TP 2.1 - GENERADORES DE NÚMEROS PSEUDO-ALEATORIOS

Santiago Cancio
Cátedra de Simulación
Universidad Tecnológica Nacional
Zeballos 1341, Rosario, Santa Fe
santiago.cancio96@gmail.com

Nicolás Fierro
Cátedra de Simulación
Universidad Tecnológica Nacional
Zeballos 1341, Rosario, Santa Fe
nicofierro1@gmail.com

Alejandro Gómez Fernández
Cátedra de Simulación
Universidad Tecnológica Nacional
Zeballos 1341, Rosario, Santa Fe
agomfer@gmail.com

24 de abril de 2021

ABSTRACT

Este trabajo práctico se basará en el estudio de diferentes generadores de números pseudo-aleatorios. Primero crearemos los generadores para luego realizar diversas pruebas con los mismos.

1. Introducción

Un número pseudo-aleatorio es un número generado en un proceso que parece producir números al azar, pero no lo hace realmente. Las secuencias de números pseudo-aleatorios no muestran ningún patrón o regularidad aparente desde un punto de vista estadístico, a pesar de haber sido generadas por un algoritmo completamente determinista, en el que las mismas condiciones iniciales producen siempre el mismo resultado.

Los generadores de números pseudoaleatorios son ampliamente utilizados en campos tales como el modelado por computadora, estadística, diseño experimental, etc. Algunas de estas secuencias son lo suficientemente aleatorias para ser útiles en estas aplicaciones.

Por lo general, el interés no radica en generar un solo número aleatorio, sino muchos, reunidos en lo que se conoce como secuencia aleatoria.

Nuestro objetivo es generar dichas sucesiones de números pseudo-aleatorios, y para lograr esto hemos programado dos generadores distintos:

- Generador Congruencial Lineal (GCL)
- Generador Método de los cuadrados medios

2. Descripción de los generadores

2.1. Generador Congruencial Lineal

Un generador congruencial lineal es un algoritmo que permite obtener una secuencia de números pseudo-aleatorios calculados con una función lineal definida a trozos discontinua. Es uno de los métodos más antiguos y conocidos para la generación de números pseudo-aleatorios. La teoría que sustenta el proceso es relativamente fácil de entender, el algoritmo en si es de fácil implementación y su ejecución es rápida.

El generador se define por la relación de recurrencia:

$$X_{n+1} = (aX_n + c) \mod m$$

donde es la secuencia de valores pseudoaleatorios, y X

m, 0 <m - el "módulo"

a, 0 <a <m - el "multiplicador"

c, 0 <= c <m - el ïncremento"

X0, 0 <= $X_0 < m - la"semilla"o"valorinicial"$

son constantes enteras que especifican el generador. Si c = 0, el generador a menudo se denomina generador congruencial multiplicativo (MCG) o Lehmer RNG. Si c <>0, el método se denomina generador congruencial mixto.

Cuando c <>0, un matemático llamaría a la recurrencia una transformación afín, no lineal, pero el nombre inapropiado está bien establecido en la informática.

2.2. Generador Método de los cuadrados medios

Es un método propuesto en los años 40 por los matemáticos John von Neumann y Nicholas Metropolis, siendo utilizado para la generación de números pseudo-aleatorios, Esto para obtener una sucesión de números que básicamente se obtienen a partir de recurrencia, los cuales son relevantes en los procesos de simulación debido a que con estos números se hace posible comprobar el correcto funcionamiento de una prueba mediante la observación del comportamiento de las variables que se puedan encontrar a lo largo de la simulación.

El método consiste en tomar un número al azar, X° de 2n cifras que al ser elevado al cuadrado resulta un número de hasta 4n cifras, de no ser así se deben agregar ceros a la izquierda de dicho resultado para que éste tenga exactamente 4n cifras.

Se denomina X1 al número resultante de seleccionar las 2n cifras centrales del resultado anterior.

Se genera el número pseudo-aleatorio U1 ubicando un punto decimal delante de las 2n cifras de X1 y así sucesivamente para los demás números pseudo-aleatorios.

3. Pruebas

Existen dos tipos de pruebas:

1. Empíricas: evalúan estadísticas de sucesiones de números.
2. Teóricas: se establecen las características de las sucesiones usando métodos de teoría de números con base en la regla de recurrencia que generó la sucesión.

Nosotros realizaremos cuatro pruebas diferentes para determinar la calidad de generación de los números pseudo-aleatorios. Las pruebas serán las siguientes:

- Prueba Monobit
- Prueba de Bondad de Ajuste Chi Cuadrado
- Prueba de Rachas
- Prueba de Poker

3.1. Prueba Monobit

El propósito de esta prueba, es determinar si el número de unos y ceros en una secuencia son aproximadamente la misma que sería de esperar para una secuencia verdaderamente aleatoria. La aparición de un cero o un uno en la secuencia debería ser igualmente probables, de modo que el defecto detectado por este método es que la secuencia contiene demasiados ceros o unos.

