COL866: Special Topics in Algorithms
Concentration inequalities and their Applications in Computer Science
II semester, 2022-23.
Minor 1 exam.
30 January 2023, Maximum Marks: 25.

Instructions: Please handwrite your solutions clearly and concisely. Try to fit the solution for each problem in one side on an A4 sized page, definitely no more than 2. Latexed solutions are highly encouraged. Upload the paper on gradescope by the deadline (no extensions will be given under any circumstances). Make sure you mark the pages related to each problem on gradescope.

Problem 1 (6 marks)

The *randomized quicksort algorithm* works as follows: Given an array A of size n, it picks an index r at random from 0 to $n-1$. It then "pivots" the array around $A[r]$, i.e., it reorganises the array such that if $A[r]$ is moved to location i_r, then $A[j] \leq A[i_r]$ for all $j < i_r$ and $A[j] > A[i_r]$ for all $j > i_r$. The subarray $A[0, \ldots, i_r - 1]$ and $A[i_r + 1, \ldots, n-1]$ are recursively sorted. An array or size 1 is considered sorted. Show that there is a $c > 0$ such that the running time of randomized quicksort is at most $cn \log n$ with probability at least $1 - 1/n$. You may assume that the time taken for the pivot operation on an array of size k is $\theta(k)$. To avoid pathological corner cases you can assume that all the elements of the array are distinct.

Hint: Try to bound the upper tail of the height of the recursion tree by distinguishing between "good" choices of pivots and "bad" ones.

Problem 2 (9 marks)

Recall that, for an integer $n \geq 2$ and $p \in (0, 1)$, the Erdös-Renyi (ER) random graph $G_{n,p}$ is defined as follows: It has n vertices $v_1, \ldots, v_n$ and for each $i \neq j$, the edge (v_i, v_j) is present with probability p independent of all other edges. Recall also that given a graph $G = (V, E)$, a set of vertices $U \subseteq V$ is called an *independent set* if there is no edge in E between the vertices in U. Let $I_{n,p}$ be the (random) size of the largest independent set in $G_{n,p}$. Prove that there are constants c_1 and c_2 (that may depend on p) such that $c_1 \log n \leq I_{n,p} \leq c_2 \log n$ with probability $1 - o(1)$.

Problem 3 (10 marks)

We throw n balls into n bins independently and uniformly at random. Then we remove all the balls that are alone in a bin. We throw the remaining balls again. The process continues like this in rounds till all the balls are removed. Prove that the number of rounds in which the process ends in at most $c \log \log n$ rounds with probability at least $1 - o(1)$. What can we say about a lower bound for the number or rounds?