Test o Contrastes para Números Aleatorios

Cesar Orlando Torres Motta[1]

[1]Universidad de los llanos, Facultad de Ciencias Básicas e Ingeniería, Villavicencio, 2016
[*]cesar.torres@unillanos.edu.co

ABSTRACT

Random numbers can be used in many ways, there are tools that have their own way of generating, excel, mixed congruential method, the method midsquare, among others. This article will explore how as pseudo-random numbers based on the mixed congruential method also how the values behave this from distributions discrete uniform so probability and to check its consistency will be generated and randomness using the Kolmogorov-Smirnov test, chi square and spurts.

Keywords: *uniformity, randomness, generator, test.*

RESUMEN

Los números aleatorios se pueden utilizar de muchas maneras, hay herramientas que tienen su propia forma de generarlos, excel, el metodo congruencial mixto, el método midsquare, entre otros. En este articulo estudiaremos la manera de como se generaran los números pseudoaleatorios partiendo del método congruencial mixto, además la forma en como se comportan los valores, esto a partir de las distribuciones de probabilidad discreta y uniforme de tal manera se pueda comprobar su uniformidad y su aleatoriedad por medio de los test de kolmogorov-smirnov, chi cuadrado y rachas.

Palabras Claves: *uniformidad, aleatoriedad, generador, test.*

Introducción

En este articulo conoceremos como comprobar la uniformidad y aleatoriedad de los generadores de números pseudoaleatorios, aplicaremos por medio de test de kolmogorov, chi cuadrado y test de rachas, de tal forma que determinaremos bondad de ajuste de las distribuciones de probabilidad entre sí.

Resultados

Los resultados obtenidos partiendo del método congruencial mixto fueron los siguientes:

- *RANDU* $x_{i+1} = 65539\, x_i \; mod \; 2^{31}$
  Realizando pruebas de uniformidad:

  En la figura 1 podemos observar es que al realizar el test, los datos que produce el generador esta muy altos de lo que pide la uniforme que es entre (0,1), la linea roja es el test de kolmogorov, la cual indica que los datos no son tan uniformes, los bloques azules muestran la cantidad de veces que cae la función.

  La figura 2 es la ilustración del estadístico de Kolmogorov–Smirnov (K–S). La línea azul es una función de distribución uniforme, la línea verde es una función empírica de distribución acumulada, y la línea roja es el estadístico K–S. Imagen: Tomada de pruebas realizadas en Python de una dsitribución uniforme y el congruencial mixto.

  Realizando el test chi cuadrado para (k=11):
  Como el Estimado = 6.92 es menor que el valor proporcionado por la tabla T = 18.307, SE ACEPTA LA UNIFORMIDAD DE LA SECUENCIA DE VALORES.
  Realizando pruebas de ALEATORIEDAD:
  Rachas = [' ', 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0,

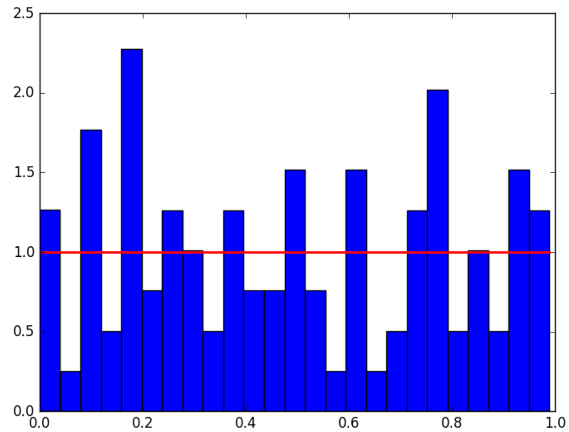

Figure 1. Test de Kolmogorov

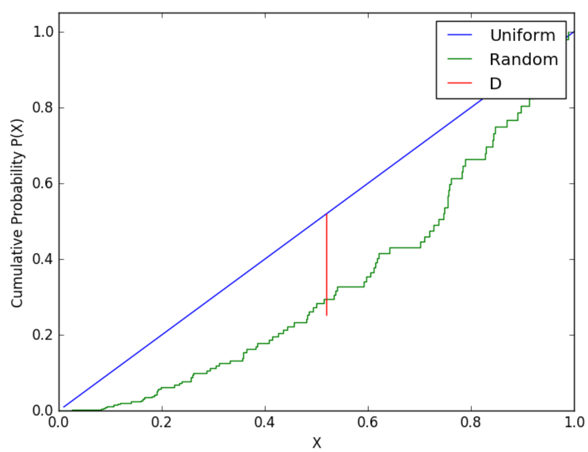

Figure 2. Test de Kolmogorov

0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1]
Numero de rachas = 61
Z = -1.27653254084

Este procedimiento se aplico con diferentes valores obteniendo resultados diversos pero manejando la misma mecánica que se empleo en RANDU. De esta manera podemos entender el comportamiento de los datos y asi mismo sacar conclusiones acerca de la uniformidad y aleatoriedad.

