GENERACION DE NUMEROS ALEATORIOS

En cualquier experimento de simulación, así como en la mayoría de los experimentos de muestreo, existe la necesidad de contar con una fuente de números aleatorios. Los números aleatorios que estudiaremos en este capítulo se refieren a números o valores de variables que siguen la distribución uniforme. La finalidad del presente capítulo es analizar las propiedades convenientes de los números aleatorios utilizados en experimentos de simulación, a fin de presentar algunos métodos que se pueden utilizar para generar esos números aleatorios y, finalmente, presentar varias pruebas estadísticas que se pueden aplicar a un determinado conjunto de esos números aleatorios para determinar cuáles son las propiedades que poseen.

Existen diferentes métodos utilizados para obtener números aleatorios, como ser:

a) **Métodos manuales**. Se pueden obtener números aleatorios extrayendo (con reposición) bolillas numeradas consecutivamente desde un bolillero, arrojando dados, sacando con reposición naipes desde una baraja, etc.
b) **Tablas de números aleatorios.** Varios libros de texto de estadística contienen grandes tablas de números aleatorios que se pueden utilizar conforme se haga necesario. Seguidamente se presenta una de ellas:

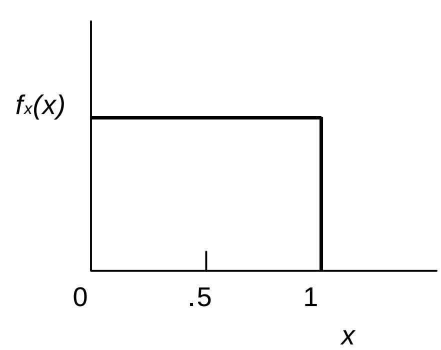
51772	74640	42331	29044	46621	62898	93582	04186	19640	87056
24033	23491	83587	06568	21960	21387	76105	10863	97453	90581
45939	60173	52078	25424	11645	55870	56974	37428	93507	94271
30586	02133	75797	45406	31041	86707	12973	17169	88116	42187
03585	79353	81938	82322	96799	85659	36081	50884	14070	74950
64937	03355	95863	20790	65304	55189	00745	65253	11822	15804
15630	64759	51135	98527	62586	41889	25439	88036	24034	67283
09448	56301	57683	30277	94623	85418	68829	06652	41982	49159
21631	91157	77331	60710	52290	16835	48653	71590	16159	14676
91097	17480	29414	06829	87843	28195	27279	47152	35683	47280

c) **Métodos de computadora analógica**. Consiste en la generación de números aleatorios mediante el uso de computadoras analógicas diseñadas específicamente para ello. Esta técnica para la creación de variables aleatorias está actualmente en desuso.
d) **Métodos de computadora digital**. Son los generados mediante relaciones matemáticas que pueden ser manifestadas o representadas por medio de una serie de proposiciones en lenguaje de computadoras digitales, las cuales cuando son ejecutadas, causen que un número sea creado a partir de la distribución uniforme de probabilidad.

PROPIEDADES DE LOS NUMEROS ALEATORIOS DISTRIBUIDOS UNIFORMEMENTE

Puesto que este curso se dedica a experimentos de simulación en computadoras digitales, consideraremos que los números aleatorios nos los proporciona algún generador de números aleatorios por medio de un programa o subrutina de computadora. El generador de números aleatorios que proporciona esos números está diseñado para producir variables que siguen la distribución uniforme. La pregunta a la que debemos responder es: “¿qué propiedades debe poseer esa serie de variables aleatorias uniformemente distribuidas?”.

En primer lugar, examinaremos la distribución uniforme de probabilidad y definiremos las propiedades que posee. En la siguiente figura se muestra la función densidad para la distribución uniforme definida sobre el intervalo (0,1).



Para los fines de este análisis, a menos que se indique otra cosa, suponemos que nos ocupamos de la distribución uniforme continua definida sobre el intervalo (0,1).

La media de esta distribución debe ser ½. Además, para cualquier muestra dada, el segundo momento o la suma promedio de los cuadrados debe ser 1/3 y el tercer momento o la suma promedio de los cubos debe ser ¼.

