УДК 004:519.6
Самарцов Д.М студент гр. ПЗ2021
Науковий керівник Литвиненко К.В к.т.н. доцент кафедри «Комп’ютерні іформаційні технології» (Дніпровський національний університет залізничного транспорту ім. В. Лазаряна)

ВИКОРИСТАННЯ МЕТОДІВ МОНТЕ-КАРЛО У ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЯХ ТА АНАЛІЗІ ВЕЛИКИХ РОЗМІРІВ ДАНИХ

Методи Монте-Карло – група методів, заснованих на одержанні великої кількості реалізацій стохастичного випадкового процесу, який формується у той спосіб, щоб його ймовірнісні характеристики збігалися з аналогічними величинами задачі, яку потрібно розв'язати. Ці методи можна застосовувати для рішення задач у різни областях науки, наприклад у фізиці, хімії, математиці, інформаційних технологіях тощо.

Базове використання методів Монте-Карло можна показати на прикладі обчислення визначених інтегралів.

Для знаходження визначеного інтегралу $\int_a^b f(x)\,dx$.[1] Потрібно згенерувати деяку кількість пар випадкових чисел $a_i b_i$ де a_iрівномірно розподілена на відрізку ab, а b_i — $0Y$.

Для визначення площі фігури обмеженою функцією $f(x)$ потрібно порахувати кількість пар чисел, які відповідають нерівності

$$f(a_i) > b_i \tag{1}$$

Позначимо кількість пар, що відповідають нерівності (1) m, а кількість згенерованих пар — n. Оскільки пари чисел розділені рівномірно, то кількість точок, що потрапили у область обмеженою функцією $y = f(x)$, де вісі $x = a_i, y = b_i$, то відношення пар, що відповідають нерівності та не відповідають, будуть відноситись так само як і площа фігури до площі квадрату у який можуть потрапляти згенеровані числа. Тоді площу фігури обмеженою функцією $f(x)$ можна визначити за формулою

$$\int_a^b f(x)\,dx = \frac{m}{n} y(a-b)$$

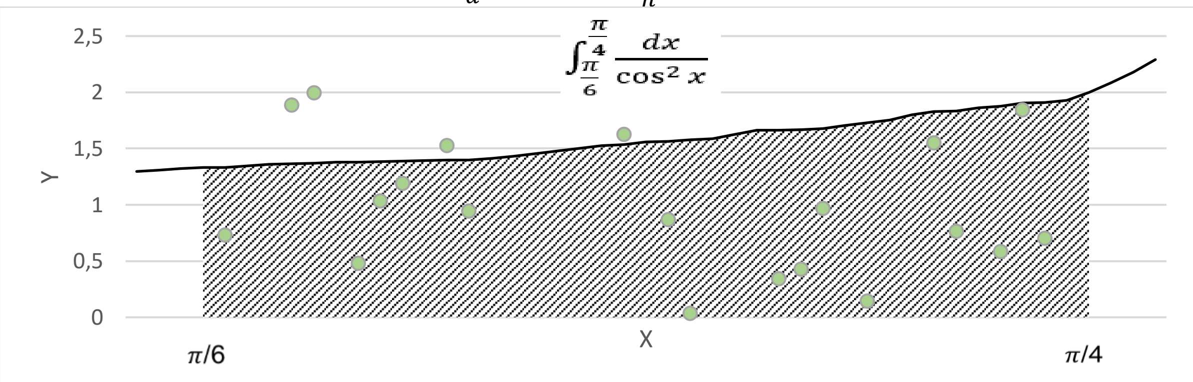

Рисунок 1 – знаходження визначеного інтегралу за допомогою методу Монте-Карло

Також методи Монте Карло можна використовувати для обчислення кратних інтегралів. [1]

Нехай функція $y = f(x_1, \dots, x_m)$ неперервна в обмеженій замкненій області G і потрібно обчислити m-кратний інтеграл I по області G

$$I = \int \cdots \int f(x_1, \dots, x_m) dx_1 \dots dx_m$$

Де I - (m + 1)-мірний обсяг вертикального циліндричного тіла в просторі $Ox_1x_2\dots x_my$, побудованого на основі G і обмеженого зверху даною поверхнею $y = f(x_1, \dots, x_m)$.

