

Capítulo 4

El Método de Monte Carlo

El **método de Monte Carlo** es un procedimiento general para seleccionar muestras aleatorias de una población utilizando números aleatorios.

La denominación Monte Carlo fue popularizado por los científicos Stanislaw Ulam, Enrico Fermi, John von Neumann, y Nicholas Metropolis, entre otros, quienes ya trabajaban sobre **muestreo estadístico**. Su nombre hace referencia al Casino de Montecarlo en Mónaco.

Este método se utiliza para calcular numéricamente expresiones matemáticamente complejas y difíciles de evaluar con exactitud, o que no pueden resolverse analíticamente. En nuestro caso analizaremos el cálculo o estimación de integrales definidas, y aproximaciones al valor de π .

El método de Monte Carlo se basa en dos resultados fundamentales de la Teoría de la Probabilidad:

1. La **Ley Fuerte de los Grandes Números**: Si $X_1, X_2, \dots, X_n, \dots$ son variables aleatorias independientes, idénticamente distribuidas, con media μ , entonces:

$$P\left(\lim_{n \rightarrow \infty} \frac{X_1 + X_2 + \dots + X_n}{n} = \mu\right) = 1, \quad (4.1)$$

o equivalentemente con probabilidad 1

$$\lim_{n \rightarrow \infty} \frac{X_1 + X_2 + \dots + X_n}{n} = \mu.$$

2. Si X es una variable aleatoria absolutamente continua, con función de densidad f , y $g : \mathbb{R} \mapsto \mathbb{R}$ es una función, entonces $g \circ X$ es una variable aleatoria y su valor esperado está dado por

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

Supongamos entonces que se quiere determinar un cierto número θ , desconocido, y que se sabe que este valor puede calcularse como

$$\theta = E[g(X)]$$

para cierta variable aleatoria X que posee una distribución \mathcal{F} . En algunos casos puede ocurrir que, por la naturaleza de la función g , o de la función de densidad de X , resulta muy difícil o imposible determinar $E[g(X)]$. El método de Monte Carlo propone encontrar un **valor estimado** de θ , donde *estimado* significa que hay una alta probabilidad de obtener un valor muy cercano a la verdadera solución. Para esto se considera una secuencia X_1, X_2, \dots de variables aleatorias, independientes, todas con la misma distribución de X , es decir, $X_i \sim \mathcal{F}$ para todo $i \geq 1$. Entonces $g(X_1), g(X_2), \dots, g(X_n), \dots$ son todas variables aleatorias independientes, con media igual $\theta = E[g(X)]$, y por la Ley Fuerte de los Grandes Números se tiene que:

$$\theta = \lim_{n \rightarrow \infty} \frac{g(X_1) + g(X_2) + \dots + g(X_n)}{n},$$

con probabilidad 1.

Notemos que el lado izquierdo de la igualdad es un número, mientras que lo que está a la derecha de la igualdad es el límite de una variable aleatoria. Por esto se indica que la probabilidad de que este límite sea igual a θ es 1.

También es importante aclarar que la Ley Fuerte de los Grandes Números no determina que existe una cota para el error

$$\left| \theta - \frac{g(x_1) + g(x_2) + \dots + g(x_n)}{n} \right|$$

para cualquier realización de las X_i :

$$X_1 = x_1, \quad X_2 = x_2, \quad X_3 = x_3, \quad \dots \quad X_n = x_n,$$

y un n suficientemente grande; sino que establece que para **casi** todas las realizaciones ocurre que

$$\lim_{n \rightarrow \infty} \frac{g(x_1) + g(x_2) + \dots + g(x_n)}{n} = \mu.$$

Dicho de una manera más coloquial, es prácticamente improbable que una realización de las X_i 's no cumpla que el límite de sus promedios

$$\frac{g(x_1) + g(x_2) + \dots + g(x_n)}{n}$$

se aproxime a θ a medida que n tiende a infinito.

4.1. Estimación de integrales definidas

4.1.1. Integración sobre $(0, 1)$

Una de las aplicaciones del método de Monte Carlo es facilitar el cálculo de integrales definidas. En realidad no se determina el valor exacto, sino que se utiliza para **estimar** el valor de la integral principalmente en casos que el cálculo analítico no es posible.

