

1. Método Monte Carlo

1.1. La Regla de las "Tres Sigmas"

Es posible *verificar*, usando la función de densidad $f(x)$ de una Distribución Normal, que

$$\int_{\mu-3\sigma}^{\mu+3\sigma} f(x) dx = 0.997$$

Cualquiera sean los valores de μ y σ . Esta última expresión puede ser expresada en términos de probabilidades

$$P[\mu - 3\sigma < X < \mu + 3\sigma] = 0.997$$

Siendo X una variable aleatoria Normal. La probabilidad es tan próxima a 1 que la última fórmula se interpreta así: *al efectuar una prueba es prácticamente imposible obtener un valor de X que difiera de μ en más de 3σ*

1.2. Teorema Central del Límite

Considere N variables aleatorias X_1, X_2, \dots, X_N independientes e idénticamente distribuidos. Si

$$\begin{aligned} E[X_1] &= E[X_2] = \dots = E[X_N] = m \\ \text{Var}(X_1) &= \text{Var}(X_2) = \dots = \text{Var}(X_N) = b^2 \end{aligned}$$

Entonces se define $Q_N = X_1 + X_2 + \dots + X_N$ y se deduce que

$$\begin{aligned} E(Q_N) &= Nm \\ \text{Var}(Q_N) &= Nb^2 \end{aligned}$$

Por otra parte, si se define una variable aleatoria **Normal** ξ_N con media y varianza de Nm y Nb^2 respectivamente, el Teorema Central del Límite **afirma** que para cualquier intervalo (a', b') se cumple que (para valores muy grandes de N)

$$P[a' < Q_N < b'] \approx \int_{a'}^{b'} f_{\xi_N}(x) dx$$

El significado real de este teorema es: *la suma Q_N de una gran cantidad de variables idénticas es aproximadamente Normal*

En realidad, este teorema es válido para situaciones mucho más generales: los sumandos X_1, X_2, \dots, X_N pueden no ser idénticos e independientes, necesitando que únicamente cada sumando por separado no influya considerablemente en la suma.

1.3. Esquema General - Monte Carlo

Suponga que se desea calcular una cantidad m que se desconoce. La idea es asociar (crear artificialmente) una variable aleatoria ξ tal que $E(\xi) = m$ con $\text{Var}(\xi) = b^2$.

Considere N variables aleatorias independientes $\xi_1, \xi_2, \dots, \xi_N$ con la misma distribución que ξ . Si N es suficientemente grande, la distribución de la suma $Q_N = \xi_1 + \xi_2 + \dots + \xi_N$ seguirá, de acuerdo al teorema central del límite, una distribución aproximadamente Normal con parámetros $a = Nm$ y $\sigma^2 = Nb^2$. Se deduce entonces que

$$P\left[m - 3\frac{b}{\sqrt{N}} < \frac{Q_N}{N} < m + 3\frac{b}{\sqrt{N}}\right] \approx 0.997$$

Ó

$$P\left[\left|\frac{1}{N} \sum_{i=1}^N \xi_j - m\right| < 3\frac{b}{\sqrt{N}}\right] \approx 0.997$$

Esta relación es el soporte para la utilización del método de Monte Carlo, pues permite calcular **aproximadamente** m y a la vez permite calcular el error de dicho cálculo

En efecto, se determina N valores de la variable aleatoria ξ y de acuerdo a la última expresión, la media aritmética de estos valores será aproximadamente igual a m . Con una gran probabilidad se puede afirmar que el error de esta aproximación no sobrepasa de $3\frac{b}{\sqrt{N}}$ --el error tiende a cero cuando N se incrementa.

La última expresión, también se le conoce como la *Ley Fuerte de los Grandes Números*, y es denotado como

$$\frac{1}{N} \sum_{i=1}^N \xi_j \rightarrow m$$

2. Números Aleatorios

El núcleo de un estudio de simulación es la capacidad de generar números aleatorios que representan el valor de una variable aleatoria Distribuida Uniformemente en $(0, 1)$. En esta sección se explicará la forma de hacerlo mediante el uso de la computadora.

