COMPUTATIONS IN CHEMICAL ENGINEERING

(BY: DIYA SARAF)

VECTORS, STACKS AND QUEUES

Vectors are like arrays which can resize themselves whenever an element is deleted or added. Since we don’t have to define their size when initializing them, they are helpful when we don’t know what the size of our output is going to be. Elements can be inserted in vectors using push_back() instead of specifying an index position for the element.

Stack is a data structure in which operations are performed in a one way order (last in first out). It can be visualised like a stack of plates where a new plate is placed on top and is also the first to be removed. Push and pop operations on a stack can be performed in constant time (O(1)).

Queues is a data structure in which insertion and deletion follow the principle of “First in, First out” (FIFO), where the first element added to the queue is the first one to be removed. A Queue is like a line waiting to purchase tickets, where the first person in line is the first person served.

LINKED LIST AND BINARY TREE

It is a data structure which consists of nodes connected together. Each node stores data and a reference (link) to the next node in the sequence. We need to define a new class for linked list and functions for all its operations. The address stores in the last node is ‘NULL’ since it does not point to any node. Unlike arrays, linked lists allow for efficient insertion or removal of elements from any position in the list.

Binary Tree is a non-linear data structure where each node has at most two children (next nodes connected to it). They are referred as left node and right node. Each node in a Binary Tree has three parts: data, pointer to left node and pointer to right node. Binary trees can be traversed in many different ways (in-order, pre order , post order and level order).

GRAPHS

Graph Data Structure is a collection of nodes connected by edges. It's used to represent relationships between different entities. Basically,
Linked list: one node to other
Binary tree: One node has two child node
Graph: Each node connected to atleast one node (can be connected to all nodes)

Some of the commonly used types of graphs are directed, undirected, weighted and unweighted. Graphs can be traversed in 2 ways: Breadth first search and dept first search. In both of them, a boolean visited array is used to mark the visited vertices to avoid processing a node more than once. Graphs can be used to model a wide variety of real-world problems.

RESEARCH PAPER-1

This research paper talks about Graph Theory Applied to Plasma Chemical Reaction Engineering. There are many applications of plasma in chemical engineering but, their lack of selectivity at atmospheric pressure is one of the main obstacles limiting the more widespread use of plasmas. schemes. The number of possible reactions in air alone can be in excess of 1800 reactions, with over 70 different species being present. This will only increase with the need to simulate plasma reactions involving more complex chemicals and mixtures.

The first step in forming a graph from a plasma chemical kinetic dataset involves organizing the data, which typically includes a list of reactions, their reactants and products, and the rate coefficients of these reactions. Only forward reactions were used and hence a directional graph was used.

Simulating the evolution of the concentrations of species over time consists of solving the rate equations for each reaction over discrete timesteps, summing the rates of creation and destruction reactions for each species, and then subtracting the latter from the former for each timestep. Rate coefficients can either be constants or functions of parameters like electron energy or temperature. Their change can be quantified by the diference between the rate coefficients at the maximum and minimum electron energies (within the energy range 1–10 eV) for DBD plasma.

Gephi, an open-source graph visualization tool, was used to create a directional weighted graph from a plasma chemical reaction system for humid air plasma. The thickness of each edge in the graph represents the rate of reaction, with the fastest rates having the thickest edges. Using the fltering functions of Gephi, information relevant to single or small numbers of species of interest can be viewed in increasing levels of complexity.

There may be a number of successive reactions between the plasma feed gas and a desired chemical product, and a number of different possible reaction pathways by which a product may be formed from a reactant. It is useful to know which one of these is the fastest. Even though the concentration of reactants change the rate of a reaction, it is the rate coefficients that determine which reaction pathway has the potential to be the fastest.

Dijkstra's algorithm can be used to find the shortest distance between two vertices on a graph. The path lengths are based on the rate coefficients of the reactions involved. Dijkstra's algorithm treats edges as weights. So, an edge with a weight of two counts as a path twice as long as an edge with a weight of one. Since the largest rate coefficients indicate the fastest reactions, they must be transformed to fit into Dijkstra's framework, where shorter paths represent faster reactions. This is achieved by taking the inverse of all rate coefficients (1/k). By this, a reaction with a low rate coefficient/slow reaction has a high weight. The resulting path lengths obtained are sorted in ascending order to find the most rapid reaction pathway.

This algorithm has certain limitations as well. It only compares the fastest reaction pathway (shortest path) between each two species. Clearly, even if all things are equal, a species may be formed more rapidly through a combination of two or more slower pathways, for example. Moreover, each reaction pathway can be affected differently by changes in conditions (temperature, electron energy, etc.), and this difference may not be apparent from a visual inspection of the connectivity matrix.