Si tuviéramos que dividir el intervalo (0,1) entre *n* clases o subintervalos, el *i*ésimo subintervalo debería contener *T/n* observaciones, siendo T el número total de observaciones hechas. O sea, si hubiéramos hecho 1000 observaciones de las variables a partir de la distribución uniforme, poniéndolas en una distribución de frecuencias que contuviera, por ejemplo, 10 intervalos espaciados por igual, podríamos esperar tener una frecuencia de 1000/10 = 100 observaciones en cada intervalo. Esta propiedad debe ser válida, sea cual sea el tamaño de los intervalos. Es decir, si tuviéramos que reducir el intervalo en el caso discreto a un valor simple, podríamos esperar tener exactamente el mismo número de observaciones registradas para cada valor en el rango.

Esta propiedad se puede extender todavía más, diciendo que la probabilidad de observar un valor en un intervalo dado permanece constante y es independiente del valor obtenido anteriormente. Veamos nuevamente esta propiedad dada en función de la distribución discreta. Indica que, en el caso discreto, la probabilidad de que salga cualquier número dado es exactamente la misma para cada número diferente e independiente del número registrado anteriormente. O sea:

$$P(x_i = X) = \text{constante} \qquad \text{para todos los } i$$

Es preciso recordar que nos ocupamos de una serie de números que se generan sucesivamente en el tiempo. Teóricamente, debemos esperar eventos independientes y para cualquier intervalo dado, una probabilidad constante de que un valor dado caerá en ese intervalo. Puede resultar que nuestro generador particular no produzca de hecho esos números independientes y tenga una variación cíclica, por ejemplo se generen varios números por debajo de la media, y así sucesivamente. Las pruebas normales de frecuencia pueden indicar un buen ajuste; sin embargo, al tener en cuenta el orden en que aparecen los números, podremos ver que el generador no era verdaderamente aleatorio y que, por ende, no creamos eventos verdaderamente independiente.

Puesto que esas propiedades existen de manera ideal en cualquier nivel de significancia, deberemos poder repartir los números y comprobar la aleatoriedad de cada dígito o grupo de dígitos. Si se desean números aleatorios con tres posiciones decimales, será útil probar el generador hasta cinco posiciones decimales. De todos modos, las series de números que aparecen deben comprobarse en cada dígito si la aplicación a que se destina así lo requiere. En cualquier dígito dado del número aleatorio, la probabilidad de que aparezca los dígitos cero a nueve es de 0.1. Por ejemplo, la probabilidad de observar el número tres en el segundo lugar decimal de nuestro número aleatorio es 0.1 para cada número generado. Esta propiedad, como ya se indicó, debe comprobarse en todos los dígitos de cada número.

GENERADORES DE NUMEROS ALEATORIOS

Antes de analizar métodos para diseñar generadores de números aleatorios o dar ejemplos de ellos, examinaremos las propiedades convenientes que deben poseer un generador de números aleatorios. Para nuestros fines, debemos tener en consideración esas propiedades en función de un programa de computadora. Sencillamente, preguntémonos: "¿cuáles son las propiedades convenientes de este programa de computadora para generar variables aleatorias uniformemente distribuidas?".

Evidentemente, la primera respuesta que ofreceremos para esa pregunta es que la serie de números que produce ese generador debe seguir la distribución uniforme ideal tan de cerca como sea posible.

Supongamos que el generador de números aleatorios produce números aceptables para nuestra aplicación. ¿Qué otras propiedades podemos esperar razonables que tenga este programa?. Las otras propiedades convenientes se dan a continuación:

1. **Debe ser rápido**; o sea, que debe generar un número en el menor tiempo posible. Este generador de números aleatorios es tan sólo una faceta de un modelo de simulación en computadora y, por ende, no debe consumir mucho tiempo.
2. **El programa mismo debe ser breve**. Esto quiere decir que no se requiera gran espacio de almacenamiento del mismo.
3. **Debe tener 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. El generador no se reciclará necesariamente mediante un regreso al primer número generado. Por ejemplo, el generador puede reciclarse y comenzar a reproducir la serie que se inicia en el décimo número generado.
4. Puesto que podemos desear duplicar el experimento varias veces, el generador debe poder reproducir las mismas series de números que se desee. Por otra parte, debe tener capacidad para producir, a voluntad, un conjunto o una serie claramente distintos.

5. **El generador debe ser de naturaleza no degenerativa.** La degeneración significa que el generador produce continuamente el mismo número. Si el generador tiene degeneración, el programa debe poder hacer correcciones y seguir adelante.