3.2. Prueba de Bondad de Ajuste Chi Cuadrado

La Prueba de Bondad de Ajuste Chi Cuadrado es el test de bondad de ajuste más utilizado. En general un test de bondad de ajuste se utiliza para discriminar si una colección de datos o muestra se ajusta a una distribución teórica de una determinada población. En otras palabras, nos dice si la muestra disponible representa razonablemente los datos que uno esperaría encontrar en la población.

La aplicación de la prueba de bondad de ajuste chi cuadrado requiere que los datos estén agrupados en categorías o clases. Si los datos originalmente no se encuentran agrupados será necesario agruparlos antes de aplicar el test de chi cuadrado para lo cual será necesario construir una tabla de frecuencia o histograma.

La fórmula de cálculo del estadístico chi cuadrado utilizado en el test de bondad de ajuste chi cuadrado corresponde a:

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

3.3. Prueba de Rachas

El contraste de rachas permite verificar la hipótesis nula de que la muestra es aleatoria, es decir, si las sucesivas observaciones son independientes. Este contraste se basa en el número de rachas que presenta una muestra. Una racha se define como una secuencia de valores muestrales con una característica común precedida y seguida por valores que no presentan esa característica. Así, se considera una racha la secuencia de k valores consecutivos superiores o iguales a la media muestral siempre que estén precedidos y seguidos por valores inferiores a la media muestral.

El número total de rachas en una muestra proporciona un indicio de si hay o no aleatoriedad en la muestra. Un número reducido de rachas (el caso extremo es 2) es indicio de que las observaciones no se han extraído de forma aleatoria, los elementos de la primera racha proceden de una población con una determinada característica mientras que los de la segunda proceden de otra población. De forma idéntica un número excesivo de rachas puede ser también indicio de no aleatoriedad de la muestra.

Si la muestra es suficientemente grande y la hipótesis de aleatoriedad es cierta, la distribución muestral del número de rachas, R, puede aproximarse mediante una distribución normal de parámetros:

$$\mu_R = \frac{2n_1n_2}{n} + 1 \quad \sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}}$$

donde n1 es el número de elementos de una clase, n2 es el número de elementos de la otra clase y n es el número total de observaciones.

$$Z = \frac{R + c - \mu_R}{\sigma_R}$$

donde c = 0,5 si $R < \mu_R$ y c = -0,5 si $R > \mu_R$.

3.4. Prueba de Poker

Esta prueba examina en forma individual los dígitos del número pseudoaleatorio generado. La forma como esta prueba se realiza es tomando 5 dígitos a al vez y clasificándolos como: Par, dos pares, tercia, full, póker, quintilla y todos diferentes. Las probabilidades para cada una de las manos del póker diferentes se muestran a continuación:

- Todos diferentes = 0.3024
- Un par = 0.504
- Dos pares = 0.108
- Tercia = 0.072
- Full = 0.009
- Poker = 0.00045
- Quintilla = 0.0001

Con las probabilidades anteriores y con el número de números pseudoaleatorios generados, se puede calcular la frecuencia esperada de cada posible resultado, la cual al compararse con la frecuencia observada, produce el estadístico:

$$\chi_0^7 = \sum_{i=1}^{7} \frac{(FO_i - FE_i)^2}{FE_i}$$

Si $\chi_0^2 < \chi_{\alpha,6}^2$. Entonces los números pasan la prueba.

4. Conclusión

Referencias

[1] Números pseudo-aleatorios: https://es.wikipedia.org/wiki/N

[2] Generador Congruencial Lineal: https://es.qaz.wiki/wiki/Linear$_c ongruential_g enerator$

[3] Método de los cuadrados medios: https://es.wikibooks.org/wiki/Método$_d e_l os_c uadrados_m edios_p ara_l a_g eneracin_d e_n meros_p seudoaleat$

[4] Prueba Monobit: http://4imedio.blogspot.com/2012/08/prueba-de-frecuencia-monobit.html

[5] Prueba de Bondad de Ajuste: https://www.probabilidadesyestadistica.com/prueba-de-bondad-de-ajuste-chi-cuadrado/

[6] Prueba de Rachas: http://www.ub.edu/aplica$_i nfor/spss/cap5-4.htm$

[7] Prueba de Poker: https://www.monografias.com/trabajos32/prueba-de-poker/prueba-de-poker.shtml: :text=PRUEBA