TRABAJO PRACTICO "GENERADORES PSEUDOALEATORIOS"

A PREPRINT

Gigena Daiana
Catedra Simulacion
Legajo 39372
UTN-FRRO
daigigena3@gmail.com

Vilchez Joaquin
Catedra Simulacion
Legajo 46483
UTN-FRRO
joaquinvilchez95@gmail.com

May 19, 2020

ABSTRACT

El siguiente informe tiene como objetivo fundamental introducirnos en la simulación de generadores de números pseudoaleatorio.

1 Introducción

Se llama números pseudoaleatorios a aquellos que surgen por medio de una función (determinista, no aleatoria) que aparentan ser aleatoria. Estos se generan a partir de un valor inicial llamado semilla, que se aplica iterativamente a la función. La sucesión de números pseudoaleatorios obtenida es sometida a diversos test para medir hasta que punto se asemeja a una sucesión aleatoria.

2 Generadores pseudoaleatorios

En el siguiente trabajo practico estudiaremos distintos tipos de generadores pseudoaleatorios, entre ellos, Generador lineal congruencial(GLC), la media de los cuadrados, otros. Creados en un entorno de trabajo de lenguaje de programación Python.

2.1 Media de los cuadrados

Este método fue propuesto por el matemático John Von Neumann (1903-1957). En el cual la sucesión se obtiene por recurrencia.

- Se inicia con una semilla de 4 dígitos.
- La semilla se eleva al cuadrado, obteniendo un numero de 8 dígitos (si no es así, se le agregan ceros).
- Los 4 números del centro serán el siguiente numero de la secuencia.

2.2 Generador congruencial lineal

Este generador fue introducido por Lehmer, comienza con un valor inicial (semilla) y los sucesivos valores se obtiene recursivamente del modo:

$$x_{_n} = (ax_{n-1} + b) mod\, m \tag{1}$$

2.3 Generador congruencial multiplicativo

$$x_{_{n+1}} = (ax_n) mod\, m \tag{2}$$

3 Test/Pruebas

Someteremos a distintos tipos de pruebas a los GCL, para verificar la calidad de los números pseudoaleatorios. Los items mas significativos en los números son independencia e uniformidad.

3.1 Prueba de la Bondad

Esta prueba consiste en clasificar nuestros números en diferentes intervalos:

$$[0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1] \tag{3}$$

Creando una tabla de frecuencias en donde ubica a cada numero aleatorio que esta dentro de dichos intervalos:

$$[0, 0.1], [0.1, 0.2], [0.2, 0.3], [0.3, 0.4]... \tag{4}$$

Esta tabla de frecuencia indica nuestras frecuencias obtenidas. Nuestras frecuencias esperadas las podemos obtener a través de esta ecuación:

$$E_i = \frac{N}{n} \tag{5}$$

Con estos valores podemos aplicar la ecuación de chi-cuadrado para poder obtener nuestro valor de significancia(Z)

$$X^2_{calculada} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{6}$$

Gracias a este valor vamos a poder indicar si nuestros números aleatorios se ajustan a una distribución de chi-cuadrada y poder verificar la validez de nuestros números generados.

3.2 Prueba de autocorrelación

En esta prueba se seleccionan todos los números aleatorios generados y se determinan ciertos parámetros para poder hacer dicha prueba, como por ejemplo: El margen de error, el numero por donde se quiere comenzar con la secuencia y la amplitud de la misma.

Para analizar la correlación se utiliza la densidad de probabilidad:

$$\rho_{im} = \frac{1}{M+1} \sum_{k=0}^{m} r_{(i+km)} * r_{|i+(k+1)m)|} \tag{7}$$

- **N** representa el tamaño de la muestra.
- **i** representa el primer numero donde empieza la amplitud de autocorrelación
- **m** representa la amplitud de la correlación
- **M** representa el entero mayor de la correlación
- **M** se obtiene de la siguiente manera, obteniendo un numero con decimales y truncándolo para obtener solamente el numero entero.

$$M = Truncar\left\{\frac{(N-1)}{m}\right\} - 1 \tag{8}$$

Luego tenemos que obtener la desviación estándar de nuestra autocorrelación a través de la siguiente ecuación

$$\sigma_{\rho im} = \frac{\sqrt{13M+7}}{12(M+1))} \tag{9}$$

Por último para obtener la significancia (Z), a través de la siguiente ecuación: Z:

$$Z = \frac{\rho_{im} - 0.25}{\sigma_{\rho im}} \tag{10}$$

3.3 Prueba de Póquer

Examina en forma individual los dígitos del numero aleatorio generado. La forma en la que esta prueba se realiza es tomando 5 dígitos a la vez (si tiene mas dígitos, solo se toman 5) y clasificándolos como:

- Par
- Dos pares diferentes
- Tres dígitos iguales
- Full
- Cuatro dígitos iguales
- Cinco dígitos iguales
- Todos diferentes

Al categorizar nuestros números aleatorios de esa manera, se genera una tabla de frecuencia de cada una de esas categorías con relación a los números aleatorios obtenidos y podemos obtener así, las frecuencias observadas, mientras que las frecuencias esperadas son la probabilidad de ocurrencia de cada categoría en particular.

