| 26 | 0,5 | 5026 | 5000 | 0,135200 | |
| 27 | 0,6 | 4971 | 5000 | 0,168200 | |
| 28 | 0,7 | 5011 | 5000 | 0,024200 | |
| 29 | 0,8 | 5028 | 5000 | 0,156800 | |
| 30 | 0,9 | 5018 | 5000 | 0,064800 | |
| 31 | 1 | 4914 | 5000 | 1,479200 | |
| 32 | TOTAL | | 50000 | 4,707600 | |
| 33 | | | | | |
| 34 | TEST CHI CUADRADO | | | | |
| 35 | CHI-CUADRADO | | 4,707600 | | |
| 36 | Estimador | | | | |
| 37 | $\chi^2_{0,05;9}$ | | 16,918978 | | |
| 38 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |
| 39 | | | | | |
| 40 | TEST DE LOS PROMEDIOS | | | | |
| 41 | MEDIA | | 0,499250 | | |
| 42 | Z | | -0,581144 | | |
| 43 | $Z_{0,025}$ | | 1,959964 | | |
| 44 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |

| | B | C | D |
|----|---|---|--------------|
| 46 | TEST DE RACHAS | | |
| 47 | U | | 33256,000000 |
| 48 | MEDIA | | 33333,000000 |
| 49 | DES.V.TÍPICA | | 94,279195 |
| 50 | Z | | -0,816723 |
| 51 | $Z_{0,025}$ | | 1,959964 |
| 52 | NO SE RECHAZA LA HIPÓTESIS DE INDEPENDENCIA | | |

Figura 1: Microsoft Excel: Pruebas de uniformidad e independencia

2.4. Tiradas de simulación

Para comprobar la exactitud de los cálculos mostrados se han realizado 100 tiradas de simulación y se han anotado, utilizando la programación de scripts que permiten las aplicaciones estudiadas (Microsoft VBA y OpenOffice.org Basic), los valores de los estadísticos calculados en cada una de las tiradas, con el fin de poder comprobar si en algún momento se llega a rechazar alguna de las hipótesis planteadas.

| | B | C | D | E | F |
|----|---|-------------------|-------------------|---------------------|-------------------|
| 2 | CLASE | FRECUENCIA | F.R.ACUM. | TEÓRICA | DIFERENCIA |
| 3 | 0,1 | 5085 | 0,1 | 0,1 | 0,002 |
| 4 | 0,2 | 4984 | 0,2 | 0,2 | 0,001 |
| 5 | 0,3 | 4930 | 0,3 | 0,3 | 0,000 |
| 6 | 0,4 | 5032 | 0,4 | 0,4 | 0,001 |
| 7 | 0,5 | 4968 | 0,5 | 0,5 | 0,000 |
| 8 | 0,6 | 4910 | 0,6 | 0,6 | 0,002 |
| 9 | 0,7 | 5041 | 0,7 | 0,7 | 0,001 |
| 10 | 0,8 | 4956 | 0,8 | 0,8 | 0,002 |
| 11 | 0,9 | 4997 | 0,9 | 0,9 | 0,002 |
| 12 | 1 | 5097 | 1 | 1 | 0,000 |
| 13 | TOTAL | 50000 | | | |
| 14 | | | | | |
| 15 | TEST KOLMOGOROV-SMIRNOV | | | | |
| 16 | Diferencia máxima | | | | 0,001940 |
| 17 | Estimador Kolmogorov-Smirnov | | | | |
| 18 | $D_{0,05;1000}$ | | | | 0,006082 |
| 19 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |
| 20 | | | | | |
| 21 | CLASE | FRECUENCIA | F. TEÓRICA | CHI CUADRADO | |
| 22 | 0,1 | 5085 | 5000 | 1,445000 | |
| 23 | 0,2 | 4984 | 5000 | 0,051200 | |
| 24 | 0,3 | 4930 | 5000 | 0,980000 | |
| 25 | 0,4 | 5032 | 5000 | 0,204800 | |
| 26 | 0,5 | 4968 | 5000 | 0,204800 | |
| 27 | 0,6 | 4910 | 5000 | 1,620000 | |
| 28 | 0,7 | 5041 | 5000 | 0,336200 | |
| 29 | 0,8 | 4956 | 5000 | 0,387200 | |
| 30 | 0,9 | 4997 | 5000 | 0,001800 | |
| 31 | 1 | 5097 | 5000 | 1,881800 | |
| 32 | TOTAL | 50000 | | 7,112800 | |
| 33 | | | | | |
| 34 | TEST CHI CUADRADO | | | | |
| 35 | CHI-CUADRADO | | | | 7,112800 |
| 36 | Estimador | | | | |
| 37 | $\chi^2_{0,059}$ | | | | 16,918978 |
| 38 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |
| 39 | | | | | |
| 40 | TEST DE LOS PROMEDIOS | | | | |
| 41 | MEDIA | | | | 0,500317 |
| 42 | Z | | | | 0,245843 |
| 43 | $Z_{0,025}$ | | | | 1,959964 |
| 44 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |
| 46 | TEST DE RACHAS | | | | |
| 47 | U | | | | 33336,000000 |
| 48 | MEDIA | | | | 33333,000000 |
| 49 | DESV.TÍPICA | | | | 94,279195 |
| 50 | Z | | | | 0,031820 |
| 51 | $Z_{0,025}$ | | | | 1,959964 |
| 52 | NO SE RECHAZA LA HIPÓTESIS DE INDEPENDENCIA | | | | |