Metodología

Actualmente existen múltiples generadores de números aleatorios en diferentes entornos y compiladores lo cual supondría para un usuario de la Simulación que no es necesario su estudio. Sin embargo, estudios sobre algunos generadores comerciales sugieren que debemos actuar con cuidado con el uso de ellos. Incluso, el uso progresivo de modelos de simulación cada vez más detallados exige generadores de números aleatorios de mayor calidad.[1]

Pruebas o contrastes empíricos:

Normalmente, pasaremos un conjunto de pruebas cuando introduzcamos un nuevo generador, o cuando deseamos analizar si un generador del que desconozcamos su algoritmo subyacente es suficientemente bueno para nuestros propósitos.[1]

Constraste o prueba de Kolmogorov-Smirnov

Consideramos el caso en que F_0 es continua. La función de distribución empírica de una muestra X_1,X_2,...,X_n se define como:[1]

$$F_n\ (x)\ \ =\ \ (\tfrac{X_i \leq x}{n})$$

Bajo la hipótesis nula $H_0 : F_x(x) = F_0(x)$, esperamos que F_n se aproxime a F_0. Definimos el estadístico bilateral de Kolmogorov-Smirnov[1]

$$D_n\ \ = \sup_{x \in \Re} |F_n(x) - F_0(x)|$$

La distribución exacta de D_n está tabulada para valores seleccionados de $n \leq 40$ y del nivel de significación ∞. Para muestras grandes, se utiliza la distribución asintótica de D_n, que viene dada, para todo $Z \leq 0$, por

$$\lim_{n \to \infty} P|(\sqrt{n} D_n \leq z) = L(z) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} + e^{-2i^2 z^2}$$

$L(z)$ está tabulada y se comprueba que la aproximación es suficientemente buena para $n \geq 35$. Intuitivamente, esperamos que D_n sea pequeño cuando la hipótesis nula es cierta. En nuestro caso particular de aleatoriedad, si $X_{(1)}$

$$D_n\ \ =\ \ \max_{i \leq i \leq n}\ [\ max\ [\ [\frac{i}{n} - X_{(i)}], [X_{(i)} - \frac{i-1}{n}]]]$$

Constraste o prueba X^2

La prueba X^2 es de bondad de ajuste. Es poco potente, por lo que permite justificar el rechazo de una hipótesis, pero proporciona escaso soporte a su aceptación. El problema de bondad de ajuste se plantea como sigue. Tenemos una muestra $X_1, X_2, \ldots, X_n$ de una población con distribución $F_x(x)$ desconocida. Deseamos contrastar la hipótesis nula $H_0 : F_x(x) = F_0(x)$, para todo $x \in \Re$, donde $F_0(x)$ está completamente especificada, frente a la alternativa $H_0 : F_x(x) \neq F_0(x)$ para algún x. Para realizar el contraste, partimos el soporte de X en k subconjuntos o clases mutuamente excluyentes. Para cada subconjunto $i, i = 1, \ldots, k$, calculamos el número f_i de observaciones que caen en la clase y el numero esperado de observaciones e_i bajo la distribución hipotética f_0. El estadístico de la prueba es:

$$X^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

cuya distribución asintótica es una X^2_{k-r-1}, donde r es el número de parámetros que se estiman a partir de las observaciones. La aproximación se considera aceptada si $min(e_i > 5)$. Intuitivamente, el estadístico tiende a ser pequeño bajo la hipótesis nula, y grande cuando ésta es falsa, con lo qu ese deduce inmediatamente un contraste de hipótesis. La potencia del contraste crece con el número k de subconjuntos considerados.[1]
En nuestro problema de aleatoriedad, f_0 es la función de distribución de la uniforme en $[0, 1]$, con lo que $r = 0$. Además, pueden cogerse k subintervalos de $[0, 1]$ de igual longitud, con lo que $e_i = n/k$.

Constraste o prueba de rachas

Dada la sucesión de observaciones $X_1, X_2, \ldots, X_n$, construimos la sucesión de símbolos binarios definida mediante 1 si $X_i X_{i+1}$.Definimos racha creciente (decreciente) de longitud. Definimos racha creciente (decreciente) de longitud, a un grupo seguido del números 1 (ó 0). Contabilizamos el número de rachas. Sabemos que su distribución asintótica, bajo la hipótesis nula de aleatoriedad, es:

$$N(\frac{2n-1}{3}, \frac{16n-29}{90})$$

References

1. RÍOS, RÍOS & MARTÍN, 2000, *Simulación, Métodos y aplicaciones.* Editorial Alfaomega, 2000.David Ríos, Sixto Ríos, Jacinto Martín.