Simulación Monte Carlo

SAMJ

2025/05/08

1 Introducción

El método de Monte Carlo es una técnica probabilística utilizada para resolver problemas matemáticos complejos, especialmente aquellos difíciles de abordar de manera exacta. Se basa en generar números aleatorios y observar el comportamiento de un sistema mediante simulaciones repetidas.

Se emplea frecuentemente para:

- Aproximar integrales definidas.
- Resolver ecuaciones diferenciales estocásticas.
- Simular sistemas físicos o financieros complejos.

2 Idea principal

Sea X una variable aleatoria de interés. Supongamos que queremos calcular una cantidad como $\mathbb{E}[f(X)]$ o $\mathbb{P}(X \in A)$. En realidad, esta última es un caso particular de la primera.

El método básico de Monte Carlo consiste en:

1. Generar n muestras i.i.d. $X_1, X_2, \ldots, X_n$ de la distribución de X.
2. Calcular el promedio muestral:
$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i).$$
3. Por la **ley de los grandes números**, se tiene que $\hat{\mu}_n \to \mathbb{E}[f(X)]$ cuando $n \to \infty$ (en algún sentido preciso, no especificado aquí).

3 Aplicaciones

Consideramos la esperanza de una función de una variable aleatoria uniforme. Si $X \sim \text{Uniforme}(a, b)$, entonces:

$$\mathbb{E}[f(X)] = \frac{1}{b-a} \int_a^b f(x)\,dx$$

Despejando:

$$\int_a^b f(x)\,dx = (b-a) \cdot \mathbb{E}[f(X)]$$

Aproximamos la esperanza mediante el promedio muestral:

$$\mathbb{E}[f(X)] \approx \frac{1}{n} \sum_{i=1}^{n} f(X_i)$$

donde los X_i son n muestras independientes de la distribución $\text{Uniforme}(a, b)$.

4 Ejemplo 1: Área bajo un semicírculo

Queremos aproximar el área bajo la mitad de un semicírculo de radio 1 en el intervalo $[0, 1]$, es decir:

$$\int_0^1 \sqrt{1-x^2}\,dx$$

Esta integral representa un cuarto del área de un círculo de radio 1, cuyo valor exacto es $\frac{\pi}{4} \approx 0.7854$.

Paso 1: Generar muestras aleatorias

Generamos n muestras independientes de $X_i \sim \text{Uniforme}(0, 1)$.

Paso 2: Evaluar la función

Calculamos $f(X_i) = \sqrt{1 - X_i^2}$ para cada muestra.

Paso 3: Calcular el promedio

Aproximamos la integral como:

$$\int_0^1 \sqrt{1-x^2}\,dx \approx \frac{1}{n} \sum_{i=1}^{n} \sqrt{1 - X_i^2}$$

```
import numpy as np

# Nmero de simulaciones
n = 100000

# Paso 1: Muestras uniformes en [0,1]
X = np.random.uniform(0, 1, n)

# Paso 2: Evaluar la funcin del semicrculo
f_X = np.sqrt(1 - X**2)

# Paso 3: Calcular el promedio
area_aprox = np.mean(f_X)

print(f"Aproximacin␣del␣rea␣bajo␣el␣semicrculo:␣{area_aprox:.4f}")
```

5 Ejemplo 2: Comparación de ingresos

Ana proviene de una población de científicos de datos en Estados Unidos, cuyos ingresos anuales siguen una distribución Gamma con parámetros $\alpha = 4$ y $\beta = 25{,}000$, lo que implica una media de $\mathbb{E}[X_A] = 100{,}000$ dólares y una desviación estándar de $\sigma_A = 25{,}000$.

Luis, por otro lado, pertenece a una población en México, donde los ingresos siguen una distribución Gamma con $\alpha = 3$ y $\beta = 6{,}667$, resultando en una media de $\mathbb{E}[X_L] = 20{,}000$ dólares y una desviación estándar de $\sigma_L = 12{,}247$.

Dado que Ana gana más que el 50% de su población y Luis más que el 90%, estimamos:

$$\mathbb{P}(X_A \geq 3X_L \mid X_A > q^A_{0.5},\, X_L > q^L_{0.9})$$

donde q^i_p denota el percentil p de la distribución correspondiente para $i \in \{A, L\}$.

```
import numpy as np
import scipy.stats as stats

# Parmetros de las distribuciones Gamma
alpha_A, beta_A = 4, 25000 # EE.UU. (Ana)
alpha_L, beta_L = 3, 6667 # Mxico (Luis)

# Nmero de simulaciones
n = 10_000_000

# Paso 1: Generar muestras
X_A = np.random.gamma(alpha_A, beta_A, n)
X_L = np.random.gamma(alpha_L, beta_L, n)

# Paso 2: Calcular percentiles
q_A_50 = np.percentile(X_A, 50)
q_L_90 = np.percentile(X_L, 90)

# Paso 3: Filtrar muestras
```

```
X_A_filtradas = X_A[X_A > q_A_50]
X_L_filtradas = X_L[X_L > q_L_90]

# Paso 4: Igualar tamaos para comparacin
min_len = min(len(X_A_filtradas), len(X_L_filtradas))
X_A_filtradas = X_A_filtradas[:min_len]
X_L_filtradas = X_L_filtradas[:min_len]

# Paso 5: Estimar la probabilidad
prob = np.mean(X_A_filtradas >= 3 * X_L_filtradas)

print(f"Probabilidad de que Ana gane al menos el triple que Luis: {prob:.4f}")
```

6 Convergencia y precisión

- **Ley de los grandes números:** garantiza la convergencia del promedio muestral hacia el valor esperado.
- **Teorema central del límite:** para n grande,

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

donde $\sigma^2 = \mathrm{Var}(X)$. Por lo tanto, el error es del orden de $O(n^{-1/2})$.

7 Ventajas y limitaciones

Ventajas

- Fácil de implementar.
- Maneja bien problemas de alta dimensión.
- Flexible y aplicable a una gran variedad de contextos.

Limitaciones

- Convergencia lenta: el error disminuye como $1/\sqrt{n}$.
- Puede ser computacionalmente costoso.
- La aleatoriedad introduce variabilidad entre ejecuciones.