Los números aleatorios y la ingeniería

En el presente trabajo se llama la atención sobre las propiedades que deben cumplir los Generadores de Números Aleatorios dado que estos son el corazón mismo de la Simulación.

Esta presentación se inicia con una introducción al tema, luego se comentan los métodos de generación, propiedades, tipos y comprobación a que deben someterse los generadores. Finalmente, se dan las conclusiones.

LUIS GERARDO ASTAIZA A.
Ingeniero Mecánico, M.I.S.
Profesor Asociado
Universidad Nacional

Indudablemente, la mayoría de los ingenieros se ha enfrentado a problemas para los cuales ha sido imposible obtener una solución analítica. En particular, problemas caracterizados por: muchas variables y sus funciones; variables aleatorias y sus distribuciones; muchos parámetros; no linealidades; restricciones diversas así como una o varias respuestas. Para solucionar este tipo de problemas se ha empleado la simulación en computadoras. Las alternativas al uso de la simulación son el análisis matemático, la experimentación con el sistema real o un prototipo de él o la dependencia de la experiencia o la intuición.

Todas, incluyendo la simulación, tienen limitaciones. Muy a menudo, el análisis matemático de sistemas complejos es imposible; la experimentación con sistemas pilotos o reales es costosa y requiere tiempo, además de que las variables pertinentes no siempre están sujetas a control. Frecuentemente, la intuición y la experiencia son las únicas alternativas a la simulación en computadoras, pero puede ser muy poco apropiada.

Así, la simulación es una técnica a la que se recurre en última instancia, pero su uso es cada vez más frecuente ya que da respuestas a pesar de sus dificultades, costo y tiempo requerido.

Sin embargo, para que esta sea posible es necesario contar con procedimientos capaces de producir números aleatorios.

Entiéndese por número aleatorio aquel que puede ser generado con igual probabilidad y en forma independiente de cualquier resultado previo. Estadísticamente, esto significa que los números son variables aleatorias, independientes y con distribución uniforme. Dada su importancia, se hace necesario conocer algunos conceptos relacionados con ellos, tales como métodos de generación; propiedades; pruebas a que son sometidos; tipos de generadores y su uso en la generación de valores de variables aleatorias que siguen, bien sea una distribución conocida (exponencial, uniforme, etc.) o una distribución empírica (datos).

METODOS DE GENERACION

El tema de los números aleatorios es complicado e implica el uso de: álgebra abstracta, teoría de

números, programación de sistemas y consideraciones de Hardware. Sin embargo, su utilización se lleva a cabo desde hace mucho tiempo. En primer lugar, están los métodos manuales, tal como el uso de la ruleta, la cual al detenerse marca con una aguja uno de los números marcados en la tabla. En segundo lugar, está el uso de tablas de números aleatorios, disponibles en textos de estadística. Finalmente, con el advenimiento de los computadores (Digital y Analógico) se abrieron nuevos horizontes para el desarrollo de métodos de generación.

A este aspecto Tocker (14) ha sugerido tres métodos para producir números aleatorios cuando se usan computadores digitales: Provisión externa; generación interna a partir de un proceso físico al azar y la generación interna de sucesiones de dígitos por medio de una relación de recurrencia. El primer método implica la grabación de tablas de números aleatorios, por ejemplo, las tablas de la Rand, en archivo magnético, a fin de tratar los números como datos de entrada para un determinado problema. Una objeción a este método es la lentitud del proceso de entrada. El segundo método hace uso de un aditamento especial de la computadora capaz de registrar los resultados de algún proceso aleatorio y además, reducir estos resultados a sucesiones de dígitos. De los procesos aleatorios que se emplean para generar dígitos con este método, se incluyen el decaimiento de los materiales radiactivos y el ruido térmico en un circuito de válvulas electrónicas. El defecto principal de este método es que los resultados no se pueden reproducir, por lo que no es posible comprobar los cálculos efectuados. El tercer método y el cual es el más aceptable para computadores digitales, implica la generación de "números pseudo aleatorios" por medio de una ecuación recursiva y algorítmica.

Esto significa que los resultados anteriores pueden ser utilizados para determinar los cálculos siguientes. Este método supera las dificultades anteriores ya que no requiere dispositivos de entrada y permite su reproducción.