RANDOM VARIABLE

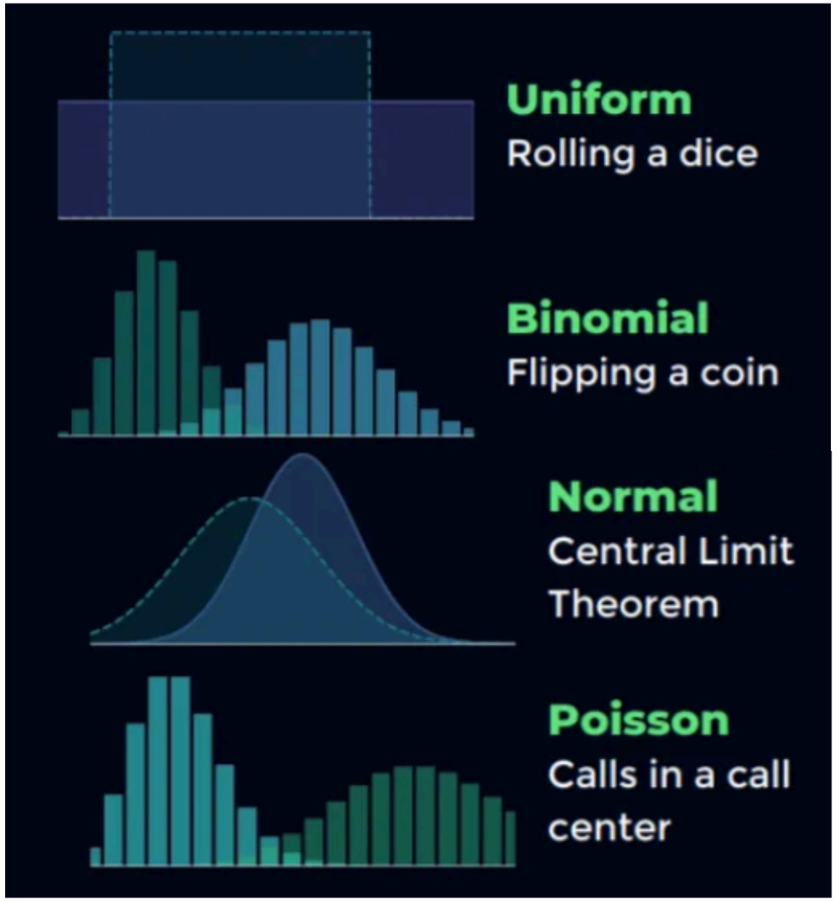
Random variable’ is ways (or a function) to map the outcome of random processes to numbers. For example, tossing three coins and defining random variable as the number of heads. Random variables have 2 types: (1) Discrete and (2) Continuous

A discrete variable is a variable which can take only can take only distinct values (for example tossing three coins and defining random ariable as the number of heads). A continuous random variable, on the other hand, can take on any value in some interval (for example the weight of a random person).

The expectation (mean), of a continuous random variable X provides a measure of the central tendency of the distribution. It is calculated using the probability density function (pdf). Variance measures the spread of the values of X around the mean. It is defined as the expected value of the squared deviation from the mean. The standard deviation is the square root of the variance and provides a measure of the average distance of the values of X from the mean.

