Laboratorio 1.
Simulación de Monte Carlo I

CI6314 - Modelamiento y Simulación de Sistemas

August 14, 2023

Overview

¿Qué es el Metódo de Monte Carlo?

El método de Monte Carlo es una aplicación de la Ley Fuerte de los Grandes Números y consiste en realizar muchas repeticiones de un experimento aleatorio hasta encontrar resultados "estables".

Es uno de los métodos más sencillos para simular variables y eventos aleatorios

¿Qué es el Método de Monte Carlo?

Sea γ un valor que se desea estimar. Si existe una función f y una familia de variables aleatorias $(x^{(1)}, \cdots, x^{(N)})$ independientes e idénticamente distribuidas, fáciles de calcular computacionalmente que satisfacen que:

$$\mathbb{E}(f(x^{(i)})) = \gamma$$

El método de Monte Carlo consiste en obtener una realización de las v.a. $(x^{(1)}(\omega), \cdots, x^{(N)}(\omega))$ y aproximar γ por la media empírica:

$$\gamma \simeq \hat{S}_N(\omega) := \frac{1}{N} \sum_{l=1}^{N} f(x^{(l)}(\omega))$$

¿Qué es el Método de Monte Carlo?

Conceptos y definiciones

Experimento: fenómeno que se está representando
Ej: Lanzar 2 dados de seis caras.

Objetivo o Evento aleatorio de interés: resultado del experimento que queremos estimar.
Ej: que la suma de ambos dados sea igual o superior a 7.

Variables aleatorias: variables usadas para representar los resultados del experimento.
Ej: tirar un dado se representa como la parte entera de una distribución uniforme entre 1 y 6.

Repeticiones: número de veces que se calcula el evento de interés.

Experimento con dados

Experimento

Lanzar 2 dados perfectamente balanceados de seis caras numeradas del 1 al 6

Objetivos

1. ¿Cuál es el valor esperado de la suma?
2. ¿Cuál es la probabilidad de sacar exactamente 7?

Ejemplo

Experimento con dados

Como son experimentos sencillos se puede calcular el valor esperado de los objetivos analíticamente:

Valor esperado de la suma

$$\mathbb{E}(d_1 + d_2) = \mathbb{E}(d_1) + \mathbb{E}(d_2) = \frac{1+6}{2} + \frac{1+6}{2} = 7$$

Probabilidad de sacar exactamente 7

$$\begin{aligned}
\mathbb{P}(d_1 + d_2 = 7) = \quad & \mathbb{P}(d_1 = 1, d_2 = 6) + \mathbb{P}(d_1 = 2, d_2 = 5) \\
& + \mathbb{P}(d_1 = 3, d_2 = 4) + \mathbb{P}(d_1 = 4, d_2 = 3) \\
& \mathbb{P}(d_1 = 5, d_2 = 2) + \mathbb{P}(d_1 = 6, d_2 = 1) \\
\mathbb{P}(d_1 + d_2 = 7) = \quad & \tfrac{6}{36} \approx 0,1666
\end{aligned}$$

Intervalo de confianza

Definimos un intervalo al nivel $C = (1 - \alpha)\%$ de confianza para la media μ :

$$c.i. = (\mu - z_C \sigma_{\bar{x}}\ ,\ \mu + z_C \sigma_{\bar{x}})$$

donde $z_C = z_{(1-\frac{\alpha}{2})}$ corresponde a la inversa de una distribución normal acumulada para un intervalo de tamaño $\alpha/2$

$$z_{(1-\frac{\alpha}{2})} = N^{-1}\left(1 - \frac{\alpha}{2}\right)$$

y $\sigma_{\bar{x}}$ es el error estándar de la media de una muestra de tamaño n.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Como no conocemos los valores reales de μ y σ, tenemos que usar estimadores basados en nuestra muestra. Un estimador para la media es la media muestral $\bar{X}$, mientras que un estimador insesgado de la varianza es $\hat{\sigma}$.

$$\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Si n es pequeño z_C se calcula a partir la inversa de una distribución t-Student con $n-1$ grados de libertad (en lugar de la normal):

$$z_C \approx t_C = \phi^{-1}\left(\frac{1+C}{2}, n-1\right)$$

Intervalo de confianza

El intervalo de confianza se calcula como:

$$c.i. = \left(\bar{X} - t_C \frac{\hat{\sigma}}{\sqrt{n}} \, , \, \bar{X} + t_C \frac{\hat{\sigma}}{\sqrt{n}} \right)$$

¿Qué tan bueno es el intervalo de confianza que calculamos?

Réplicas necesarias

Para mejorar la precisión de la simulación debemos aumentar el número de replicas, pero ¿cuántas son suficientes?

Determinar un número adecuado de réplicas dependerá de la **varianza** $\hat{\sigma}$ de la muestra, el **nivel de confianza** α deseado y del **error** que estamos dispuestos a asumir (ancho del intervalo). Recordemos que el ancho del intervalo es:

$$e = z_C \frac{\hat{\sigma}}{\sqrt{n}}$$

Reordenando, la cantidad mínima de réplicas necesarias son

$$n \geq z_C^2 \frac{\hat{\sigma}^2}{e^2}$$

Réplicas necesarias

Con el resultado anterior podemos definir la cantidad mínima de réplicas necesarias de la siguiente forma:

1. Ejecutamos n_0 réplicas iniciales.
2. Estimamos $\hat{\sigma}$ con la muestra actual y calculamos n.
3. Ejecutamos las $n - n_0$ réplicas faltantes

Nota 1: Si las réplicas iniciales producen una mala estimación de la varianza, lo esperable es que se esté sobreestimando, por lo tanto no va a aumentar el n final (no es problema)

Nota 2: Para definir e necesitamos conocer el orden de magnitud del nuestro resultado esperado.

Réplicas necesarias

¿Cuántas réplicas se necesitan para estimar los ejercicios anteriores con un error de 0,1 y un nivel de confianza del 90%?

¿Se obtienen resultados más precisos?

¿Hay algún cambio según el n_0 inicial considerado?

Costo de la precisión

Determinar el nivel de precisión y confianza adecuado es fundamental para tener resultados relevantes, pero puede ser muy caro. Consideremos para el problema de la suma estimando $\sigma^2 = 2,3$

Table: Número de réplicas mínimas para distintos niveles de valores de α y e.

α	z	0,1	0,05	0,01	0,001
0,9	1,64	622	2.489	62.227	6.222.750
0,95	1,96	884	3.534	88.354	8.835.355
0,97	2,17	1.083	4.333	108.314	10.831.372
0,98	2,33	1.245	4.979	124.474	12.447.357
0,99	2,58	1.526	6.104	152.603	15.260.262

Jugar CRAPS

Empleé el Método de Monte Carlo para simular un juego de CRAPS, con las siguientes reglas (instrucciones del juego):

1. El jugador lanza dos dados de 6 caras.
   - Gana si los dados suman 7 u 11.
   - Pierde si los dados suman 2, 3 o 12.
   - En otro caso se guarda el resultado (punto) y sigue jugando.
2. El jugador sigue lanzando los dados hasta que:
   - Si repite el resultado de la primera tirada original (punto) gana
   - Si obtiene un 7 pierde.

¿Cuál es la probabilidad de ganar? ¿Cuál es el número esperado de tiradas para un jugador?