Inicialmente los números aleatorios se generaban de forma manual o mecánica, utilizando ruedas giratorias, lanzamiento de dados, barajas de cartas, etc. Sin embargo, actualmente se hace uso del computador para la obtención de ellas y a estas se les denomina números pseudoaleatorios.

1.1.1. Generación de Números Pseudoaleatorios

Se describirá un método numérico para generar secuencias de números aleatorios. Dado que se utilizará relaciones determinísticas para generar estas secuencias se argumenta que estos números no son verdaderamente aleatorios.

Un método para generar números pseudoaleatorio comienza con un valor inicial denominado de semilla x_0 y luego se calcula de manera recursiva los valores sucesivos x_n , $n \geq 1$ realizando el siguiente cálculo

$$x_n = a x_{n-1} \text{ módulo } m$$

Donde a y m son enteros positivos previamente especificados, y que significa que $a x_{n-1}$ se divide entre m y el residuo se considera como el valor de x_n . Es decir, cada x_n es 0, 1, 2, ..., $m-1$. Y la cantidad $\frac{x_n}{m}$ se considera como una aproximación del valor de una variable aleatoria Uniforme en (0, 1).

En el método anterior, se tiene que x_n después de un cierto número finito (a lo más m) de valores generados, *alguno debe repetirse*, y una vez que esto ocurre toda la secuencia comienza a repetirse. Es decir, se desea encontrar valores a y m tales que, para cualquier semilla x_0 , el número de valores que se puedan generar antes de que suceda esta repetición sea *grande*. Existe una rama denominada de Teoría de la Simulación en donde se estudia *criterios* para la determinación adecuada de a y m teniendo en consideración la codificación del lenguaje de máquina. Para una máquina de 32 bits se ha probado que las elecciones de $m=2^{31}-1$ y $a=7^5=16807$ producen la propiedad deseada.

El método anterior es denominado de *Método Congruencial* y puede ser extendido para

$$x_n = (a x_{n-1} + c) \text{ módulo } m$$

Denominado de *Método Congruencial Mixto*. La mayor parte de lenguajes de computadora tienen integrado un generador de números pseudoaleatorios que se le puede invocar para fines específicos.

EJEMPLO 1

En este ejemplo se verá como se utiliza el *método congruencial mixto* para generar números aleatorios. Se seleccionará parámetros apropiados para producir secuencias de números aleatorios aceptables y no aceptables y se mostrará como detectar las diferencias.

Iniciaremos con un conjunto de parámetros para a , c y m que produzcan una secuencia periódica pequeña

```
a<-7; c<-3; m<-10
```

Una vez que la secuencia debe estar entre 0 y $m-1$ se espera tener 1 a m secuencias en R

Se generará un arreglo bidimensional con los índices de fila que correspondan a semillas iniciales entre 0 y 9

```
z<-seq(0,9)
```

La fórmula del generador está dado por

```
a<-7; c<-3; m<-10
z<-matrix(0,10,10)
z[,1]<-seq(0,9)
```

```
for (j in 1:9) {
  for (i in 1:10) {
    z[i,j+1] <- (a*z[i,j] + c) %% m
  }
}
z
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	0	3	4	1	0	3	4	1	0	3
[2,]	1	0	3	4	1	0	3	4	1	0
[3,]	2	7	2	7	2	7	2	7	2	7
[4,]	3	4	1	0	3	4	1	0	3	4
[5,]	4	1	0	3	4	1	0	3	4	1
[6,]	5	8	9	6	5	8	9	6	5	8
[7,]	6	5	8	9	6	5	8	9	6	5
[8,]	7	2	7	2	7	2	7	2	7	2
[9,]	8	9	6	5	8	9	6	5	8	9
[10,]	9	6	5	8	9	6	5	8	9	6

Existen tres secuencias cíclicas, 2 de longitud 4 - {1 0 3 4} y {5 8 9 6} y uno de longitud 2 - {2 7} con diferentes semillas iniciales. Lo que se desea claramente es una secuencia donde el periodo es largo y se aproxime a $m-1$ y no deba exhibir secuencias periódicas para cualquier semilla inicial.

A continuación se examinará un conjunto de números uniformes generados por los parámetros

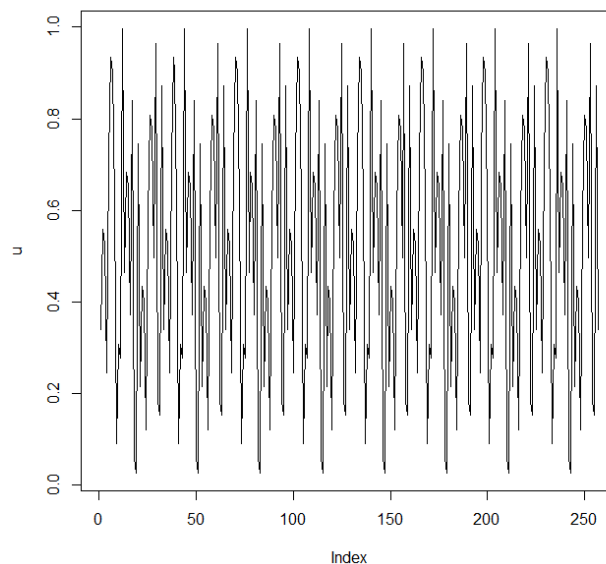
```