Así pues, en general, podemos decir que un generador de números aleatorios debe ser un programa breve y rápido que produzca una larga secuencia de números aleatorios (que pasen las pruebas estándar) antes de comenzar a reciclarse y que tenga una naturaleza algorítmica. Como veremos, para cualquier experimento dado puede haber varios programas distintos que satisfagan esos criterios.

Al principio, se mencionó que un generador de números aleatorios era casi siempre un programa o subrutina de computadora. Debe resultar evidente que, si sólo se requieren cinco números aleatorios para un experimento dado de simulación, sería mucho más económico leerlos de archivo, después de tomar los valores de un conjunto de tablas. Además, si un generador dado de números aleatorios llega a generar perfectamente 30 números aleatorios antes de reciclarse (o sea, si tiene un período de 30 números) y sólo necesitamos 20 para nuestro experimento, ese generador será apropiado para nuestros fines.

Antes indicamos que el generador debe ser de naturaleza algorítmica. Se trata de un requisito muy práctico para el programa generador. El que sea algorítmico quiere decir que se utiliza el resultado del cálculo anterior para determinar el siguiente. De hecho, se trata de una cualidad útil, porque le permite al generador funcionar independientemente de todas las demás partes del programa de simulación. En general, lo único que se debe hacer es proporcionar un valor inicial o de partida al generador, que sigue adelante, generando una secuencia de valores a partir de ese inicial.

Para ser precisos, debemos observar que los métodos presentados aquí para la creación de generadores de números aleatorios son determinísticos, puesto que todos ellos incluyen una técnica repetitiva o una relación de congruencia, expresada por medio de una fórmula. Cualquier serie de números creados de este modo no podrá ser “verdaderamente aleatoria”. En teoría, sólo algunos fenómenos físicos pueden ser procesos aleatorios verdaderos. En vista de esto, muchos autores hablan de “generadores de números seudoaleatorios” en lugar de “generadores de números aleatorios”.

METODOS PARA LA GENERACION DE NUMEROS ALEATORIOS

El primer método para la generación de números aleatorios se conoce como técnica de **“cuadrado medio” (midsquare).** Lo propuso Von Neumann y Metropolis hacia 1946. En el método de cuadrado medio cada número sucesivo se genera tomando los *n* dígitos centrales del cuadrado del número anterior de *n* dígitos.

$$r_{m+1} = (r_m)^2 \qquad n \text{ dígitos medios}$$

Por ejemplo, supóngase que deseamos generar números aleatorios de tres dígitos. Nuestro primer valor es 239. Esto nos da la secuencia siguiente de números: 239➔712➔694➔163➔656➔033➔108➔166➔755. Los problemas que se presentan en este método se deben a que tiende a degenerar con rapidez. Dependiendo del valor inicial, el método puede degenerar al cabo de 20 términos.

En este método se producirá eventualmente la repetición de algunas series y la secuencia de números aleatorios se reciclará. A su vez, debido a las operaciones que se deben realizar este método no es muy rápido.

Otro método de generación de números aleatorios se denomina técnica del **producto medio** y es muy similar a la técnica de cuadrado medio en que el número aleatorio resultante se toma de los *n* dígitos centrales del resultado de una multiplicación previa. En notación matemática:

$$r_{m+1} = r_m \times r_{m-1} \qquad n \text{ dígitos medios}$$

La técnica para el producto medio implica la elección de dos números aleatorios r_1 y r_2, cada uno de ellos con *P* dígitos. Luego, se multiplica r_1 por r_2 para obtener *U*. Se hace r_3 igual a los *P* dígitos centrales de *U*. A continuación r_4 es igual a r_3 multiplicado por r_2, y así sucesivamente. Una modificación de este método consiste en utilizar un multiplicador constante, en lugar de números aleatorios; o sea:

$$r_{m+1} = K \times r_m \qquad n \text{ dígitos medios}$$

Este método es similar al de cuadrado medio. No obstante, los dos tienen períodos más largos y los números que producen parece estar distribuidos más uniformemente. Sin embargo, como sucede en el caso de la técnica de cuadrado medio esta técnica parece degenerar finalmente hacia algún valor constante. Tanto el método de cuadrados medios como el de producto medio tienen un período relativamente breve que se ve afectado considerablemente por los valores escogidos inicialmente.

