Generación de Números Aleatorios

Números "*elegidos al azar*" son útiles en diversas aplicaciones, entre las cuáles podemos mencionar:

- **Simulación o métodos de Monte Carlo**: se simula un proceso natural en forma computacional. Estas aplicaciones se realizan en muy variados campos con el fin de emular distintos comportamientos: física (por ejemplo, para simular colisiones entre partículas), ingeniería (diseño de obras hidráulicas, puentes, etc.), inversiones de capital, redes, servicios a clientes, call centers, etc. La simulación a través de la computadora es una herramienta poderosa para comprender la naturaleza de sistemas complejos.

- **Muestreo**: con el fin de seleccionar una submuestra de una población.

- **Análisis Numérico**: algunas técnicas para resolver problemas de análisis numérico complejos han sido desarrolladas usando números aleatorios.

- **Programación**: la generación de valores aleatorios puede ser útil para poner a prueba la efectividad de un algoritmo. También son útiles en criptología.

A pesar de que fue en la década del 40 que las primeras computadoras modernas fueron desarrolladas, la simulación ya existía en forma embrionaria aún antes de que la computadora apareciera en escena. Así, por ejemplo, en la segunda mitad del siglo XIX, se realizaban experiencias arrojando agujas al azar sobre una superficie reglada con el fin de estimar el número π. En 1908 W. S. Gosset, bajo el seudónimo de Student, realizaba un muestreo experimental con el fin de descubrir la distribución de un estimador de la correlación en una distribución normal bivariada. En ese momento los números aleatorios se generaban mediante métodos observacionales (mecanismos físicos) tales como tirar un dado, extraer una carta de un mazo o mediante una ruleta.

Dado el esfuerzo que significaba generar números aleatorios cada vez que eran necesarios, parece razonable que se hayan construido tales números y luego tabulado. Tippett (1927) publicó una tabla con 41600 números aleatorios "tomados en forma aleatoria de informes censales". Cada número era uno de los enteros 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 y el usuario tomaba varios de ellos y agregaba un punto decimal para formar un número aleatorio entre 0 y 1. Desde ese momento una serie generadores de números aleatorios fueron propuestos. La primera máquina fue usada en 1939 por Kendall y Babington-Smith con el fin de producir una tabla de 100000 de dígitos aleatorios y en 1955 la RAND Corporation utilizó extensamente una tabla de 1000000 dígitos aleatorios que fue obtenida a partir de una ruleta electrónica especialmente diseñada. ERNIE fue una famosa máquina de números aleatorios que fue usada por la lotería británica, es decir la *British Premium Savings Bonds Lottery*.

Poco después de la aparición de las computadoras, se comenzó a buscar maneras eficientes de obtener números aleatorios, pues aún cuando se podían usar las tablas existentes éste era un recurso limitado, ya sea por el espacio de memoria necesario como

por resultar, en algunos casos, cortas. Si bien máquinas como ERNIE podrían haber trabajado junto con una computadora, una solución en la que la computadora provee todo parecía más satisfactoria. La búsqueda se orientó, entonces, a la producción de números aleatorios usando operaciones aritméticas de una computadora. John von Neumann sugirió en un principio, alrededor de 1946, usar el método del "cuadrado medio". Su idea era calcular el cuadrado del número aleatorio anterior y tomar los dígitos del medio del número calculado. Así, por ejemplo, si queremos generar un número aleatorio de 10 dígitos y el número anterior es

5772156649 ⟹ 3331779238059490 9201

el nuevo número será 7923805949.

La primera pregunta que cabe hacer es porqué motivo un número generado por este procedimiento que es determinístico, va a resultar aleatorio. La respuesta es que el número no es aleatorio, pero parece serlo, en el sentido en que en una aplicación la relación real entre un número y el siguiente no tiene ningún significado físico. Por lo tanto, el carácter no aleatorio no es una característica indeseable y podría ser que el "cuadrado medio" resultase ser un buen "*batido*" del número anterior. Es claro, de todas formas, que un mecanismo de esta naturaleza no podría haber reemplazado a ERNIE.

Las secuencias de números generadas en forma determinística reciben el nombre de secuencias pseudo-aleatorias o quasi-aleatorias, si bien nosotros nos referiremos a ellas como secuencias aleatorias, sobreentendiendo que sólo "parecen" aleatorias. Números aleatorios generados en forma determinística en una computadora funcionan muy bien en muchísimas aplicaciones, a condición de que el método de generación sea bueno.

Volviendo a la propuesta de von Neumann, ésta no parece ser una buena fuente de números aleatorios. Podría suceder que la secuencia caiga en un ciclo corto de repeticiones, siendo el caso extremo el del cero el cual, si aparece en la secuencia, seguirá repitiéndose siempre. A partir de los años 50 se realizaron diversas experiencias con el método propuesto por von Neumann. Trabajando con números de 4 dígitos en lugar de 10, G. E. Forsythe probó con 16 números iniciales. Con 12 de ellos terminó con el ciclo 6100, 2100, 4100, 8100, 6100, etc. Y con otras dos terminó en cero. En efecto,

```
6100**2 = 37210000
2100**2 =  4410000
4100**2 = 16810000
8100**2 = 65610000
```

Metrópolis realizó muchas pruebas con los números del "*middle-square*" , en especial con sistemas de números binarios. Mostró que en secuencias de 20 dígitos, hay 13 ciclos diferentes en los que la secuencia puede caer, el más largo de los cuales tiene longitud

142. Estas falencias del "*middle-square*" son algunas de las consideraciones que debemos hacer ante un generador de números aleatorios.