Figura 2: OpenOffice.org Calc: Pruebas de uniformidad e independencia

Con OpenOffice.org Calc, las pruebas de la χ^2 y de los promedios rechazan la hipótesis de uniformidad en 3 casos, mientras que el test de rachas rechaza la hipótesis de independencia en 2 casos y el de Kolmogorov-Smirnov rechaza la hipótesis de uniformidad en 1 caso. Efectuando el mismo análisis con Excel, el test que peores resultados arroja es el de los promedios ya que en 11 ocasiones se rechaza la hipótesis de uniformidad. El resto de test tienen una probabilidad de rechazo entre el 5 y el 6 %.

| | B | C | D | E | F |
|----|---|-------------------|-------------------|---------------------|-------------------|
| 2 | CLASE | FRECUENCIA | F.R.ACUM. | TEÓRICA | DIFERENCIA |
| 3 | 0,1 | 4971 | 0,09942 | 0,1 | 0,001 |
| 4 | 0,2 | 4957 | 0,19856 | 0,2 | 0,001 |
| 5 | 0,3 | 5132 | 0,3012 | 0,3 | 0,001 |
| 6 | 0,4 | 4864 | 0,39848 | 0,4 | 0,002 |
| 7 | 0,5 | 4934 | 0,49716 | 0,5 | 0,003 |
| 8 | 0,6 | 5080 | 0,59676 | 0,6 | 0,001 |
| 9 | 0,7 | 5091 | 0,70058 | 0,7 | 0,001 |
| 10 | 0,8 | 5079 | 0,80216 | 0,8 | 0,002 |
| 11 | 0,9 | 4915 | 0,90046 | 0,9 | 0,000 |
| 12 | 1 | 4977 | 1 | 1 | 0,000 |
| 13 | TOTAL | 50000 | | | |
| 14 | TEST KOLMOGOROV-SMIRNOV | | | | |
| 15 | Diferencia máxima 0,002840 | | | | |
| 16 | Estimador Kolmogorov-Smirnov | | | | |
| 17 | $D_{0,05;1000}$ 0,006082 | | | | |
| 18 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |
| 19 | | | | | |
| 20 | | | | | |
| 21 | CLASE | FRECUENCIA | F. TEÓRICA | CHI CUADRADO | |
| 22 | 0,1 | 4971 | 5000 | 0,168200 | |
| 23 | 0,2 | 4957 | 5000 | 0,369800 | |
| 24 | 0,3 | 5132 | 5000 | 3,484800 | |
| 25 | 0,4 | 4864 | 5000 | 3,699200 | |
| 26 | 0,5 | 4934 | 5000 | 0,871200 | |
| 27 | 0,6 | 5080 | 5000 | 1,280000 | |
| 28 | 0,7 | 5091 | 5000 | 1,656200 | |
| 29 | 0,8 | 5079 | 5000 | 1,248200 | |
| 30 | 0,9 | 4915 | 5000 | 1,445000 | |
| 31 | 1 | 4977 | 5000 | 0,105800 | |
| 32 | TOTAL | 50000 | | 14,328400 | |
| 33 | TEST CHI CUADRADO | | | | |
| 34 | CHI-CUADRADO 14,328400 | | | | |
| 35 | Estimador | | | | |
| 36 | $c_{n,\alpha}^2$ 16,918978 | | | | |
| 37 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |
| 38 | | | | | |
| 39 | | | | | |
| 40 | TEST DE LOS PROMEDIOS | | | | |
| 41 | MEDIA 0,500680 | | | | |
| 42 | Z 0,526481 | | | | |
| 43 | $Z_{0,025}$ 1,959964 | | | | |
| 44 | NO SE RECHAZA LA HIPÓTESIS DE UNIFORMIDAD | | | | |

| | B | C | D |
|----|---|---|--------------|
| 46 | TEST DE RACHAS | | |
| 47 | U | | 33348,000000 |
| 48 | MEDIA | | 33333,000000 |
| 49 | DESV. TÍPICA | | 94,279195 |
| 50 | Z | | 0,159102 |
| 51 | $Z_{0,025}$ | | 1,959964 |
| 52 | NO SE RECHAZA LA HIPÓTESIS DE INDEPENDENCIA | | |

Figura 3: Gnumeric: Pruebas de uniformidad e independencia

3. CONCLUSIONES

La utilización de software de código abierto se presenta como una alternativa para realizar todo tipo de simulaciones, obteniendo resultados igual de válidos que con una aplicación comercial como puede ser Microsoft Excel. La generación de números pseudoaleatorios mediante las funciones que incorporan OpenOffice.org Calc y Gnumeric proporciona valores uniformes e independientes que pueden ser aplicados en cualquier técnica de simulación.

El trabajo aquí presentado puede y debería ser complementado utilizando los

trabajos de Marsaglia, así como la nueva versión de OpenOffice.org Calc, que se encuentra en fase de desarrollo en el momento de realizarse el mismo.

REFERENCIAS BIBLIOGRÁFICAS

- [1] ALVAREZ MADRIGAL, M. (1997). *Curso virtual de Investigación de Operaciones*. Instituto Tecnológico y de Estudios Superiores de Monterrey. Campus Estado de México.
<http://webdia.cem.itesm.mx/ac/alvarez/acceso.html>
- [2] COSS BU, R. (1982). *Simulación. Un enfoque práctico*. Limusa, México D.F.
- [3] ESCUDERO, L. F. (1973). *La simulación en la empresa*. Ediciones Deusto, Bilbao.
- [4] LEVENE, M. (1952). "On the power function of tests of randomness based on runs up and down". *Annals of Mathematical Statistics*, Vol. 23, pp. 34-56.
- [5] MARSAGLIA, G. (1995). *The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness*.
<http://www.stat.fsu.edu/pub/diehard/>
- [6] NAYLOR, T. ET. AL. (1966). *Computer simulation techniques*. John Wiley & Sons, New York.
- [7] RUBINSTEIN, R. Y. (1981). *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York.