Veamos primero el caso en que se desee calcular el valor θ de una integral definida en el intervalo $[0, 1]$:

$$\theta = \int_0^1 g(x) dx. \quad (4.2)$$

Si ahora consideramos una variable aleatoria uniforme U , $U \sim \mathcal{U}(0, 1)$, entonces la función de densidad de U es

$$f(x) = I_{(0,1)}(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{c.c} \end{cases},$$

y por lo tanto (4.2) se puede escribir como un valor esperado:

$$\theta = \int_{-\infty}^{\infty} g(x) f(x) dx = \int_0^1 g(x) f(x) dx = E[g(U)].$$

Ahora, por la Ley Fuerte de los grandes números podemos considerar una sucesión de N variables aleatorias uniformes $U_i \sim \mathcal{U}(0, 1)$, independientes, y aproximar el valor θ con el límite de promedios:

$$\lim_{N \rightarrow \infty} \frac{g(U_1) + g(U_2) + \cdots + g(u_N)}{N} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N g(U_i) \simeq \theta.$$

En la práctica, esta integral puede aproximarse con una muestra de tamaño N suficientemente grande: $U_1 = u_1, U_2 = u_2, \dots, U_N = u_N$ y estimar θ con:

$$\theta = \frac{1}{N} \sum_{i=1}^N g(u_i).$$

Ejemplo 4.1. Para estimar el valor de la integral

$$\int_0^1 (1 - x^2)^{3/2} dx,$$

consideramos una realización de N variables aleatorias uniformes independientes:

$$U_1 = u_1, \quad U_2 = u_2, \quad \dots \quad U_N = u_N$$

y aproximamos el valor de la integral por

$$\int_0^1 (1 - x^2)^{3/2} dx \sim \frac{1}{N} \sum_{i=1}^N (1 - u_i^2)^{3/2}.$$

```

from random import random

def g(u): ##función a integrar en (0,1)
    return (1 - u ** 2) ** (1.5)
##estimación de la integral de g con Nsim simulaciones
def MonteCarlo(g, Nsim):
    Integral = 0
    for _ in range(Nsim):
        Integral += g(random())
    return Integral/Nsim

```

El valor exacto de esta integral es $\frac{3\pi}{16}$, aproximadamente 0.5890486226. (Ver Figura 4.1 y Cuadro 4.1).

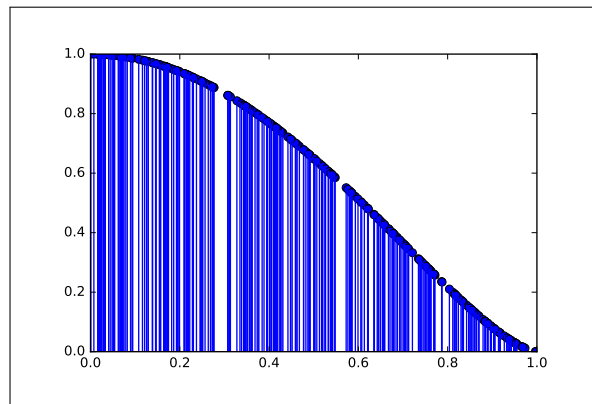


Figura 4.1: $N = 300$, Valor estimado=0.6067846103

| # Simulaciones | Valor estimado |
|----------------|--------------------|
| 1000 | 0.5788611891947463 |
| 10000 | 0.5863982284155244 |
| 100000 | 0.5906798799945882 |
| 1000000 | 0.5882139637408895 |

Cuadro 4.1: Estimaciones de $\int_0^1 (1 - x^2)^{3/2} dx$

4.1.2. Integración sobre un intervalo (a, b)

Para estimar el valor de una integral definida, sobre un intervalo (a, b) , con a y b reales, se aplica un cambio de variables para transformarla en una integral entre 0 y 1. Esto es, si

$$\theta = \int_a^b g(x) dx,$$

con $a < b$, entonces definimos la variable y :

$$y = \frac{x - a}{b - a}, \quad dy = \frac{1}{b - a} dx$$

y así el valor de θ puede calcularse como:

$$\int_a^b g(x) dx = \int_0^1 g(a + (b - a)y)(b - a) dy = \int_0^1 h(y) dy.$$

donde

$$h(y) = g(a + (b - a)y)(b - a), \quad y \in (0, 1).$$

Ejemplo 4.2. Para estimar el valor de la integral

$$\int_{-1}^1 e^{x+x^2} dx,$$

realizamos un cambio de variable:

$$y = \frac{x + 1}{2}, \quad dy = \frac{1}{2} dx.$$

Luego

$$\int_{-1}^1 e^{x+x^2} dx = \int_0^1 e^{(2y-1)+(2y-1)^2} 2 dy = \int_0^1 e^{(2y-1)(2y)} 2 dy.$$

Una estimación con N simulaciones estará dada por el valor de una expresión:

