Lecture 32:
Monte Carlo Method
&
Hypothesis Testing - Part I

Satyajit Thakor
IIT Mandi

Monte Carlo method

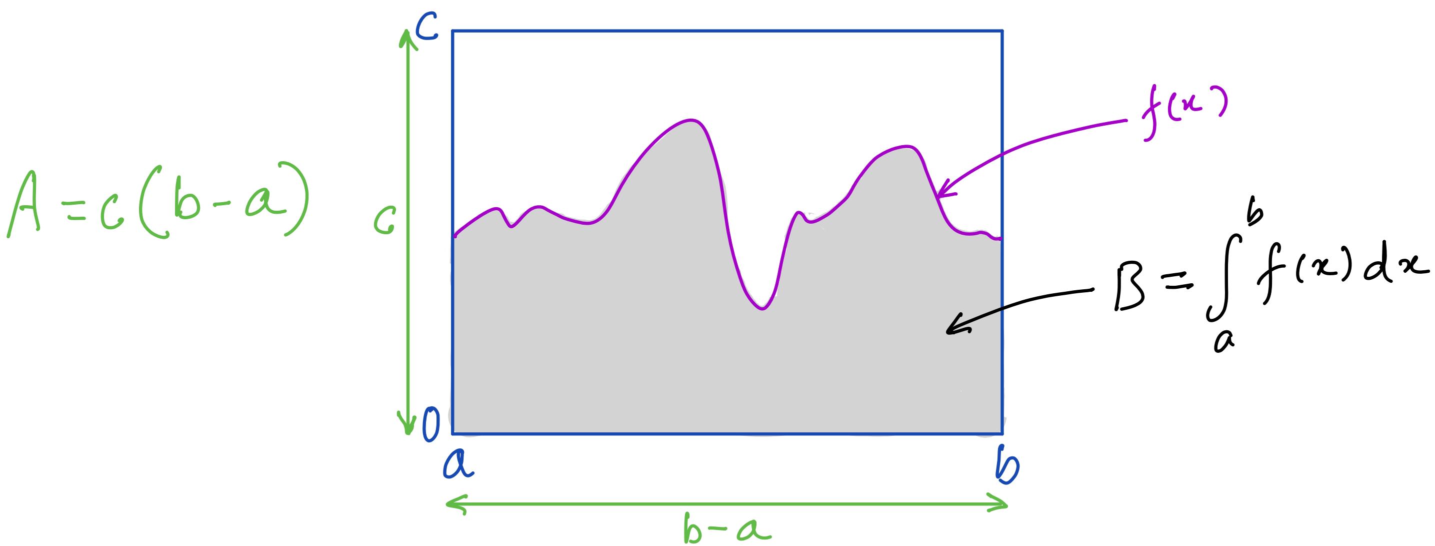
- Monte Carlo method is a set of algorithms which use randomness to solve problems either exactly or approximately.
- We will study Monte Carlo integration - a method used to approximate a definite integral using random point generation.
- Consider the problem of approximating the definite integral

$$\int_a^b f(x)dx$$

  where, f is some complicated function so that exact integration is not possible with any existing method.
- Assume that $0 \leq f(x) \leq c$ and hence the integral is finite.
- We can generate and use random numbers to approximate this integral!

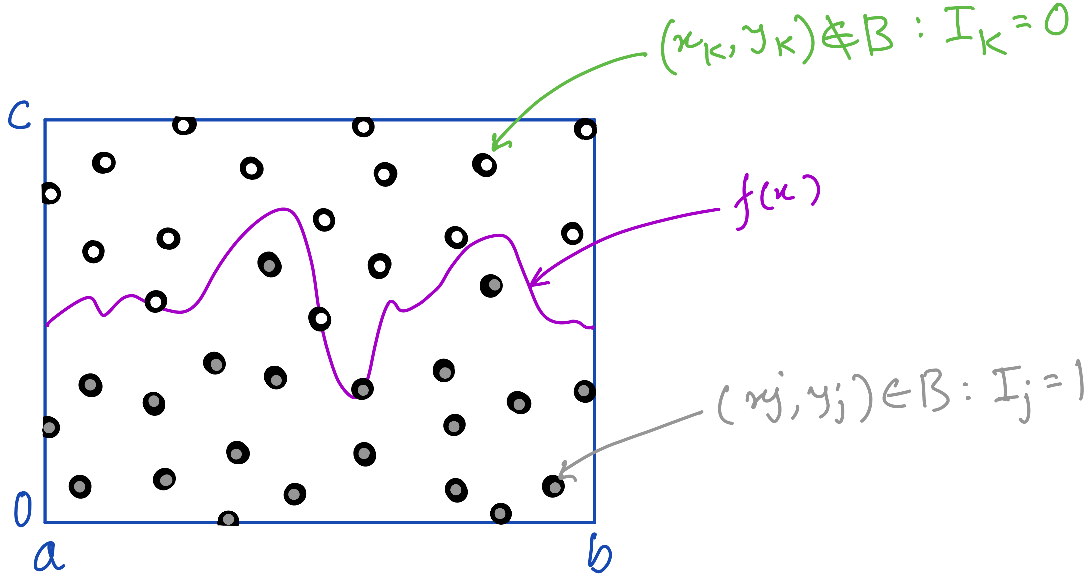
Monte Carlo method

- We can generate and use random numbers to approximate this integral!
- Let A be the area of the rectangle in the (x, y)-plane defined by $a \leq x \leq b$ and $0 \leq y \leq c$.
- Let B be the region under the curve $y = f(x)$ for $a \leq x \leq b$,
- Thus, the area of B is the desired integral.



