CS203B: Mathematics for Computer Science - III
Semester I, 2018-19, CSE, IIT Kanpur

Assignment 2

Deadline : 6:00PM, 30 August

Important Guidelines:

1. It is only through the assignments that one learns the most in any course. You are advised to refrain from searching for a solution on the net or from a book or from other fellow students. If you are cheating the instructor, you are cheating yourself first. The onus of learning from a course lies first on you and then on the quality of teaching of the instructor. So act wisely while working on this and other assignments.

2. This assignment has 2 problems and solution of each problem has two components (1) programming, and (2) preparing report. The report has to be of single page only. Handwritten reports will not be accepted. You must type it using any word processor (For example : Latex, Microsoft Word, Google Doc). You will have to upload the code on moodle before the deadline.

1 Distributing clients among servers evenly

The aim of this problem is to make you realize the power or randomization to solve a very difficult problem in distributed environment.

There are n clients. There are n servers of identical computational powers. Each client has a job which it wants to be executed on any server. Each job will take unit time. The aim is to assign each client to a server such that all jobs are finished as soon as possible. This problem is quite trivial if we have a centralized authority that knows each client and server. This authority could map each client to a unique server in 1-1 mapping. As a result, all jobs will finish in unit time.

Unfortunately, there is no such centralized authority. Moreover, no client can communicate with any other client. Each client just knows the address of every server and only action it can do is to send its job to one of them. Each server may receive one or more jobs and it then executes them in some order.

Think over this problem for a while now to realize its nontriviality before proceeding further.

Now consider the following 2-step protocol followed by each client:

1. A client selects its server randomly uniformly and independent of other clients. Let it be y.
2. The client and sends its job to server y.

Let Z be the random variable for the maximum number of job received by any of the n servers.

The Objectives

There are 3 objective of this assignment. But only the 2nd objective will be graded.

Objective 1 Make an intelligent guess on the expected value of Z.

Objective 2 Write a computer programme to estimate the probability distribution of Z. Pick a sufficiently large value of n. Make sure that n is at least 10^7. Note that you will have to run the corresponding experiment sufficiently large number of times to estimate accurately the probability distribution of Z.

Objective 3 Ponder order to provide a justification for the experimental result you obtain.

The tasks of the assignment

The code

You have to design a simple, neat, and efficient program that estimates the probability distribution of Z. You will have to upload its code before the deadline of the assignment.

The report

Prepare a report of single page. The report should have only one plot that shows the distribution of Z. You may write any additional observation that you derived from the experiment. But be precise. It must not be more than a few sentences.

2 Spreading of fake news

Suppose there is a town with n persons. Each one of them is lazy but fond of spreading fake news. On some day, a person hears some fake news. The next morning, he/she uniformly randomly picks a phone number from the telephone directory, and communicate the fake news to the person with that number. So now there are 2 persons who know the fake news. As you can imagine, the fake news starts spreading according to the following protocol.

Every morning, each person who knows the fake news, picks a phone number uniformly randomly from the telephone directory, and communicate the fake news to the other person.

You may observe that there is a possibility that a person calls the same person multiple times. Moreover, a single person may get a call from multiple persons on the same day. Hence, every phone call *may not* necessarily increase the number of persons knowing the fake news. Let X be a random variable for the number of days it takes for each person of the town to know the fake news. It can be seen that for every m, however large it may be, $\mathbf{P}[X > m]$ is surely non-zero. It can also be established very easily that $\log_2\lceil n \rceil$ is the least number of days to spread the fake-news to the entire town. We are interested in the probability distribution of the random variable X.

The Objectives

There are 3 objective of this assignment. But only the 2nd objective will be graded.

Objective 1 Make an intelligent guess on the expected number of days it will take to spread the fake news.

Objective 2 Write a computer programme to estimate the probability distribution of X. Pick a sufficiently large value of n. Make sure that n is at least 10^6. Note that you will have to run the corresponding experiment sufficiently large number of times to estimate accurately the probability distribution of X.

Objective 3 Ponder order to provide a justification for the experimental result you obtain.

2. The tasks of the assignment

The code

You have to design a simple, neat, and efficient program that estimates the probability distribution of X. You will have to upload its code before the deadline of the assignment.

The report

Prepare a report of single page. The report should have only one plot that shows the distribution of X. You may write any additional observation that you derived from the experiment. But be precise. It must not be more than a few sentences.

Hints and directions

1. The code for each problem will require a single array and a counter. It will be extremely simple.
2. Make sure to repeat the experiment sufficiently large number of times to ensure that you get an accurate estimate of the distribution of X and Z. For this you should complete the program at least 2-3 days before the deadline.