Tocker ha aclarado que la principal objeción a esta solución radica en los aspectos un tanto filosóficos respecto a que una sucesión de dígitos, generada mediante una regla puramente determinística, resulta ser antítesis directa de una sucesión aleatoria. Sin embargo, esta objeción puede superarse, al menos parcialmente, al tomar el punto de vista un tanto pragmático de que una sucesión puede considerarse aleatoria si satisface un conjunto de pruebas estadísticas de aleatoriedad previamente determinadas.

PROPIEDADES

Antes de considerar las pruebas estadísticas, debemos mencionar las propiedades que debe presentar un "buen" generador de números aleatorios. En primer lugar, los números producidos deben lucir como obtenidos de una distribución uniforme U(0,1) o sea: su media debe ser 1/2, segundo momento o suma promedio de los cuadrados debe ser 1/3 y tercer momento o la suma promedio de los cubos debe ser 1/4. Adicionalmente, si tuviésemos que dividir el intervalo (0,1) en K subintervalos el J^{th} subintervalo debería contener N/K números, donde N es el tamaño de la muestra. En segundo lugar, los números aleatorios deben ser independientes. Esto es, debemos esperar sucesos independientes y para cualquier intervalo dado, una probabilidad constante de que un valor dado caerá en ese intervalo. Estas propiedades deben cumplirse independientemente del tamaño de la muestra.

Finalmente, además de producir números aleatorios que cumplan las propiedades mencionadas, el generador debe ser:

1. Rápido; o sea, que debe generar un número en el menor tiempo posible.
2. De programa breve; o sea no debe requerir grandes cantidades de memoria.
3. Capaz de producir un conjunto de números aleatorios diferentes o reproducir una serie de números.
4. De naturaleza no degenerativa. La degeneración significa que el generador produce continuamente el mismo número. Si este fenómeno se presenta, el programa debe poder hacer correcciones y seguir adelante.
5. Presentar un período largo. El período de un generador de números aleatorios es una medida de la cantidad de números que se generan antes de que reaparezca la misma secuencia de números.

TIPOS DE GENERADORES

Los generadores de números aleatorios los podemos clasificar en: Generadores basados en métodos de congruencia; generadores compuestos y generadores de Tausworthe. Los generadores basados en métodos de congruencia se pueden explicar en la siguiente forma. Consideremos la relación:

$$U = V \pmod{T} \qquad [1]$$

donde U y V son dos números reales cualesquiera. Si la diferencia entre estos números es aún divisible por un número entero T, entonces U se define como congruente a V con módulo T. El significado de la ecuación anterior puede explicarse de la siguiente manera. Si la expresión $\frac{U - V}{T}$ es un entero, entonces U es congruente a V con módulo T. En general hay muchos valores de T que satisfacen la relación anterior. En la mayoría de los casos uno está interesado en el valor más grande T que satisface esta definición. Dados los valores de V y T esto se determina fácilmente a través de

$$U = V - \left(\frac{V}{T}\right) T \qquad [2]$$

donde: (V/T) es un valor truncado. Ejemplo, suponga:

$$U = 12 \pmod 5 =$$

$$U = 12 - \left(\frac{12}{5}\right)(5) = 2$$

La gran mayoría de los generadores de números aleatorios en uso hoy en día son los generados lineales basados en métodos de congruencia definidos por la relación:

$$Z_i = f(Z_{i-1}, Z_{i-2} \ldots) \pmod m \qquad [3]$$

donde f es una función determinística de los Z_i anteriores. Los Z_i están comprendidos entre 0 y m-1 y los números aleatorios U(0,1) están definidos por

$$U_i = Z_i / m$$

A continuación mencionaremos algunos casos particulares. Para una discusión más detallada ver [Knuth 5].

A. GENERADORES CONGRUENTES LINEALES

Definidos por la relación:

$$Z_i = (a\, Z_{i-1} + C) \pmod m \qquad [4]$$

donde **a** es un multiplicador, C el incremento y Z_0 la semilla o valor inicial son valores no negativos y deben satisfacer

$$0 < m,\ c < m,\ Z_o < m \text{ y } a < m$$

La expresión [4] nos dice que la suma $(a\, Z_{i-1} + C)$ se divide por m y Z_i lo hacemos igual al residuo. Ejemplo:

$$Z_i = (7Z_{i-1} + 13) \pmod{14} \quad \text{con } Z_0 = 3$$

la tabla 1, da

El comportamiento cíclico de los generadores congruenciales lineales es inevitable. De acuerdo con la expresión [4], siempre que Z_i asume un valor previo, exactamente la misma secuencia se genera. La longitud de un ciclo se denomina el período (p) de un generador. Si p = m, el generador es de período completo.