a<-137; c<-0; m<-256;
x[1]<-87;
for (j in 1:256){
  x[j+1] <- (a*x[j] +c) %% m
}
u<- x/256
plot(u,type="l")

```

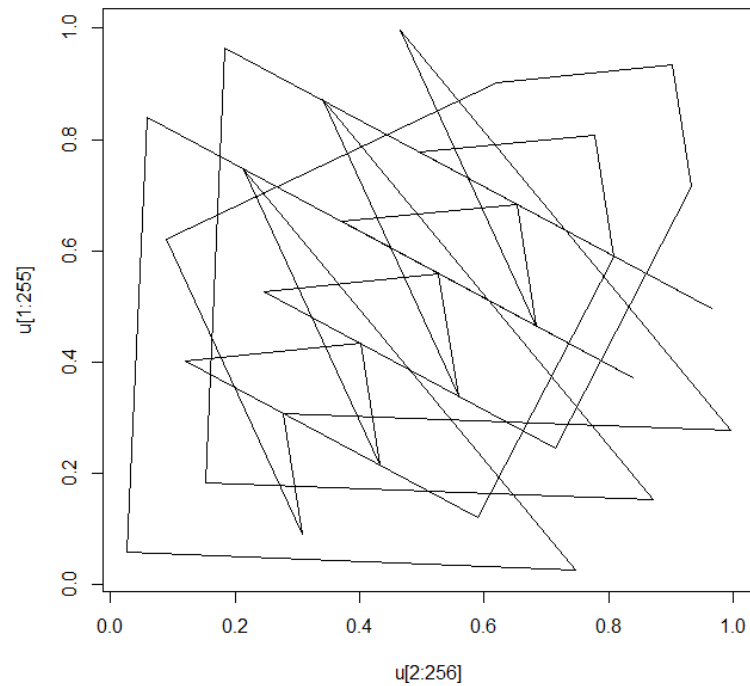
Se debe ejecutar otros test para verificar que no se tiene una secuencia periódica. Si se grafica los valores u_i vs. i descubrimos que una secuencia periódica existe y su periodo es aproximadamente 40.