Luego de obtener esos datos podemos aplicarlos a una distribución de chi-cuardado para poder obtener nuestro valor de significancia

$$X^2 = \sum_{i=1}^{7} \frac{(O_i - E_i)^2}{E_i} \tag{11}$$

Luego, con este valor de Z podemos determinar si los números se ajustan a una distribución de chi-cuadrado o no y determinar la validez de nuestros números aleatorios.

3.4 Prueba de corrida arriba abajo

Consiste en comparar el numero sucesivo con el anterior, para ver si es mayor (se le asigna 1) o menor (se le asigna 0). Se arma una colección con esos ceros y unos y se obtiene la cantidad de corridas.

- **a =** numero de corridas
- Media:

$$\mu_a = \frac{2N-1}{3} \tag{12}$$

- Varianza:

$$\sigma_a^2 = \frac{16N - 29}{90} \tag{13}$$

- Z:

$$Z = \frac{a - \mu_a}{\sigma_a} \tag{14}$$

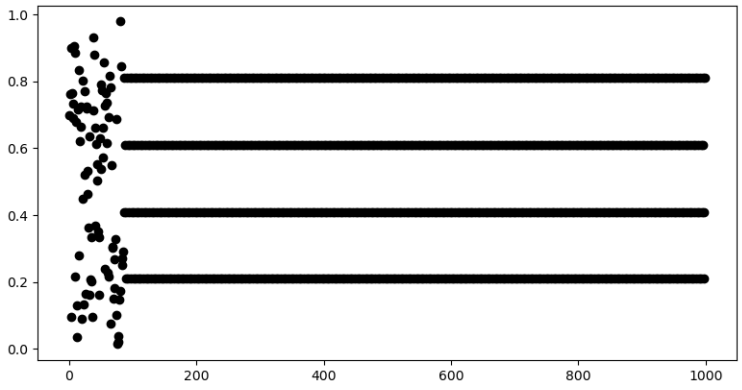
4 Resultados

Luego de someter al GCL a las distintas pruebas estos fueron los resultados:

4.1 Media de cuadrados

Elegimos una semilla de cuatro dígitos y la cantidad de tiradas. En nuestro caso elegiremos la semilla 9268 y 1000 tiradas.

Los resultados son los siguientes:

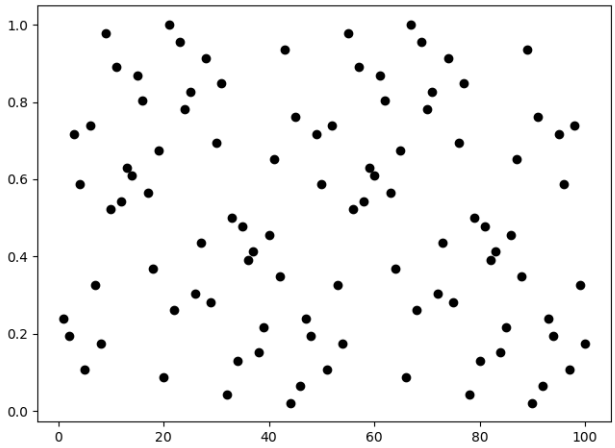


Como se puede observar en el gráfico, el método de la media de cuadrados muestra números aleatorios en la primer etapa (aproximadamente 100 tiradas) de manera correcta, hasta que llega un momento en que se vuelve completamente repetitivo y no muestra mas números aleatorios.

4.2 Método multiplicativo (GCL)

Elegimos una semilla de 2 dígitos, una constante multiplicativa, un modulo y una cierta cantidad de tiradas. En nuestro caso elegimos la semilla 97, constante multiplicativa 35, modulo 47 y 100 tiradas.

Los resultados son los siguientes:



4.2.1 Prueba de bondad

Para poder testear nuestros números aleatorios, los sometemos a pruebas como por ejemplo la de bondad explicada anteriormente.

Los resultados son los siguientes:

Intervalos	**Oi**	**Ei**
[0.0, 0.1]	8	10
[0.1, 0.2]	13	10
[0.2, 0.3]	9	10
[0.3, 0.4]	11	10
[0.4, 0.5]	10	10
[0.5, 0.6]	9	10
[0.6, 0.7]	10	10
[0.7, 0.8]	10	10
[0.8, 0.9]	10	10
[0.9, 1]	10	10

Luego, con estos valores que tenemos en la tabla podemos calcular el valor de Z, el cual nos da 1.6 , con $\alpha = 0.05$ que es un equivalente a un 95% de confianza , llegamos a la conclusión de que: Los números son independiente según la Prueba de Bondad.

