COL866: Special Topics in Algorithms
Concentration inequalities and their Applications in Computer Science
II semester, 2022-23.
Minor 2 exam.
26 February 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 ($2 + 9 + 3 = 14$ **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 $V = \{v_1, \ldots, v_n\}$ and for each $i \neq j$, the edge (v_i, v_j) is present with probability p independent of all other edges. In L12 we showed that $G_{n,p}$ is connected with probability at least $1 - o(1)$ whenever p is $\omega(\log n/n)$. We now assume that time is discrete and consider a temporal version of $G_{n,p}$:

> Given an integer $L > 0$ and $\delta \in (0, 1)$, each vertex $v_i \in V$ independently chooses $T_i \subseteq \{0, \ldots, L - 1\}$ such that $|T_i| = \lceil \delta L \rceil$. For each $j \in \{0, \ldots, L - 1\}$ we say $V_j = \{v_i : j \in T_i\}$. Each pair of vertices $u, v \in V_j$ is connected at time j with probability p independent of all other pairs.

We refer to the above model as $G_{n,p,\delta,L}$. V_j is the set of vertices that are "awake" in time slot j. In each time slot some subset of vertices is awake and forms an ER random graph of its own with parameter p. Although above we have written it as if there are only L time slots, we can assume that the graph created at any time slot $j \in \{0, \ldots, L - 1\}$ is repeatedly created for each j' such that $(j' - j)$ mod $L = 0$.

We say that $G_{n,p,\delta,L}$ is *temporally connected* if for any pair of vertices $v_{\ell_1}, v_{\ell_2} \in V$ there is a sequence of vertices $v_{\ell_1} = u_0, u_1, \ldots, u_k = v_{\ell_2}$ such that for every $0 \leq i < k$, there is some $j \in \{0, \ldots, L - 1\}$ such that u_i and u_{i+1} are both in V_j and there is a path connecting them in the subgraph formed at time j.

Problem 1.1 (2 **marks**)
Argue that when $\delta > 1/2$ then $G_{n,p,\delta,L}$ is temporally connected with probability at least $1 - o(1)$ whenever $p = \omega(\log n/n)$.

Problem 1.2 (9 **marks**)
What is the threshold value $f(\delta, n)$ such that $G_{n,p,\delta,L}$ is temporally connected with probability $1 - o(1)$ when $\delta \in (0, 1/2)$? Prove your answer rigorously and completely. Use Chernoff bounds where required.

Problem 1.3 (3 **marks**)
Comment on the lowest value of δ for which it is possible to achieve temporal connectivity with probability $1 - o(1)$. What can we say about a value of δ for which temporal disconnectedness is achieved with probability $1 - o(1)$? Note that this subproblem you need not write a full rigorous proof but at least try and make some arguments.

Problem 2 (6 **marks**)
Let X be a non-negative random variable with finite second moment. Show that $\mathrm{E}\left[e^{-\lambda(X - \mathrm{E}[X]}\right] \leq e^{\lambda^2 \mathrm{E}[X^2]/2}$ for all $\lambda > 0$. Further, if $X_1, \ldots, X_n$ are independent non-negative random variables with

finite second moments, if $S = \sum_{i=1}^{n}(X_i - \mathrm{E}[X_i])$ and $\nu = \sum_{i=1}^{n} \mathrm{E}\left[X_i^2\right]$, argue that for any $t > 0$,

$$\Pr\{S \le -t\} \le \exp\left(\frac{-t^2}{2\nu}\right).$$

Now derive a version of the second moment method using this inequality. Is this stronger than the second moment method derived from the Paley-Zygmund Inequality? Write an answer to this question using an illustrative example. Keep in mind the answer doesn't have to be a simple Yes/No answer.

Problem 3 (5 marks)
For some $p \in (0,1)$, let X be a random variable that takes value 0 with probability $1-p$, 1 with probability $p/2$ and -1 with probability $p/2$. Given t i.i.d. random variables $X_1, \ldots, X_t$ distributed as X, $S_t = \sum_{i=1}^{t} X_i$ gives us the position of the "lazy" random walk on $\mathbb{Z}$ at time t if it started at 0. We are interested in characterising the drift of the walk away from the origin. Specifically, given some $\varepsilon \in (0,1]$ and integer $t > 0$, we are interested in the value n such that $\Pr\{|S_t| \ge n\} \le \varepsilon$. What is the smallest value of n that Hoeffding's bound can give us? Compare this with the smallest value that Bernstein's inequality (the simple version that follows from Bennett's inequality) can give us. Under what condition on p is there an asymptotic difference between the two?