En orden a tener período completo es necesario seleccionar las constantes **a**, **m** y **c** de acuerdo con lo previsto por Hull y Dobell [4].

1. **c** debe ser impar y primo en relación con m.
2. **a** = 1 (mod p) si p es un factor primo de m.
3. **a**= 1 (mod 4) si 4 es un factor de m.

En general, podemos utilizar las siguientes reglas:

1. **c** se puede usar cualquier constante. Sin embargo, para garantizar buenos resultados, seleccione **c** tal que **c** mod 8 = 5 (para un computador binario) o tal que **c** mod 200 = 21 (para un computador decimal).
2. **a** debe ser un entero impar, no divisible por 3 ó 5.

   Los valores más convenientes para **a** son de la forma:

   $a = 2^s + 1 \quad s \geq 2$ computador binario
   $a = 10^s + 1 \quad s > 1$ computador decimal

3. **m** debe ser lo suficientemente grande, usualmente por conveniencia se define de acuerdo con el tamaño de la palabra del computador.

   $m = 2^b$ computador binario
   $m = 10^d$ computador decimal

   donde **b** es el número de bits y **d** el número de dígitos decimales en una palabra del computador que se utilice.

   La expresión [4] con **c** # 0 se denomina el método congruencial lineal mixto. Con **c** = 0 se conoce como el método congruencial multiplicativo.

   Según Hull y Dobell [4] los mejores resultados para este caso en un computador binario se logran cuando:

   a. $a = 8t \pm 3$
   2. t es un entero positivo
   3. Z_o es impar
   4. $m = 2^b$, donde $b > 2$

   Aquí el período es la cuarta parte de los enteros o **a** m-1

En lugar de hacer **m** = 2_b, recientemente se ha propuesto que **m** sea el número primo más grande menor de 2^b. Si este valor se elige y **a** es un elemento primitivo de módulo **m** [5] entonces el período **p** = m - 1

Tabla 1.

i	Z_i	U_i	i	Z_i	U_i	i	Z_i	U_i
1	4	0.266	6	6	0.400	11	5	0.333
2	11	0.733	7	10	0.666	12	3	0.200
3	0	0.000	8	8	0.533	13	4	0.266
4	13	0.866	9	9	0.600	14	11	0.733
5	14	0.933	10	1	0.066	15	0	0.000

Para un computador decimal elegimos:

1. **a** = 200 t ± p
2. **t** es un entero positivo
3. Zo es impar no divisible por 5
4. **m** = 10^d

   p es cualquiera de los valores 3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61, 67, 69, 77, 83 y 91.

   Si **d** = 10, una buena selección de **a** será la dada por **a** = 100.000 ± 3.

A continuación presentamos una forma general de subrutina. FORTRAN para ilustrar el método congruencial multiplicativo.

```
    SUBROUTINE RANDU (IX, IY, RN)
    IY = IX*a
    IF(IY) 5, 6, 6
5.  IY = IY + m
6.  RN = IY
    RN = RN*m-1
    IX = IY
    RETURN
    END
```

Note que el analista debe sustituir los valores de **a, m** y **m**$^{-1}$ en el momento de elaborar la subrutina.

Para un computador de 32 bits (b = 31) podemos usar:

```
    a = 65539
5   IY = IY + 2147483647 + 1
    RN = RN* 04656613 E - 9
```

B. GENERADORES COMPUESTOS

Están basados en la combinación de otros generadores separados con la esperanza que el generador final presente mejor comportamiento estadístico. Uno de los más conocidos es el desarrollado por McLaren y Marsaglia [7] y ampliado por Marsaglia y Bray [8]. Inicialmente un vector **V** = (V_1 V_2... Vk) es llenado (secuencialmente) con los primeros **k** U_i obtenidos del primer generador congruencial lineal (k = 128). Luego el segundo generador es utilizado para generar un entero aleatorio **I** distribuido uniformemente en los enteros 1, 2... **k** y **V**I es entregado como el primer número aleatorio de la serie; el primer generador entonces reemplaza la **Ith** posición en **V** con su siguiente U_i y el segundo generador selecciona la siguiente posición del **V** actualizado etc.

C. GENERADORES DE TAUSWORTHE

Están basados en un artículo producido por Tausworthe [13]. Estos métodos están relacionados con sistemas criptográficos y operan directamente con bits para formar números aleatorios.