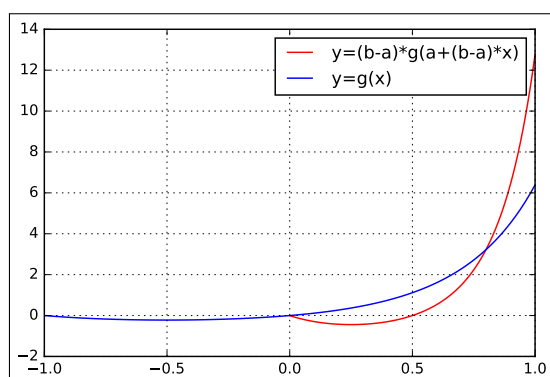
$$\frac{2}{N} \sum_{i=1}^N e^{(2u_i-1)+(2u_i-1)^2} = \frac{2}{N} \sum_{i=1}^N e^{(2u_i-1)(2u_i)},$$

donde u_1, u_2, \dots, u_n es una realización de las variables aleatorias U_1, U_2, \dots, U_N , todas uniformes en $(0, 1)$ e independientes entre sí.

Notar que si $U \sim U(0, 1)$, entonces $2U - 1 \sim U(-1, 1)$.

```
## Función a integrar
```

```
def funciong(x):
    return np.exp(x ** 2 + x)
```

Figura 4.2: Gráfico de $g(x) = e^{x+x^2}$ con y sin cambio de variables

```
##Estima la integral de funciong entre a y b con Nsim simulaciones
def IntegralMonteCarlo(funciong, a, b, Nsim):
    Integral = 0
    for _ in range(Nsim):
        Integral += g(a + (b-a) * random())
    return Integral * (b-a)/Nsim
```

| # simulaciones | valor estimado |
|----------------|----------------|
| 1000 | 3.60200514128 |
| 10000 | 3.59901204683 |
| 100000 | 3.56130047278 |
| 1000000 | 3.59370623674 |
| 10000000 | 3.58717156846 |

Cuadro 4.2: Estimaciones de la integral

4.1.3. Integración sobre $(0, \infty)$

En el caso de la estimación de una integral en el intervalo $(0, \infty)$:

$$\theta = \int_0^{\infty} g(x) dx,$$

también se aplica un cambio de variables, transformando biyectivamente el intervalo $(0, \infty)$ en $(0, 1)$. Un cambio de variables posible es el siguiente:

$$y = \frac{1}{x+1}, \quad dy = -\frac{1}{(x+1)^2} dx = -y^2 dx.$$

Luego se tiene que:

$$\int_0^\infty g(x) dx = - \int_1^0 \frac{g(\frac{1}{y} - 1)}{y^2} dy = \int_0^1 \frac{g(\frac{1}{y} - 1)}{y^2} dy = \int_0^1 h(y) dy,$$

con

$$h(y) = \frac{1}{y^2} g\left(\frac{1}{y} - 1\right).$$

Ejemplo 4.3. Para estimar la siguiente integral:

$$\int_0^\infty \cos(x) e^{-x} dx$$

se aplica el método de Monte Carlo para la estimación de la integral con el cambio de variables propuesto:

$$\int_0^1 \frac{\cos(\frac{1}{x} - 1) e^{-(\frac{1}{x} - 1)}}{x^2} dx \sim \theta \sim \frac{1}{N} \sum_{i=1}^N \frac{\cos(\frac{1}{u_i} - 1) e^{-(\frac{1}{u_i} - 1)}}{u_i^2}.$$

| # simulaciones | valor estimado |
|----------------|----------------|
| 1000 | 0.540652067791 |
| 10000 | 0.503039650709 |
| 100000 | 0.4991965288 |
| 1000000 | 0.499299312179 |

Cuadro 4.3: Estimaciones de la integral

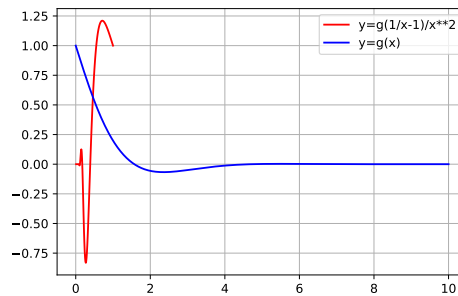


Figura 4.3: Gráfico de $g(x) = \cos(x) e^{-x}$, con y sin cambio de variables

4.2. Estimación de integrales múltiples

El método de Monte Carlo para el cálculo de integrales en una variable no es muy eficiente comparado con otros métodos numéricos que convergen más rápidamente al valor de la integral.

