Problem Set 4

Due in class, Thursday, March 29

Instructions

- You should think about *each* problem by yourself for *at least 30 minutes* before choosing to collaborate with others.
- You are allowed to collaborate with fellow students taking the class in solving the problems (in groups of at most 3 people for each problem). In fact this is encouraged so that you interact with and learn from each other. However, *you must write up your solutions on your own.* If you collaborate in solving problems, you should clearly acknowledge your collaborators for each problem.
- Reference to any external material besides the course text and material covered in lecture is not allowed. In particular, you are not allowed to search for answers or hints on the web. You are encouraged to contact the instructors or the TA for a possible hint if you feel stuck on a problem and require some assistance.
- Solutions typeset in LaTeX are preferred.
- Feel free to email the instructors or the TA if you have any questions or would like any clarifications about the problems.
- You are urged to start work on the problem set early.

---

1. (10 points) Let N be a non-negative integer-valued random variable. Prove that

$$\Pr[N > 0] \geq \frac{\mathbb{E}[N]^2}{\mathbb{E}[N^2]} \ .$$

(Hint: One approach is to use Cauchy-Schwarz inequality, $\mathbb{E}[NB]^2 \leq \mathbb{E}[N^2]\mathbb{E}[B^2]$, applied to a suitable choice of random variable B)

2. (20 points) A clique in a graph is a set of vertices which are all adjacent to each other. In this exercise (based on Exercise 3.54 of the notes), we will study cliques in $G(n, 1/2)$ random graphs.

   (a) Prove that with high probability, a $G(n, 1/2)$ random graph has no clique of size $(2+\epsilon)\log_2 n$ for any constant $\epsilon > 0$.

   (b) Prove that with high probability, a $G(n, 1/2)$ random graph has a clique of size $(2-\epsilon)\log_2 n$ for any constant $\epsilon > 0$. (You may use **without proof** that for $k = \Theta(\log n)$, the function

$$f(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}} \ ,$$

   achieves its the maximum value in the range $2 \leq i < k$ for $i = 2$.)
   Hint: Express the number N of cliques of size k as a sum of $\binom{n}{k}$ indicator random variables, and then use the second moment method.

(c) Show that for any $\epsilon > 0$, a clique of size $(2-\epsilon)\log_2 n$ can be found in a $G(n, 1/2)$ random graph (with high probability) in time $n^{O(\log n)}$.

(d) Given an $O(n^2)$ time algorithm for finding a clique of size $\Omega(\log n)$ in a $G(n, 1/2)$ random graph (with high probability over the choice of the graph). Hint: Use a greedy algorithm.

3. (10 points) (Based on exercise 3.29 in the book) Consider a model of a random subset $N(n,p)$ of positive integers $\{1, 2, \ldots, n\}$ where $N(n,p)$ is the set obtained by independent at random including each $i \in \{1, 2, \ldots, n\}$ in the set with probability p. Similarly to graphs, one can define a notion of "increasing" properties of $N(n,p)$ and argue that every increasing property of $N(n,p)$ has a (weak) threshold.

   (a) What is the asymptotic threshold value of p for $N(n,p)$ to contain an even number? Prove your answer.

   (b) What's your guess for the asymptotic threshold value of p for $N(n,p)$ to contain three numbers x, y, z such that $x + y = z$? Briefly explain the reasoning for your particular choice (formal proof not required).

4. (15 points) Let $\mathsf{NAE} : \{0,1\}^2 \to \{0,1\}$ be the binary constraint on Boolean variables given by:

$$\mathsf{NAE}(x,y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

   Consider a random instance $\mathcal{C}(n,m)$ of a satisfiability problem on n variables $\{x_1, x_2, \ldots, x_n\}$ with m NAE constraints, where each constraint is picked independently at random from among the $\binom{n}{2}2^2$ possibilities (for choosing two variables, and the associated literals). Example constraints are $\mathsf{NAE}(x_1, x_2)$, $\mathsf{NAE}(x_3, \overline{x_5})$, etc.

   Let $\epsilon > 0$ be an arbitrary constant.

   (a) What is the expected number of satisfying assignments N to a random instance of $\mathcal{C}(n,m)$?

   (b) Prove that when $m = (1+\epsilon)n$, the probability that $\mathcal{C}(n,m)$ is satisfiable tends to 0 as $n \to \infty$.

   (c) Fix two assignments $a, b \in \{0,1\}^n$ that differ in k coordinates ($0 \leq k \leq n$). For a random NAE constraint, what is the probability that *both* a, b satisfy it?

   (d) Use the above to compute an expression for the second moment $\mathbb{E}[N^2]$ of the number of satisfying assignments of a random instance $\mathcal{C}(n,m)$. (You do not need to simplify your answer.)

   (e) **(Extra credit)** Prove that when $m = (1-\epsilon)n$, the probability that $\mathcal{C}(n,m)$ is satisfiable is at least a constant as $n \to \infty$.

5. (15 points) Let D be a random directed graph on n vertices in which for each vertex u, we choose a vertex v at random and add the directed edge $u \to v$ to the graph. (We could have $v = u$ in which case there is a loop at u.) For a vertex v, let $r(v)$ be the number of vertices reachable from v by a directed path in D.

   (a) For each $k \in \{1, 2, \ldots, n\}$, what is the probability that $r(v) = k$?

   (b) Prove that for a fixed vertex v, the probability that $r(v) \leq \sqrt{n}/10$ is at most $1/3$, and the probability that $r(v) \geq 10\sqrt{n}$ is also at most $1/3$.