Estos métodos, que se solían utilizar ampliamente, han cedido su lugar a los **métodos de congruencia** basados en la relación:

$$r_{i+1} \equiv ar_i + C(m\acute{o}d\ m), \qquad 0 \le r_i \le m$$

Esta relación implica que la suma $ar_i + C$ se debe dividir por m y r_{i+1} es igual al residuo. La relación indica "*r_{i+1} es congruente con $ar_i + C$ módulo m*". Como ilustración, sea $m = 25$, $a = 6$ y $C = 1$; asimismo, sea $r_0 = 1$.

$$r_1 \equiv 6 \times 1 + 1 (\text{mód } 25) \qquad r_1 \equiv 7$$
$$r_2 \equiv 6 \times 7 + 1 (\text{mód } 25) \qquad r_2 \equiv 18$$
$$r_3 \equiv 6 \times 18 + 1 (\text{mód } 25) \qquad r_3 \equiv 9$$

El método lo propuso inicialmente Lehmer en 1949. Con $C = 0$, se denomina **método multiplicativo de congruencia**. La forma que se muestra con $C \neq 0$ se denomina **método de congruencia mixta**. Otra forma:

$$r_{i+1} \equiv r_i + r_{i-1} (\text{mód } m)$$

se conoce como **método congruente aditivo**. De hecho, se trata de una secuencia de Fibonacci cuando $r_0 = 0$ y $r_1 = 1$.

Se ha demostrado que el método congruente multiplicativo, cuando:

$$r_{i+1} \equiv ar_i (\text{mód } m)$$

se comporta muy bien desde el punto de vista estadístico (Gorenstein, 1967).

Por lo común, en una computadora binaria se considera que m es alguna potencia de 2. En este mismo tipo de computadoras, el periodo máximo es $m/4$, donde $m = 2^b$ $(b > 2)$ y se logra con r_0 impar y $a = 8t \pm 3$, donde $t = 1,2,3$. En este método suele ser conveniente hacer que b tome el valor del número de bits de una palabra binaria de la computadora que se utilice.

Un avance respecto a lo anterior fue realizado por MacLaren y Marsaglia (1965) que propusieron una combinación de dos generadores congruentes para producir secuencias aleatorias. En este método se utiliza un generador para mezclar la secuencia producida por el otro.

Los métodos analizados corresponden a los primeros desarrollos. En la actualidad existen gran cantidad de ellos y día a día aparecen nuevos métodos cada vez más complejos y sofisticados.

PRUEBAS SOBRE LA UNIFORMIDAD DE LA DISTRIBUCION

Uno de los análisis básicos que se deben hacer siempre es el de la validación de la uniformidad de la distribución. Para ello se pueden aplicar dos pruebas básicas: la **prueba de *ji* cuadrada** y la de **Kolmogorov-Smirnov**. Ambas pruebas se interesan por el grado de acuerdo que existe entre la distribución de una muestra de números aleatorios generados y la distribución uniforme teórica. Además, las dos pruebas están basadas en una hipótesis nula de que hay una diferencia no detectable entre una distribución muestral y la teórica. Las dos pruebas se basan en el agrupamiento de datos mustrales en clases, dentro del intervalo (0,1). Parece ser que la prueba de Kolmogorov-Smirnov es más poderosa en casi todos los casos que la prueba *ji* cuadrada.

Ambas son muy fáciles de programar en una computadora y se pueden diseñar como subrutinas para fines de prueba.

Las dos pruebas anteriores nos permiten determinar el grado de uniformidad de nuestra muestra pero no nos dice nada respecto del orden en el que aparecen los números. Para verificar la aleatoriedad del orden de aparición de los números se deben realizar las **pruebas de corridas**. Se considera una corrida a una secuencia de números ascendente o descendente. Mediante ciertos criterios se considera que una secuencia de números aleatorios no es tal si en la misma existen pocas o excesivas corridas.

Las anteriores son las pruebas más comunes sin embargo existen otras entre las que podemos mencionar, la **prueba de huecos** (determina la significancia de los intervalos entre la repetición de cierto dígito), la **prueba del póquer** (analiza la frecuencia con la que se repiten los dígitos en números aleatorios individuales), etc.