El método opera así:

1. Define una secuencia b_1 b_2... de dígitos binarios
2. Haga

   $b_i = (C_1 b_{i-1} + c_2 b_{i-2} \ldots + C_R b_{i-R}) \pmod 2$

   donde C_1 C_2... C_R son cero o uno. Por lo general solamente dos son diferentes de cero.
3. Agrupe **k** bits para formar un entero binario de longitud **k** con valor entre 0 y 2 , el cual es dividido por 2^k, para dar el número U(0,1).

   Estos generadores, el período máximo es 2^{R-1}

   Para inicializar el algoritmo es necesario especificar los primeros **R** dígitos binarios en alguna forma, lo cual es equivalente a especificar Z_0.

COMPROBACION DE UN GENERADOR DE NUMEROS ALEATORIOS

Dado que un algoritmo genera una secuencia de números pseudoaleatorios, debemos preguntarnos ¿qué tan cerca están ellos de ser aleatorios en su comportamiento?

Para responder a este interrogante un gran número de pruebas estadísticas han sido propuestas. Entre las pruebas más comúnmente utilizadas están las siguientes:

A. Pruebas sobre la uniformidad de la distribución.

   Para ello se pueden aplicar dos pruebas básicas:

   La prueba de Chi-cuadrado y la prueba de Kolmogorov-Smirnov. Ambas pruebas se interesan por el grado de acuerdo con que existen entre la distribución de una muestra de números aleatorios generados y la distribución uniforme. Además, las dos pruebas están basadas en la hipótesis nula.

   Ho: Los U_i son variables aleatorias independientes e idénticamente distribuidas con función de distribución U(0,1).

Para evaluar la prueba de Chi-cuadrado con todos los parámetros conocidos procedemos así:

1. Dividimos el intervalo [0,1] en **K** subintervalos adyacentes y generamos U_1 U_2... U_n (como regla general **K** debe ser por lo menos 100 y n/K por lo menos 5).
2. Evaluamos

   Nj = número de **Ui** en el **Jth** intervalo

   P_j = 1/k proporción esperada en el **Jth** intervalo.
3. Finalmente, la prueba estadística es

$$X^2 = \sum_{j=1}^{K} \frac{(N_J - nP_j)^2}{nP_j}$$

   Si H_0 fuera verdad, esperamos que sea pequeño. Por tanto, rechazamos H_0 si x^2 es demasiado grande.

   Para **n** grande la última expresión presenta una distribución Chi-cuadrado con **k-1** grados de libertad. Así que, rechazamos la hipótesis nula a un nivel α si

$$x^2 > x^2_{k-1,\,1-\alpha}$$

B. PRUEBAS DE CORRIDAS

Sirven para constatar el número de corridas mayores o menores de alguna constante (usualmente el valor medio) o crecientes y decrecientes.

La prueba implica el conteo del número actual de ocurrencias de corridas de diferente longitud y comparar estos valores a los esperados usando Chi-cuadrado.

C. PRUEBAS DE AUTOCORRELACION

Estas examinan la tendencia de los números de ir seguidos por otros números.

D. PRUEBAS DE HUECOS

Cuenta el número de dígitos que aparecen entre repeticiones de un dígito particular. Si el dígito **x** va seguido de **k** dígitos distintos, antes de que vuelva a aparecer **x**, se dice que existe un hueco de tamaño **k**.

E. PRUEBA DE POKER

Es análoga a las manos de poker. Se utiliza para analizar la frecuencia con la que se repiten los dígitos en números aleatorios individuales. Por ejemplo, si nos ocupamos en números aleatorios de cinco dígitos, nos interesará examinar la frecuencia con que ocurre lo que sigue en los números individuales.

- Los cinco dígitos son diferentes
- Hay exactamente un par (dígitos repetidos)
- Dos pares diferentes
- Tres dígitos iguales
- Tres dígitos iguales, más un par
- Cuatro dígitos iguales
- Cinco dígitos iguales

F. PRUEBAS DE SERIES

Se emplean para comprobar el grado de aleatoriedad entre los números sucesivos en una serie se utiliza la estadística Chi-cuadrado.

G. PRUEBAS DE ESTADISTICAS DE ORDEN

Verifica el valor máximo o mínimo de **n** números consecutivos o el rango de **n** valores consecutivos.

Para discusión detallada de cada una de estas pruebas consultar la bibliografía.

Las pruebas anteriores son las denominadas pruebas empíricas, las cuales son pruebas estadísticas basadas en los **Ui** producidas por un generador. Sin embargo, estas pruebas presentan la desventaja de ser locales, o sea, las pruebas examinan únicamente el segmento del ciclo que se utilizó en producir los números aleatorios. A pesar de esta desventaja, la naturaleza local de estas pruebas permite examinar los números aleatorios utilizables en una simulación.