4.2.2 Prueba de autocorrelación

Esta prueba nos pide un margen de error, en donde nosotros colocamos 0.1, y nos pide indicar donde queremos comenzar con la autocorrelación, nosotros elegimos el 2do numero de la serie de números aleatorios y por ultimo, la amplitud que se desea a partir de ese numero, nosotros elegimos 5. Por lo tanto, vamos a comenzar la autocorrelación a partir del segundo numero y luego cada 5 números a partir de ese numero base, es decir, 2, 7, 12, etc.

Esta prueba aplica la función de densidad de probabilidad, la cual nos devuelve el valor de 0.19 y también la desviación estándar, la cual nos devuelve el valor de 0.06

Con esos valores, podemos calcular el valor de Z de la siguiente manera

$$Z = \frac{0.19 - 0.25}{0.06} \tag{15}$$

Dejándonos un valor de Z de:

Z = 0.45
Con el siguiente criterio $|Z_0| > Z_{1-\alpha/2}$ se rechaza y con el siguiente criterio $|Z_0| \leq Z_{1-\alpha/2}$ se acepta. Por lo tanto con los datos obtenidos podemos decir que según la Prueba de autocorrelacion los números son independientes.

4.2.3 Prueba de poker

Se acepta esta prueba como dijimos anteriormente, toma nuestros números aleatorios y revisa sus últimos 5 dígitos para poder categorizarlo de una cierta manera y luego poder obtener una frecuencia de aparición de tal numero en dicha categoría. Los resultados son los siguientes:

Categoria	**Oi**	**Ei**
Todos distintos	35	30.25
Par	57	50.4
2 Pares	4	10.8
Tercia	0	0.9
Tercia y par	0	7.2
Poker	2	0.45
Quintina	2	0.01

Luego de obtener estos datos, debemos ingresar los grados de libertad deseados para llevarlos a una distribución chi-cuadrado y calcular el valor de la significancia Z.

Nosotros elegimos 10 grados de libertad, y el valor de la significancia es:
Z = 0.84 , con $\alpha = 0.05$ y el siguiente criterio de aceptación/rechazo $X_0^2 < X_{\alpha,6}$ Podemos concluir que según esta prueba los números son independientes.

4.3 Prueba de las corridas

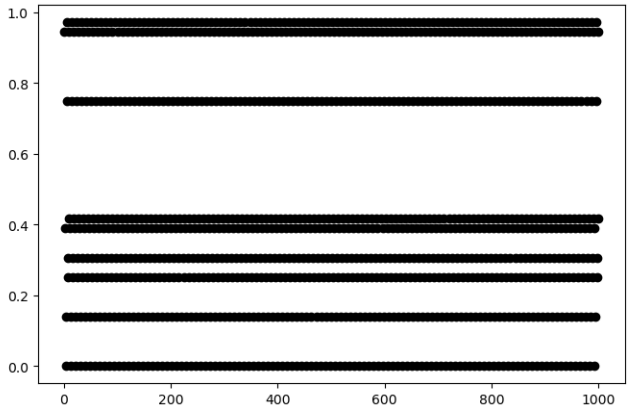
Al someterlo a esta prueba obtuvimos:

[0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0
] con un total de a=81 corridas
Desvio = 66.33
Varianza= 17.45
Significancia Z=0.84
Con $\alpha = 0.05$ y el criterio de aceptación $|Z| \leq Z_{1-\alpha/2}$ y el criterio de rechazo $|Z| > Z_{1-\alpha/2}$
Podemos decir que la secuencia de números es independiente y por la tanto aleatoria.

4.4 Método multiplicativo lineal

Para este método, nos pide una semilla de 2 dígitos, en nuestro caso elegimos 89, una constante multiplicativa, elegimos 34, un numero impar, el cual elegimos 5, un modulo, elegimos 37 y cantidad de tiradas, que elegimos 1000.

Los resultados son los siguientes:



Como se puede observar en este método, los números no son aleatorios, debido a que viéndolo con grandes tiradas, sigue un mismo patrón de generación y no es del todo efectivo.

5 Conclusión

Con la simulación realizada pudimos observar que el Generador congruencial lineal, otorga valores independientes de una variable aleatoria Uniforme [0,1] como pudimos corroborar al someterlo a distintas pruebas/test. Los valores obtenidos son valores no correlacionados, es decir,independientes en termimos estadísticos(cuando un valor no se ve afectado por los valores que toma otro valor).