En principio consideraremos métodos para generar números con distribución uniforme en el intervalo (0,1). Ésto podemos lograrlo generando enteros X_n entre 0 y un número natural m y luego tomando la fracción:

$$U_n = \frac{X_n}{m}$$

Usualmente m es un número muy grande. El más popular de los generadores de números aleatorios es el *Método Lineal de Congruencias*, que es un caso especial del método introducido por Lehmer en 1949.

Dados cuatro números m, a, c y X_0, formamos la secuencia de números aleatorios X_n de la siguiente forma

$$X_{n+1} \equiv (aX_n + c) \mod m, \qquad n \geq 0$$

es decir que X_{n+1} es el resto entero de dividir $aX_n + c$ por *m* (y por lo tanto es un entero entre 0 y *m*-1). Esta es una secuencia lineal congruente. Tengamos en cuenta que

m es el módulo m>0
a es el multiplicador $0 \leq$ a <m
c es el incremento $0 \leq$ c <m
X_0 es la semilla o valor inicial

En el caso en que c = 0, el método recibe el nombre de multiplicativo secuencial.

Por ejemplo, si m =10 y X_0 = a = c = 7, entonces la secuencia obtenida es

7, 6, 9, 0, 7, 6, 9, 0.....

En cambio, si m = 8, para la misma elección del resto de las constantes, la secuencia sería:

0, 7, 0, 7....

Ésto muestra que la elección de los números m, a y c es crucial y que siempre se caerá en un loop, es decir en un ciclo de repeticiones, que se llama *período*. Es claro que cuanto más grande sea m, mayor es la posibilidad de que el período sea largo.

En realidad, las distintas elecciones de los parámetros son sometidas a una batería de tests con los que se chequean las propiedades de los números generados.

Como ya observamos más arriba, con estos algoritmos se generan números aleatorios que se comportan como si proviniesen de una distribución U(0,1). La pregunta que es razonable hacerse es "porqué ésto es suficiente". El siguiente teorema nos da una respuesta.

Teorema: Sean U una variable aleatoria con distribución $U(0,1)$ y G una función de distribución acumulada continua y estrictamente creciente. Si $X = G^{-1}(U)$, entonces la función de distribución acumulada de X es G, es decir $F_X = G.$

Dem:

Recordemos que si $U \sim U(0,1)$, entonces su función de distribución es de la forma

$$F_U(u) = \begin{cases} 0 & \text{si } u \le 0 \\ \text{u} & \text{si } 0 < u < 1 \\ 1 & \text{si } u \ge 1 \end{cases}$$

Por lo tanto, como G es una función estrictamente creciente y su imagen pertenece al intervalo (0,1), entonces

$$F_X(x) = P(X \le x) = P(G^{-1}(U) \le x) = P(U \le G(x)) = F_U(G(x)) = G(x)$$

con lo que queda demostrado el teorema.

Ejemplo: En el caso de una variable $X \sim E(\lambda)$, la función de distribución acumulada es de la forma

$$F_X(x) = \begin{cases} 0 & \text{si } x \le 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0 \end{cases}$$

Dado $y \in (0,1)$, la inversa de F_X es

$$F_X{}^{-1}(y) = -\frac{1}{\lambda}\ln(1-y)$$

Luego, si $U \sim U(0,1)$,

$$-\frac{1}{\lambda}\ln(1-U) \sim E(\lambda)$$

Si la distribución G tiene saltos o es constante de a trozos, no existirá su inversa. Sin embargo se puede demostrar que existe una H con las propiedades requeridas en el teorema anterior, de manera que, aunque sin demostración, enunciaremos el siguiente resultado.

Teorema: Sean U una variable aleatoria con distribución $U(0,1)$ y G una función de distribución acumulada. Existe una función H tal que $H(U)$ tiene distribución acumulada G.

Ejemplos: Queremos generar una variable con distribución de Bernoulli de parámetro *p* a partir de una v.a. uniforme. Podemos aplicar el siguiente procedimiento. Generamos $U \sim U(0,1)$ y definimos:

$$X = \begin{cases} 1 & \text{si } 0 < U \leq p \\ \\ 0 & \text{si } p < U \leq 1 \end{cases}$$

En efecto, la nueva variable X toma sólo dos valores (0 y 1) y dado que $p \in (0,1)$

$$P(X = x) = P(U \leq p) = p$$

y por lo tanto X tiene la distribución deseada.

Notemos que en lugar del intervalo $(0, p)$ podríamos haber tomado cualquier intervalo en $(0,1)$ de longitud p.