En contraste, las pruebas teóricas no son pruebas en el sentido estadístico sino que utilizan los parámetros numéricos de un generador para garantizar de una manera global su desempeño sin disponer de los números aleatorios.

Estas pruebas son matemáticamente complejas y refinadas. Dos de estas pruebas son la espectral y la cúbica.

GENERACION DE VARIABLES ALEATORIAS

Finalmente, uno de los papeles más importantes de los números aleatorios U(0,1) es el hecho que constituyen el ingrediente básico de cada método de variables aleatorias a partir de cualquier distribución. De hecho, existen muchas técnicas para generar variables aleatorias y el algoritmo particular utilizado depende, a propósito, de la distribución de la cual deseamos generar; sin embargo, aproximadamente todas las técnicas se pueden clasificar de acuerdo con su base teórica en: Método de Transformada Inversa, Composición, Rechazo-Aceptación y el Uso de Propiedades Especiales [6].

A continuación presentamos un ejemplo usando el Método de Transformada Inversa: Generar valores de una variable aleatoria **X** con distribución exponencial de parámetro β .

La función de distribución es:

$$F_{(x)} = \begin{cases} 1 - e^{-x/\beta} & \text{Si } x \geq 0 \\ 0 & \text{de lo contrario} \end{cases}$$

Si hacemos U = F(x), donde $U \sim U(0,1)$ tenemos:

$$X = F^{-1}(u) = -\beta \ln (1 - u)$$

Luego, para generar la variable aleatoria requerida:

1. Generamos $U \sim U(0,1)$
2. Hacemos $X = F^{-1}(u)$ y regresamos.

CONCLUSION

Sin lugar a duda, la capacidad de diseñar algoritmos que producen números aleatorios con características tales que pueden representar adecuadamente verdaderos números aleatorios, es lo que permite resolver problemas complejos con el uso de la simulación. Sin embargo, dada la aproximación, es realmente imposible modelar con precisión las características de la distribución uniforme mediante un buen generador.

Esto quiere decir que algunas propiedades no se cumplirán, las cuales pueden no influir mucho en los resultados de determinado estudio, dando lugar a que el criterio de aceptación de un generador dado debe basarse en la aplicación que se le vaya a dar. Es precisamente responsabilidad del analista realizar las pruebas pertinentes.

BIBLIOGRAFIA

Emshoff, James R., y Roger L. Sisson: **Design and use of Computer Simulation Models.** Mac Millan Publishing Co. New York, 1970.

Fishman, G. S.: **Concepts and Methods in Discrete Event Digital Simulation,** Wiley, New York, 1973.

Fishman, G. S. **Principles of Discrete Event Simulation,** Wiley, New York, 1978.

Hull, T.E., y A. R. dobell: **Random Number Generator,** SIAM Rey, 4:230-254, 1962.

Knuth, D. E., **The Art of Computer Programming,** Vol. 2, Addison-Wesley, Reading, Mass., 1969.

Law, Averill M. y W. David Kelton: **Simulation Modeling and Analysis.** Mc Graw-Hill Book Company, 1982.

Mac Laren, M. D. y G. Marsaglia: **Uniform Random Number Generators,** J. Ass. Comput. Mach., 12; 83-89 (1965).

Marsaglia, G., y T. A. Bray: **One-Line Random Number Generators and their use in Combinations.** Commun. Ass. Comput. Mach 11:757-759 (1968).

Naylor, Thomas H., Balintfy, Joseph L., Burdick, Donald S., y Chu, Kong: **Computer Simulation Techniques.** New York: John Wiley & Sons 1966.

Naylor, Thomas H.: **Experimentos de Simulación en Computadores con Modelos de Sistemas Económicos.** Editorial Limusa, México, 1977.

Schmidt, J. W. y R. E. Taylor: **Análisis y Simulación de Sistemas Industriales.** Editorial Trillas, México, 1979.

Shannon, Robert E.: Systems Simulation: The Art and Science. Prentice Hall, Inc. Englewood Cliffs, New Jersey, 1975.

Tausworthe R. C.: **Random Numbers Generated by Linear Recurrence Modulo Two,** Math. Comput., 19: 201-209 (1965).

Tocker, K. D.: **The Application of Automatic Computers to Sampling Experiments,** Journal of the Royal Statistical Society, B16 (1954), 39-61.