```
plot(u)
```



Otra gráfica informativa puede ser hecha el cual muestre el intervalo entre números adyacentes en la secuencia. Si se grafica $u_j + 1$ vs. u_j

`plot(u(2:256),u(1:255))`

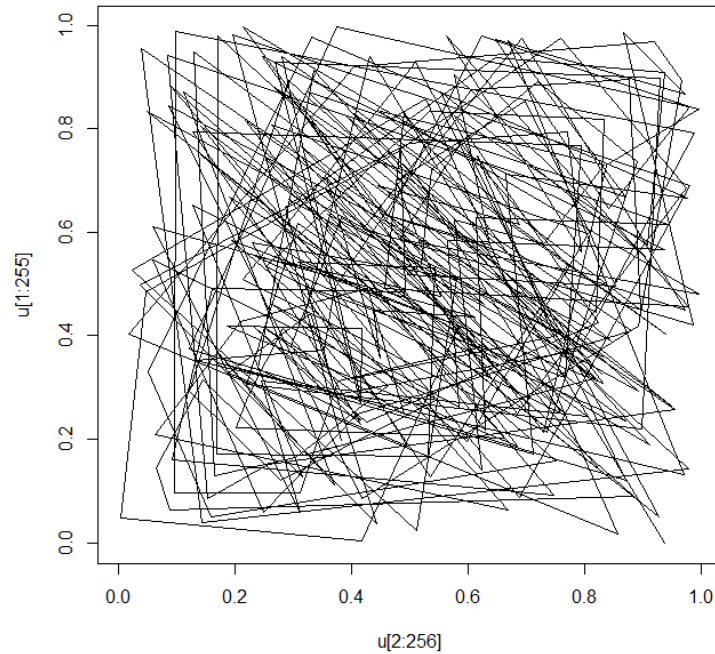


Esta secuencia de gráficos puede realizarse para diferentes N y la longitud en la secuencia puede ser observada incrementando N hasta que el gráfico no cambia. Esto indica que las secuencias están sobreponiéndose una con otra cuando N se incrementa y contiene secuencias repetidas.

Si se usa parámetros apropiados

```
a=314159269; c=453806245; m=2^31-1;
x(1)=87;
for j=1:256
    x(j+1) = mod(a*x(j)+c,m);
end
u=x/m;
```

`plot(u(2:255),u(1:254))`



Cuando N se incrementa se observa que la curva dibuja todo el espacio con ninguna repetición indicando la ausencia de periodicidad en la secuencia.

1.1.2. Uso de Números Aleatorios para Evaluar Integrales

Suponga que se desea calcular la siguiente integral

$$\theta = \int_0^1 g(x) dx$$

Si se asume que $x \sim U(0,1)$, entonces la *función de densidad* de probabilidad de x es dado por $f(x)=1$ para todo $x \in (0,1)$. Luego, es posible expresar a θ como

$$\theta = \int_0^1 1 \cdot g(x) dx = \int_0^1 f(x) g(x) dx = E[g(x)]$$

Es decir, de manera artificial se está utilizando Monte Carlo para determinar *aproximadamente* el valor de θ . Es decir, si se genera de manera independiente u_1, u_2, \dots, u_k mediante la Distribución $U(0,1)$, esto significa que $g(u_1), g(u_2), \dots, g(u_k)$ son valores de variables aleatorias independientes e idénticamente distribuidas con media θ . Por la *Ley Fuerte de los Grandes Números*, se tiene con probabilidad 1

$$\frac{1}{k} \sum_{i=1}^k g(u_i) \longrightarrow E(g(u)) = \theta$$

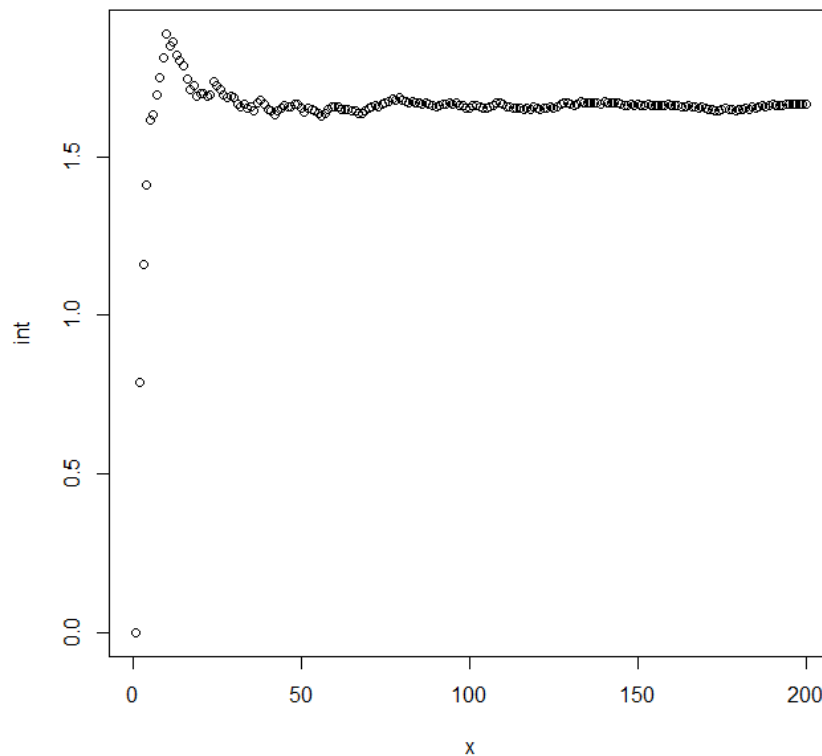
Luego, es posible aproximar θ generando una gran cantidad de números aleatorios u_i y considerando como una aproximación de θ el promedio de $g(u_i)$.

EJEMPLO 2

Emplear Monte Carlo para aproximar el valor de la siguiente integral

$$\int_0^1 e^x dx$$