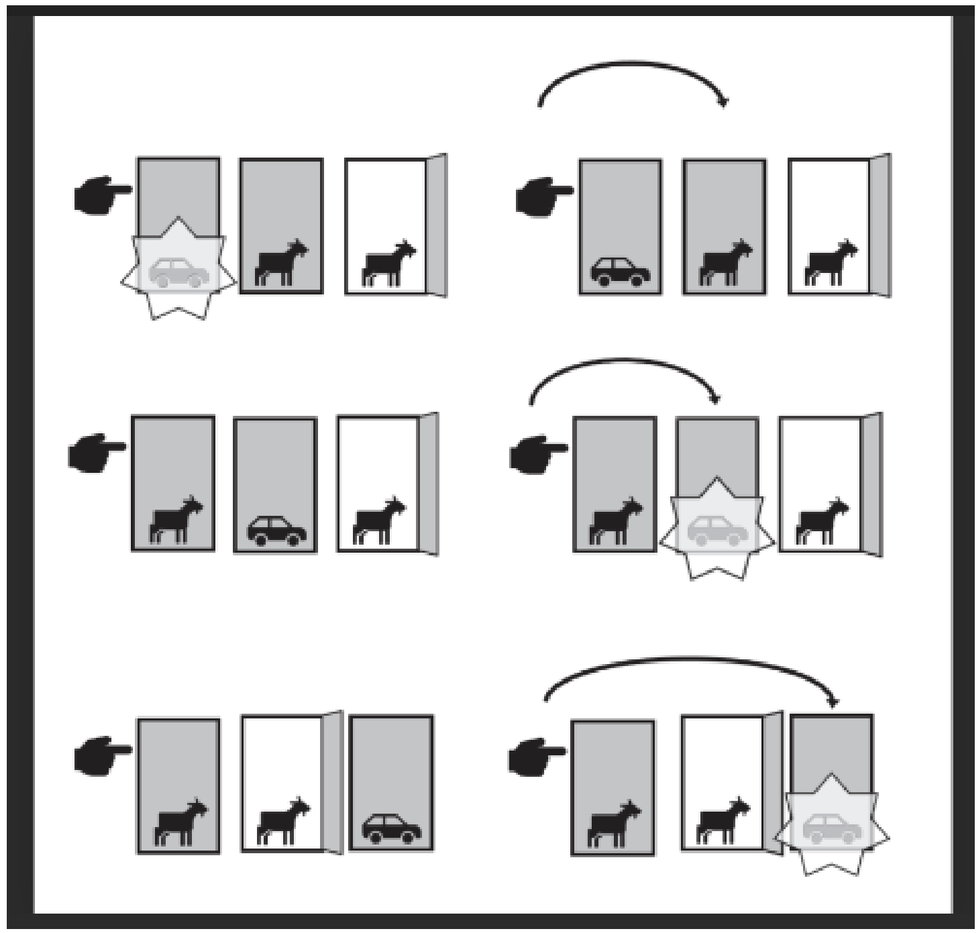
DISTRIBUTIONS

A distribution is simply a collection of data. It is usually represented in form of graph. Bernoulli, Binomial, Poisson, Uniform & Gaussian are some of the most common distributions. Bernoulli's distribution assumes that there is only one trial and only 2 outcomes. A binomial distribution is a collection of n Bernoulli trials assuming that the probability of each outcome remains the same throughout all the trials. Poisson distribution is a discrete distribution that models the probability of many events occurring in a fixed interval of time or space. A uniform distribution is a type of probability distribution in which all outcomes are equally likely. It describes how the values of a variable are distributed and is characterized by its bell-shaped curve. It is symmetric around its mean.



MONTE CARLO

Monte Carlo methods, or Monte Carlo experiments, are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. For example: The Monty Hall Problem. We basically simulate the Monty Hall game many times to estimate the probabilities of winning by switching versus not switching. Randomly select the winning door and the contestant's initial choice. Then, the door to be revealed is set as the door which is not the chosen door and not the door with a prize in it. After that, probability of winning on switching or not switching is found (after repeating this process 1000s of time using a for loop).



ACCEPT-REJECT RATIO

The Accept-Reject method is a Monte Carlo technique used to generate random samples from a target probability distribution f(x) when direct sampling is difficult. The acceptance-rejection method can be applied to discrete distributions to sample from a target distribution with probability mass function (pmf) {pj} using a proposal distribution with pmf {qj}. For continuous distribution, this method relies on a proposal distribution g(x) from which sampling is easier, and a constant c such that $f(x) \leq c \cdot g(x)$ for all x.
G and F should be around the same area and G should have fatter tails.
Steps/algorithm followed b y accept-reject method: Select a proposal distribution g(x). Generate a random sample X from the proposal distribution g(x). Generate a uniform random variable U from the interval [0,1]. Accept X as a sample from f(x) if $U \leq f(x)/(c \cdot g(X))$. Otherwise, reject X. Repeat the process until the desired number of samples is obtained.

RESEARCH PAPER-2

While solving complex equations can be extremely difficult or even impossible, the Monte Carlo methods devise discrete statistical “games” from which we can extract estimates.

The method requires these steps:
1. Defining a distribution of possible inputs for each input random variable.
2. Generating inputs randomly from those distributions.
3. Performing a deterministic computation using that set of inputs.
4. Aggregating the results of the individual computations into the final result

In an open distillation, the probability of a molecule to vaporize depends not only on the number of molecules of a certain substance, Mi , relative to the total number of molecules, but also on its relative volatility , i.e. how readily a substance vaporizes relative to the others. This research paper is about using stochastic monte carlo approach for multicomponent batch distillation.

We first assume initial mole fraction for all three components. Total molecules are taken as 10000. Mole fraction can be multiplied by total number of molecules to calculate number of molecules for each component. The algorithm is run m times (m=220). Array of relative volatility of each component is initialised. For each step (1 to m+1), firstly total molecules of each type are calculated. Then, for each type of molecule, their expectation is calculated by their number of molecules and relative volatility. Maximum expectation and ratio of expectation of each type of component with maximum expectation is then found.

The ratios are then compared with random number between 0 and 1. If ratio is greater than random number, vaporization of molecule of that component occurs and total molecules of that component are subtracted from 1. If ratio is smaller than random number, the number of molecules of that component remain same. Thus, the random number generator is actually used to compute the evolution of the number of molecules of each type in the liquid and so its composition with time (outputs), given the initial composition and relative volatilities of the substances (inputs). This is repeated n times (=38). The higher the n value, the more accurate the output will be. Every new step starts with the solution from the previous one i.e. for each step, after n times, mole fraction for each component is calculated. This continues m times and concentration of each component keeps on updating.

THANK YOU