Sin embargo, para la estimación de integrales múltiples este método cobra mayor importancia ya que computacionalmente es menos costoso.

Nuevamente, una integral múltiple de una función en varias variables definida en un hiper-cubo de lado 1 puede estimarse con el método de Monte Carlo.

Para calcular la cantidad

$$\theta = \int_0^1 \cdots \int_0^1 g(x_1, \dots, x_l) dx_1 \dots dx_l$$

utilizamos el hecho que

$$\theta = E[g(U_1, \dots, U_l)]$$

con U_1, \dots, U_l independientes y uniformes en $(0, 1)$. Esto es así porque su distribución conjunta está dada por:

$$f(x_1, x_2, \dots, x_l) = \mathbb{I}_{(0,1) \times (0,1) \times \dots \times (0,1)}(x_1, x_2, \dots, x_l),$$

y entonces

$$\theta = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_l) f(x_1, \dots, x_l) dx_1 \dots dx_l.$$

Si se tienen N muestras independientes de estas l variables,

$$(U_1^1, \dots, U_l^1), \quad (U_1^2, \dots, U_l^2), \quad \dots \quad (U_1^N, \dots, U_l^N)$$

podemos estimar el valor de θ como

$$\theta \sim \frac{1}{N} \sum_{i=1}^N g(U_1^i, \dots, U_l^i)$$

4.2.1. Estimación del valor de π

Una aplicación de Monte Carlo en su uso para la estimación de integrales múltiples, es el cálculo estimado del valor de π . Recordemos que el área de un círculo de radio r es $\pi \cdot r^2$. Si tomamos $r = 1$, entonces π está dado por el valor de una integral:

$$\pi = \int_{-1}^1 \int_{-1}^1 \mathbb{I}_{\{x^2+y^2 < 1\}}(x, y) dx dy.$$

Si X e Y son v.a. independientes, uniformes en $(-1, 1)$, ambas con densidad

$$f_X(x) = f_Y(y) = \frac{1}{2} \cdot \mathbb{I}_{(-1,1)}(x),$$

entonces su densidad conjunta es igual al producto de sus densidades:

$$f(x, y) = f_X(x) \cdot f_Y(y) = \frac{1}{4} \cdot \mathbb{I}_{(-1,1) \times (-1,1)}(x, y).$$

En particular, (X, Y) resulta un vector aleatorio con distribución uniforme en $(-1, 1) \times (-1, 1)$, y tenemos que

$$\pi = \int_{-1}^1 \int_{-1}^1 \mathbb{I}_{\{x^2+y^2 < 1\}}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 4 \cdot \mathbb{I}_{\{x^2+y^2 < 1\}}(x, y) f(x, y) dx dy.$$

Entonces $\frac{\pi}{4} = E[g(X, Y)]$ donde $g(x, y) = \mathbb{I}_{\{x^2+y^2 < 1\}}(x, y)$. Así, para estimar π podemos generar secuencias de pares (X_i, Y_i) , $i \geq 1$, donde X_i e Y_i son variables aleatorias uniformes en $(-1, 1)$, y luego estimar el valor de π como:

$$4 \cdot \frac{1}{N} \sum_{j=1}^N \mathbb{I}_{x^2+y^2 \leq 1}(x_i, y_i).$$

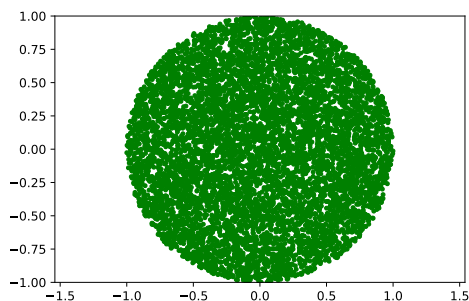
En otras palabras, π será estimado por la proporción de pares (X, Y) que caigan dentro del círculo de radio 1, multiplicado por 4.

Notemos que si $U_1, U_2 \sim U(0, 1)$, entonces

$$X = 2U_1 - 1 \quad Y = 2U_2 - 1$$

verifican $X, Y \sim U(-1, 1)$.

```
def valorPi(Nsim):
    enCirculo = 0.
    for _ in range(Nsim):
        u = 2 * random() - 1
        v = 2 * random() - 1
        if u ** 2 + v ** 2 <= 1:
            enCirculo += 1
    return 4 * enCirculo/Nsim
valorPi(Nsim)
```



| # Simulaciones | Valor estimado |
|----------------|----------------|
| 1000 | 3.16 |
| 10000 | 3.1216 |
| 100000 | 3.14292 |
| 1000000 | 3.141056 |
| 10000000 | 3.1420524 |