Перетворимо інтеграл так, щоб нова область інтегрування (Ω) цілком містилася всередині одиничного m-мірного куба.

Нехай область інтегрування G розташована в m-вимірному паралелепіпеді $a_k \le x_k \le b_k$, де $(k = 1, \dots, m)$.

Зробимо заміну змінних

$$x_k = a_k + (b_k - a_k)\xi_k$$

У цьому випадку m-мірний паралелепіпед перетворюється в m-мірний одиничний куб $0 \le \xi_k \le 1$ з $(k = 1, \dots, m)$, нова область інтегрування Ω буде належати цьому одиничному кубу.

$$\frac{D(x_1, \dots, x_m)}{D(\xi_1, \dots, \xi_m)} = \begin{vmatrix} b_1 - a_1 & 0 & \dots & 0 \\ 0 & b_2 - a_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_m - a_m \end{vmatrix} =$$

$$(b_1 - a_1) \cdot (b_2 - a_2) \cdots (b_m - a_m)$$

Таким чином

$$I = (b_1 - a_1) \cdot (b_2 - a_2) \cdot \ldots \cdot (b_m - a_m) \cdot J$$

Де

$$J = \int \cdots \int f[a_1 + (b_1 - a_1)\xi_1, \dots, a_m + (b_m - a_m)\xi_m] d\xi_1 \dots d\xi_m$$

Та область інтегрування Ω буде належати m-мірному одиничному кубу.

Нехай

$$F(\xi_1, \dots, \xi_m) = f[a_1 + (b_1 - a_1)\xi_1, \dots, a_m + (b_m - a_m)\xi_m]$$

Виберемо N рівномірно розподілених на відрізку [0, 1] випадкових послідовностей чисел:

$$\begin{matrix} \xi_1^{(1)}, & \xi_2^{(1)}, & \dots, & \xi_m^{(1)} \\ \xi_1^{(2)}, & \xi_2^{(2)}, & \dots, & \xi_m^{(2)} \\ \dots & \dots & \dots & \dots \\ \xi_1^{(N)}, & \xi_2^{(N)}, & \dots, & \xi_m^{(N)} \end{matrix}$$

Розглянемо випадкові точки $M_i\left(\xi_1^{(i)}, \xi_2^{(i)}, \dots, \xi_m^{(i)}\right)$. Вибравши досить велике N число точок $M_1, \dots, M_N$ перевіряємо, які з них належать області Ω. Нехай n точок належать області Ω.

Взявши велику кількість (n) точок M_i, $F_{\text{ср}}$ можна представити рівними

$$\tilde{F}_{\text{ср}} = \frac{1}{n} \sum_{i=1}^{n} F(M_i)$$

Тоді оцінку інтеграла J можна визначити за формулою

$$\tilde{J} = \tilde{F}_{\text{ср}} \cdot \Omega$$

Де Ω – m-мірний об’єм області інтегрування Ω. Якщо вичислити об’єм складно, то можна використовувати визначення геометричної вірогідності.

$$\tilde{\Omega} = \frac{n}{N}$$

Тоді ми можемо виразити оцінку інтеграла I:

$$\tilde{I} = \frac{V}{N} \sum_{i=1}^{n} F(M_i)$$

Методи Монте-Карло використовуються у аналізі великих об’ємів даних (Big Data та Data science). За їх допомогою можна вирішувати задачі: оцінки суми: завдання оцінки великих об’ємів даних, коли не можливо дістати данні з кожного елементу, але всі дані важливі для надання більш точної оцінки. Не доступні елементи будуть згенеровані випадково; відновлення зображень: за допомогою згенерованих даних можна замінити відсутні чи зламані частини зображення та накласти їх на реальне зображення. Метод відновлення зображення базується на алгоритмі заповнення матриці шляхом оптимізації обмеженої ядерної норми матриці [2].

Перелік посилань

1. Lapege, G.P. "VEGAS: An Adaptive Multi-dimensional Integration Program". / Г.П. Лапедж Cornell Preprint CLNS (1980) стр 80-447.

2. Cai, J.F., Cand`es, E.J., Shen, Z.: A singular value thresholding algorithm for matrix comple-tion. SIAM Journal on Optimization 20(4), 1956–1982 (2010)