CS648A : Randomized Algorithms
Semester II, 2020-21, CSE, IIT Kanpur

Programming Assignment 2

Deadline : 11:55 PM, 27 February 2022.

Most Important guidelines

- It is only through the assignments that one learns the most about the algorithms and data structures. You are advised to refrain from searching for a solution on the net or from a notebook or from other fellow students. Remember - **Before cheating the instructor, you are cheating yourself**. The onus of learning from a course lies first on you. So act wisely while working on this assignment.
- Refrain from collaborating with the students of other groups. If any evidence is found that confirms copying, the penalty will be very harsh. Refer to the website at the link: https://cse.iitk.ac.in/pages/AntiCheatingPolicy.html regarding the departmental policy on cheating.

General guidelines

1. This assignment is to be done in groups of 2 students. You have to form groups on your own. You are strongly advised not to work alone.
2. **Naming the file**:
   The submission file has to be given a name that reflects the information about the type of the assignment, the number of the assignment, and the roll numbers of the 2 students of the group. If you are submitting the solution of Programming Assignment x, you should name the file as **Prog_x_Rollnumber1_Rollnumber2.pdf**.
3. **Each student of a group** has to upload the same submission file separately. Be careful during the submission of an assignment. Once submitted, it can not be re-submitted.
4. Deadline is strict. Make sure you upload the assignment well in time to avoid last minute rush.

Random Graphs

Let n be any positive integer. Suppose we wish to study a probability space (Ω, P), where the sample space consists of all possible undirected graphs on n distinct vertices. How would you define the probability of each sample point (a graph) in this sample space ? Think for a few minutes before proceeding ...

A very simple and intuitive way to define the probability of each graph is to define the probability of each edge. Let p be any positive number ≤ 1. For each pair of vertices $u, v \in V$, we join them with an edge with probability p independent of any other pair of vertices. This process generates a *random graph* on n vertices. Notice that the graph resulting from this process can be any one of the $2^{\binom{n}{2}}$ graphs on n vertices. Moreover, the probability of any given graph on n vertices is also uniquely defined in the probability space (Ω, P). This model of random graphs is commonly called $G(n, p)$ model.

Random graphs have been extensively researched in mathematics and have been found to be of great applications in theoretical as well as applied computer science.

In this assignment, you will study the connectivity parameter of a random graph. In precise words, you have to analyse the expected size of the largest connected component in a random graph in $G(n, p)$ model for $n = 10^7$ and p varying from $[0, 2/n]$. Let $S_{n,p}$ be the expected size of the largest connected component in $G(n, p)$. The aim of this assignment is to explore the answer of the following question :

Question: How does $S_{n,p}$ vary when p increases from 0 to $2/n$ for a fixed n ?

Part 1

Before doing any experimentation and without searching on Google, what will be your natural guess for the answer to the above question ? State it and also write 2-3 sentences that form the basis of your guess.
Note: You will get full marks for this question if you attempt it irrespective of your guess and its justification. So there is no incentive for you to copy the answer from any source. Just answer the question honestly using your creative and analytical abilities.

Part 2

Write a computer program to evaluate $S_{n,p}$. Use this computer program to calculate $S_{n,p}$ when p increases from 0 to $2/n$ and $n = 10^7$?
Note: In this part, you need to plot $S_{n,p}$ versus p in the interval $[0, 2/n]$. For this purpose, you should carefully pick sufficient number of values in this interval and for each value p thus picked, you should run the computer program sufficiently large number of times to calculate $S_{n,p}$.

Part 3

Draw inference about the behavior of $S_{n,p}$ around $1/n$ from the plot you drew in Part 2. Write this inference in at most 4-5 sentences.

Part 4

Think of theoretical explanation for the behavior of $S_{n,p}$ versus p. There exist very messy, highly technical, and less intuitive ways to explain it. Interestingly, there is also a simple, formal, and intuitive explanation as well. Strive for that ... We might discuss it in some lecture in the course.
Note: This part needs not to be submitted. However, you are strongly encouraged to ponder over the explanation for the behavior of $S_{n,p}$ as a function of p.