(CS1.406) Advanced Algorithms (Spring 2022) Due: 30.01.2022

Problem Set 2

Instructions:

- Discussions amongst the students are not discouraged, but all writeups must be done individually and must include names of all collaborators.
- Referring sources other than the lecture notes is discouraged as solutions to some of the problems can be found easily via a web search. But if you do use an outside source (eg., text books, other lecture notes, any material available online), do mention the same in your writeup. This will not affect your grades. However dishonesty of any sort when caught shall be heavily penalized.
- Be clear in your arguments. Vague arguments shall not be given full credit.

1. A function $f : \mathbb{R} \mapsto \mathbb{R}$ is said to be convex if for any x_1, x_2 and $0 \leqslant \lambda \leqslant 1$, the following inequality is satisfied.

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda \cdot f(x_1) + (1-\lambda) \cdot f(x_2).$$

Let Z be a random variable that assumes values in the interval $[0,1]$ and let $p = \mathbb{E}[Z]$. Define a Bernoulli random variable X such that $\mathbb{P}[X = 1] = p$ and $\mathbb{P}[X = 0] = 1 - p$. Then show that for any convex function f, $\mathbb{E}[f(Z)] \leqslant \mathbb{E}[f(X)]$.

2. Let B be a random bipartite graph on two independent sets of vertices U and V, each with n vertices. For each pair of vertices $u \in U$ and $v \in V$, the probability that the edge (u, v) between them is present is $p(n) = \frac{\ln n + c}{n}$ (for some $c \in \mathbb{R}$) and the presence of any edge is independent of all other edges. Show that the probability that B contains an isolated vertex is asymptotically equal to $e^{-2e^{-c}}$.

3. (a) Let S and T be two disjoint subsets of a universe U such that $S = T = n$. Suppose we select a random set $R \subseteq U$ with a probability of p. We say that the random sample R is good if the following conditions hold – $R \cap S =\neq$ and $R \cap T \neq \emptyset$. Show that for $p = 1/n$, the probability that R is good is larger than some positive constant.

   (b) Suppose now that the random set R is chosen by sampling elements of U with only pairwise independence. Show that for a suitable value of p, the probability that R is good is larger than some positive constant.

4. A simple model of stock market suggests that each day a stock with price q will increase by a factor of $r > 1$ to qr with a probability of p and will fall to q/r with a probability of $1 - p$. Assuming that the starting price of the stock is 10 INR,

   (a) what is the expected value of the price of stock after d days?

   (b) what is the variance of the price of the stock after d days?

5. The weak law of large numbers states that if $X_1, X_2, X_3, \ldots$ are independent and identically distributed random variables with mean μ and standard deviation σ, then for any constant $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left[\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right] = 0.$$

Use Chebyshev's inequality to prove the weak law of large numbers.

6. In an industrial facility with n workers, COVID tests have to be performed for everybody. If each worker is tested separately, it could be expensive. However, pooling can decrease costs. One way to do pool testing is to pool the samples of a group of k workers and analyzed together – if the test comes back negative, then one test suffices for k workers, else we could test all the k workers separately and thus we would need a total of $k + 1$ tests for k workers in such a case. Suppose each worker tests positive (independently) with a probability of at most p (for a very small p).

   (a) What is the probability that the test for a pooled sample of k workers will be positive?

   (b) What is the expected number of tests necessary?

   (c) Describe how to find the best value of k.

   (d) For what values of p is pooling better than just testing every worker.