```
integral<- function(k){  
  ## k: Numero de iteraciones  
  
  s<- c(rep(0,each=k))  
  
  x<-1:k  
  for (i in 2:k) {  
    s[i]<-s[i-1]+exp(runif(1,0,1));  
  }  
  int<-s/x  
  plot(x,int)  
  integral<-int[k]  
  integral  
}  
  
> integral(200)
```



EJEMPLO 3

[Estimación de π]

Suponga que el vector aleatorio (X, Y) se distribuye Uniformemente en el cuadrado de área 4 con centro en el origen. Es decir, es un punto aleatorio en la región especificada por la Figura 8.

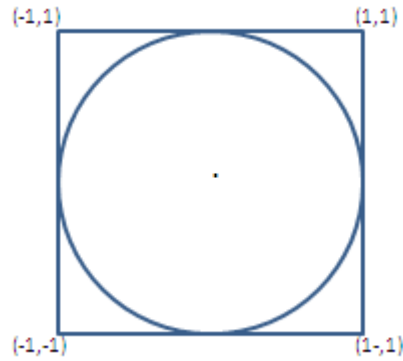


Figura 8.-

Considere la probabilidad de que este punto aleatorio del cuadrado esté dentro del círculo inscrito de radio 1. Como (X, Y) está distribuido Uniformemente en el cuadrado, se tiene que

$$P[(X, Y) \text{ este en el círculo}] = P(x^2 + y^2 \leq 1) = \frac{\text{área Círculo}}{\text{área Cuadrado}} = \frac{\pi}{4}$$

Por lo tanto, si se genera una gran cantidad de puntos aleatorios en el cuadrado, la proporción de puntos que caen dentro del círculo será aproximadamente $\frac{\pi}{4}$. Ahora

Si X e Y fuesen independientes e idénticamente distribuidos (i.i.d) uniformemente en $< -1, 1 >$, la f.d.p. sería $f(x, y) = f(x) \times f(y) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Además si $u^* \sim U(a, b)$ entonces $u = \frac{u^* - a}{b - a} \sim U(0, 1)$. Es decir, $u^* = (b - a) * u + a \sim U(a, b)$. Si hacemos $a = -1$ y $b = 1$ tenemos finalmente que $u^* = (2 * u - 1) \sim U(-1, 1)$.

Para aproximar $P[X^2 + Y^2 < 1]$ debemos hacer $X = u^*$ e $Y = u^*$ y generar $\{(X^i, Y^i)\}_{i=1}^k$ y calcular $\{g(X^i, Y^i)\}_{i=1}^k$, donde $g(X^i, Y^i) = I[(X^i)^2 + (Y^i)^2 < 1]$ y aproximar $P[X^2 + Y^2 < 1] = E[I[X^2 + Y^2 < 1]] = 1 \times P[X^2 + Y^2 < 1] + 0 \times P[X^2 + Y^2 \geq 1]$ por el promedio $\frac{1}{k} \sum_{i=1}^k I[(X^i)^2 + (Y^i)^2 < 1]$