Monte Carlo method

- Method: take random samples from A, then calculate the proportion of the samples that also fall into the area B.
- Generate iid points $(X_1, Y_1), \ldots, (X_n, Y_n)$ uniformly over A. Now let I_j be the Bernoulli r.v. such that $I_j = 1$ if $(X_j, Y_j) \in B$ and 0 otherwise.



Monte Carlo method

- Then

$$p = E(I_j) = P(I_j = 1) = \frac{B}{A} = \frac{\int_a^b f(x)dx}{c(b-a)}.$$

- By the WLLN, for large n we can approximate p as

$$\frac{1}{n}\sum_{j=1}^{n} I_j.$$

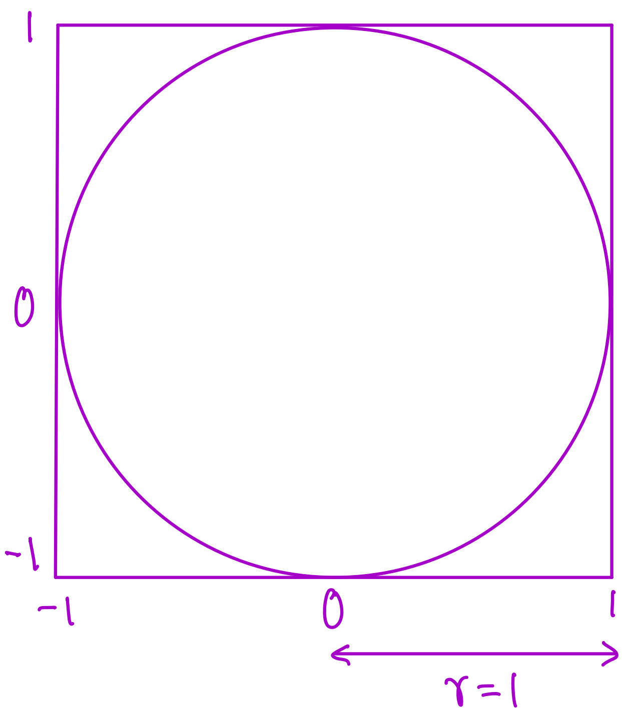
- Hence,

$$\frac{1}{n}\sum_{j=1}^{n} I_j \approx \frac{\int_a^b f(x)dx}{c(b-a)}$$

$$\Rightarrow \int_a^b f(x)dx \approx \frac{c(b-a)}{n}\sum_{j=1}^{n} I_j$$

Monte Carlo method

- Example: Approximate the value of the number π.



- Generate samples $(x_1, y_1), \ldots, (x_n, y_n)$ iid uniformly in the square.

- Let $I_j = \begin{cases} 1 & \text{if } x_j^2 + y_j^2 \leq 1 \\ 0 & \text{otherwise.} \end{cases}$

- Then, $\pi r^2 = \pi$

$$\approx \underbrace{4}_{\text{area of the square}} \cdot \frac{1}{n} \sum_{j=1}^{n} I_j$$

Hypothesis testing

- We, the scientists/engineers do not only come across problems of point/interval estimation in practice.
- Often, given a sample, we need to decide whether certain "statement" is true.

- For example, a medical researcher may decide on the basis of experimental evidence whether coffee drinking increases the risk of cancer in humans.
- Here "coffee drinking increases the risk of cancer in humans" is a statement or conjecture.
- A conjecture is a statement which a scientist proposes but correctness of which is yet to be established.
- A data scientist attempts to "test" correctness of the statement based on the available sample/evidence.

Hypothesis testing

- A statistical hypothesis is a conjecture concerning a population.
- Whether a hypothesis is true or false is never known with absolute certainty unless we examine the entire population or underlying distribution.

- We take a random sample from the population of interest and use the data contained in this sample to provide evidence that either supports or does not support the hypothesis.
- Evidence from the sample that is inconsistent with the stated hypothesis leads to a rejection or nullification of the hypothesis.

Hypothesis testing

- Another simple story to understand hypothesis testing: A person is arrested based on a suspicion of committing a crime.
- The court has the following hypothesis to test: The person is innocent.
- The goal of the court is to nullify this hypothesis based on evidence.
- In the court, evidence is presented and examined.
- Based on the evidence, the jury either fails to reject the hypothesis (the person is innocent) or rejects it (the person is guilty of the crime).
- Important thing to note:
- If the jury fails to reject the hypothesis based on the evidence presented, it does not imply that the person is innocent.
- It only implies that the evidence was insufficient to convict.