Trabajo práctico 1

Aplicaciones del método Montecarlo

Introducción

El método de *Monte Carlo* es una técnica poderosa que utiliza simulaciones para resolver problemas matemáticos o físicos, entre los que se pueden mencionar la simulación de variables aleatorias, resolución de problemas numéricos complejos, estimación de parámetros, etc. En los siguientes ejercicios se propone la resolución de algunos de los problemas típicos que pueden abordarse mediante este método.

Ejercicio 1: Simulación de muestras de una distribución normal

Uno de los posibles usos para el método Monte Carlo es la generación de muestras de una variable aleatoria con cierta distribución. Como caso particular, consideramos en este ejercicio la distribución normal estándar que se puede obtener a partir de la transformación de Box Muller. Este método permite transformar dos variables aleatorias uniformes independientes en dos variables normales independientes.

(a) Utilizando la transformación de Box Muller, genere dos variables aleatorias con distribución normal $Z_1 \sim N(0;1)$ y $Z_2 \sim N(0;1)$ a partir de dos variables uniformes $U_1 \sim U(0;1)$ y $U_2 \sim U(0;1)$. Para ello utilice las siguientes transformaciones:

$$Z_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2), \quad Z_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2).$$

Considere una simulación de $N = 10000$ muestras para ambas variables uniformes. Estime la media y varianza de Z_1 y Z_2 y grafique los histogramas correspondientes. Realice también un gráfico de dispersión (scatter plot) de Z_2 versus Z_1 y calcule el coeficiente de correlación de Pearson entre ambas variables simuladas para verificar su independencia.

(b) Demuestre que dada una variable aleatoria normal estándar $Z = N(0;1)$, la media de una variable normal definida a partir de la transformación $X = \sigma Z + \mu$ es una variable normal $\sim N(\mu;\sigma^2)$ de media $\mathbb{E}[X] = \mu$ y varianza $\mathbb{V}[X] = \sigma^2$.

(c) Aplicando Box Muller para la obtener muestras normales estándar, genere tres nuevas variables aleatorias normales $X_1 \sim N(0;2)$, $X_2 \sim N(1;2)$ y $X_3 \sim N(1;4)$. Estime en cada caso la media y varianza para su comparación con los valores teóricos. Grafique los histogramas de cada variable superpuesto a las curvas de densidad de probabilidad teóricas para verificar la coincidencia.

Ejercicio 2: Resolución de Integrales

Otra forma en la que se puede aplicar el método de Monte Carlo es en la resolución de problemas numéricos. Un caso típico es la resolución de integrales definidas, especialmente útil para la resolución de integrales muy complejas. Supongamos una integral genérica de la forma:

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

el método de Monte Carlo se basa en generar N muestras aleatorias $x_1, x_2, ..., x_N$ independientes y uniformemente distribuidas en el intervalo $[a, b]$ que representan los posibles valores de la variable de integración x. Luego, la integral se puede aproximar mediante el promedio de los valores de la función evaluada en esas muestras como:

$$I \approx \frac{b-a}{N} \sum_{i=1}^{N} g(x_i).$$

Esta técnica es particularmente útil para la resolución de integrales en múltiples dimensiones. No obstante, en este ejercicio vamos a introducir el método aplicado a la resolución de una integral más simple, en una sola dimensión. Como ejemplo, vamos a suponer que queremos calcular la probabilidad de que una variable aleatoria X se encuentre en un intervalo determinado, es decir $\mathbb{P}(a < X < b)$. Suponga que X es una variable aleatoria normal de media $\mu = 2$ y varianza $\sigma^2 = 3$ con función de densidad de probabilidad:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(a) Calcule $\mathbb{P}(\mu - \sigma < X < \mu + \sigma)$ con $N = 10000$ muestras.

(b) Calcule $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma)$ con $N = 10000$ muestras.

(c) Calcule $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma)$ con $N = 10000$ muestras.

(d) Calcule $\mathbb{P}(\mu - \sigma < X < \mu + \sigma)$ para distintos valores de N (10, 10^2, 10^3, 10^4, 10^5, 10^6). Para comparar los resultados, grafique el error cuadrático medio $MSE = \mathbb{E}[(\hat{I}_N - I)^2]$ en función N, donde $\hat{I}_N$ es el resultando de la integral para N muestras e I el valor teórico supuesto de la integral (asuma para este caso $I = 0{,}682687273250961$).

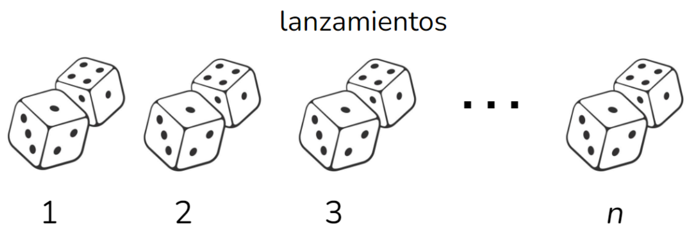
Para estimar el MSE, podemos hacer múltiples realizaciones (repetir el cálculo de la integral). Teniendo en cuenta la Ley de los Grandes Números, el promedio de $(\hat{I}_N - I)^2$ converge a su valor esperado conforme el número de términos $M \to \infty$, aplicando la siguiente ecuación:

$$\widehat{MSE} = \frac{1}{M} \sum_{k=1}^{M} (\hat{I}_{k,N} - I)^2$$

En este caso, suponga $M = 50$ realizaciones.

Ejercicio 3: Simulación de lanzamiento de Dados

Suponga un juego que consiste en el lanzamiento de dos dados en simultáneo. Cuando los números de ambos dados sumen 7 u 11, se considera un punto. El objetivo es realizar un total de n lanzamientos, llamémoslo *turno*, y contabilizar la cantidad puntos acumulados por turno. En este ejercicio estudiaremos la estadística de esta jugada aplicando Monte Carlo para emular múltiples lanzamientos.



(a) Suponiendo que tenemos dos dados no pesados (es decir que cada cara tiene la misma probabilidad de salir que las demas) considerando que un evento de éxito será cuando la suma de ambos dados dé 7 u 11. Determine la probabilidad de ocurrencia de ese evento e indique qué tipo de distribución lo modela adecuadamente.

(b) Genere 10000 realizaciones que simulen un lanzamiento de dos dados con el que se evalúe el evento de éxito. Realice un histograma y compárelo con la función de probabilidad teórica de acuerdo a la distribución propuesta en el ejercicio anterior.

(c) Suponga que ahora deseamos saber el puntaje total en n lanzamientos (es decir, la cantidad de éxitos por turno). Determine qué distribución es adecuada para modelar el puntaje obtenido por turno y defina la expresión de su función de probabilidad.

(d) Genere 10000 realizaciones simulando en cada una $n = 20$ lanzamientos para contabilizar el puntaje en cada turno (de acuerdo a la variable aleatoria propuesta en el punto anterior). Realice un histograma y compárelo con la función de probabilidad teórica para dicha variable.

Normas y material entregable

- **Informe**: El informe debe entregarse en formato PDF (**no se aceptarán otros formatos**) y con nombre: `TP1_GXX.PDF` (donde `XX` es el número de grupo). No debe agregarse código en el informe.
- **Código**: Los archivos de código utilizados deben ser en formato `.m` de Matlab/Octave (o alternativamente `.py` si usara lenguaje Python). El código debe incluirse junto al informe en un archivo ZIP (con mismo nombre que el informe) que deberá subirse al campus.