COMP 480/COMP 580 – Probabilistic Algorithms and Data Structures
Oct.17, 2023

Lecture 16

lecturer: Anshumali Shrivastava *Scribe by: Zitong Li, Jieming Luan, Ziang Ling*

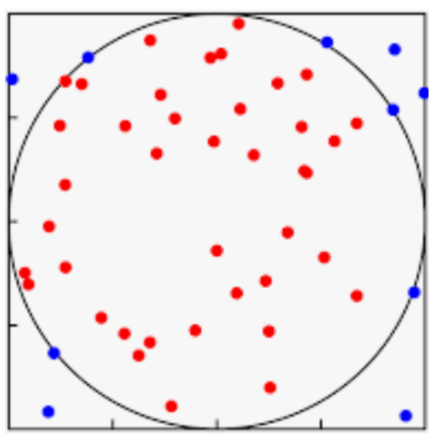
1. Monte Carlo Simulation

1.1 Definition

Monte Carlo Simulation is a computational approach that uses random sampling and probability to approximate the prediction of multiple results for a certain system or situation. When a sample is randomly chosen, then the sample will fall in one of the groups. The Monte Carlo Simulation would choose a large number of random samples to estimate the expected value based on the results of these groups.

1.2 Estimation for π

To estimate the value of π, we can use Monte Carlo Simulation. First, we would inscribe a circle within a square. Then, we generate a large number of random points within the square by selecting random coordinates. For each point, we need to check if it falls within the circle or not. Now we can get the number of points that fall inside the circle (a) and total number of points (A). The ratio of points inside the circle to the total number of points approximates the ratio of the quarter-circle's area (πR^2) to the square's area ($4R^2$), which is π/4.



As a result, the ratio μ = $\frac{a}{A}$ = $\frac{\pi}{4}$. The μ is the expected value we need to reach, and we can multiply this ratio by 4 to estimate the value of π.

1.3 Analysis

Let's say that μ represents probability that we randomly sample from the universe and the probability that points are inside our intertest region is called $\hat{\mu}$. Now, we have randomly sample N points and corresponding N indicators of points: $I_1, I_2, \ldots, I_N$. If the point is in the region, the indicator would be 1, otherwise, it would be 0.

Therefore, our estimator for this quantity is:

$$\frac{1}{N}\sum_{i=1}^{N} I_i = \hat{\mu}$$

Our goal is to estimate a or μ in this experiment and here we use an unbiased estimator. After we have computed its variance and standard confidence interval, we would get the probability about μ, which is:

$$P_r\left[|\widehat{\mu} - \mu| \geq \epsilon\mu\right] \leq \delta$$

$\epsilon\mu$ means epsilon distortion. So, if we want this quantity satisfy the statement we mentioned above, then the N need to satisfy this statement which is a probabilistic guarantee:

$$N \geq \frac{3\ln\frac{2}{\delta}}{\epsilon^2\mu}$$

The cost we are paying are N samples and certificates. Things like we generate N samples from the region A and then use N certificates to know whether it was accepted or not.

2. DNF Counting

2.1 Definition

Disjunctive Normal Form (DNF) is a logical formula as a disjunction (OR) of one or more conjunctions (AND) of variables. And the DNF counting is to demonstrate how will exponentially hard problem with the right design choice of you.

In a DNF formula, logical expression is conjunction of clauses, and each clause is a disjunction of xi, where xi is Boolean variables, such as $(x_7 \wedge x_2 \wedge \overline{x_3}) \vee (x_1 \wedge x_5 \wedge \overline{x_6}) \vee \ldots \vee (x_{n-2} \wedge x_n \wedge \overline{x_{n-1}})$. DNF counting seeks to count the distinct combinations of variable assignments that make the formula true.

2.2 DNF Properties

Given we have a DNF formula like $(x_1 \wedge x_2 \wedge x_3) \vee (x_4 \wedge x_5 \wedge x_6) \vee \ldots \vee (x_{n-2} \wedge x_{n-1} \wedge x_n)$. To simplify the statement, we can regard them as $C_1 \vee C_2 \vee \ldots \vee C_m$.

Let's say that SC_i is the set of all assignments that satisfy clause C_i. Then,

$$|SC_i| = 2^{n-k_i}$$

where k_i is the number of satisfying the clause. Now let's take the union of all SC_is, for I from 1 to m, which is

$$\bigcup_{i=1}^{m} SC_i = S$$

S means the set of assignments that satisfy all clauses. Then we can create a new set:

$$SC_i^{new} = \{(i, x) : x \in SC_i\}$$

And we can know that $|U| = \sum_{i=1}^{n} |SC_i| = \bigcup_{i=1}^{m} {SC_i}^{new}$. Because we may have a satisfying clause which can satisfy multiple clauses, so union U would be as big as m times of union S. Therefore, our cardinality has to be less than m times of S, which means:

$$|U| \leq m \cdot |S|$$

Now, what we want is to create a subset with the same cardinality as S. We hope the set is not too big and we can randomly sample from it. So we find this set which shows below:

$$\bar{\bar{S}} = \{(i, x) : (i, x) \in U \;\&\; i \text{ is the smallest one}\}$$

And we can say that $|\bar{\bar{S}}| = |S|$.