Problem Set 4 Solution

by Yuan Zhou

1. Let B be the indicator variable for the event $N > 0$, i.e. $B = 1$ when $N > 0$ and $B = 0$ when $N = 0$. It is easy to check that $NB = N$. By Cauchy-Schwartz, we have

$$\mathbf{E}[N]^2 = \mathbf{E}[NB]^2 \leq \mathbf{E}[N^2]\,\mathbf{E}[B^2],$$

which implies that

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

2. (a) Fix a set of $(2+\epsilon)\log_2 n$ vertices, the probability that there is an edge between each two of them is at most

$$2^{-((2+\epsilon)\log_2 n)((2+\epsilon)\log_2 n - 1)/2} \leq 2^{-(2+\epsilon)\log_2^2 n}.$$

Since there are $\binom{n}{(2+\epsilon)\log_2 n} \leq \frac{2^{(2+\epsilon)\log_2^2 n}}{(2\log_2 n)!}$ such set of vertices. By a union bound, the probability that there exists a clique of size $(2+\epsilon)\log_2 n$ is at most

$$2^{-(2+\epsilon)\log_2^2 n} \cdot \frac{2^{(2+\epsilon)\log_2^2 n}}{(2\log_2 n)!} \leq \frac{1}{(2\log_2 n)!} = n^{-w(1)}.$$

Therefore, with probability $(1 - n^{-w(1)})$, there is no clique of size $(2+\epsilon)\log_2 n$.

(b) Set $k = (2-\epsilon)\log_2 n$ for notational convenience. Let N_i be the indicator variable for the event that the i-th subset of size k forms a clique. $(1 \leq i \leq \binom{n}{k})$. Let $N = \sum_i N_i$ be the number of cliques of size k. We also use the notation S_i to denote the i-th subset. By Problem 1, we know that

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

We are going to lower bound $\Pr[N > 0]$ by estimating calculating $\frac{\mathbf{E}[N]^2}{\mathbf{E}[N^2]}$.
For $\mathbf{E}[N]^2$, we have

$$\mathbf{E}[N]^2 = \left(\sum_i \mathbf{E}[N_i]\right)^2 = \left(\binom{n}{k} 2^{-k(k-1)/2}\right)^2 = \binom{n}{k}^2 2^{-k(k-1)}.$$

For $E[N^2]$ we have

$$\mathbf{E}[N^2] = \sum_{i,j} \mathbf{E}[N_i N_j] = \sum_i \sum_{t=0}^{k} \sum_{j:|S_i \cap S_j| = t} \mathbf{E}[N_i N_j].$$

Now, let $f(t) = \sum_{j:|S_i \cap S_j|=t} \mathbf{E}[N_i N_j]$ so that $\mathbf{E}[N^2] = \binom{n}{k} \sum_{t=0}^{k} f(t)$. We see that

$$f(t) = \binom{k}{t}\binom{n-k}{k-t} 2^{-k(k-1)+t(t-1)/2}.$$

In all, we have

$$\frac{\mathbf{E}[N]^2}{\mathbf{E}[N^2]} = \frac{\binom{n}{k}^2 2^{-k(k-1)}}{\binom{n}{k}\sum_{t=0}^{k}\binom{k}{t}\binom{n-k}{k-t}2^{-k(k-1)+t(t-1)/2}} = \frac{\binom{n}{k}}{\sum_{t=0}^{k}\binom{k}{t}\binom{n-k}{k-t}2^{t(t-1)/2}}.$$

Therefore,

$$\begin{aligned}
&1 - \frac{\mathbf{E}[N]^2}{\mathbf{E}[N^2]} \\
=&\frac{\sum_{t=2}^{k}\binom{k}{t}\binom{n-k}{k-t}\left(2^{t(t-1)/2}-1\right)}{\sum_{t=0}^{k}\binom{k}{t}\binom{n-k}{k-t}2^{t(t-1)/2}} \qquad \left(\text{since } \binom{n}{k} = \sum_{t=0}^{k}\binom{k}{t}\binom{n-k}{k-t}\right) \\
\le&\frac{\sum_{t=2}^{k}\binom{k}{t}\binom{n-k}{k-t}2^{t(t-1)/2}}{\sum_{t=0}^{k}\binom{k}{t}\binom{n-k}{k-t}2^{t(t-1)/2}} \\
\le&\frac{\sum_{t=2}^{k}\binom{k}{t}\binom{n-k}{k-t}2^{t(t-1)/2}}{\binom{n}{k}}. \qquad \left(\text{again, by } \binom{n}{k} = \sum_{t=0}^{k}\binom{k}{t}\binom{n-k}{k-t}\right) \\
\le&k \cdot \left.\frac{\binom{k}{t}\binom{n-k}{k-t}2^{t(t-1)/2}}{\binom{n}{k}}\right|_{t=2} \\
&\text{(by the fact given in the problem statement, for } k < (2-\epsilon)\log_2 n) \\
\le&k \cdot \frac{k^4}{n^2} = o(1).
\end{aligned}$$

(c) Enumerate over all subsets of size $(2-\epsilon)\log_2 n$ and check whether any of them forms a clique.

(d) We start with a set $S = \emptyset$ and $T = [n]$. We want to add vertices to S while keeping S a clique as the algorithm proceeds, and finally output S. T is to be maintained as the common neighbors of vertices in S. The algorithm repeats the following procedure, as long as $T \neq \emptyset$: choose an arbitrary vertex $i \in T$, move it to S, and remove all the vertices in T that are not connected to i. When this produce terminates, we see that S is always a clique.

It is easy to see that the algorithm runs in $O(n^2)$ time for any graph. Now we are going to show that with high probability, S contains $\Omega(\log n)$ vertices.

By a Chernoff bound, we see that for any $|T| \ge \sqrt{n}$, and any $i \in T$, the probability that i is connected to $|T|/3$ vertices in T is at least $1 - \exp(-\sqrt{n}/100)$. Therefore, by a union bound, the probability that the algorithm proceeds for at least $\log_3 \sqrt{n}$ steps is at least $1 - (\log_3 \sqrt{n})\exp(-\sqrt{n}/100) \ge 1 - \exp(-\sqrt{n}/200)$. This also lower bounds the probability that $|S| \ge \log_3 \sqrt{n}$.

3. (a) The weak threshold is $\Theta(1/n)$. The probability that $N(n,p)$ contains an even number is $1-(1-p)^{n/2}$. We see that when $p = o(1/n)$, this probability becomes $1 - \exp(-o(1)) = o(1)$; when $p = \omega(1/n)$, this probability becomes $1 - \exp(-\omega(1)) = 1 - o(1)$.

(b) Since there are $\Theta(n^2)$ triples (x, y, z) such that $x + y = z$, while each of these triples appears in $N(n, p)$ with probability p^3, the expected number of these triples that appear in $N(n, p)$ is $\Theta(n^2/p^3)$. For this number to be a constant, we need $p^3 = \Theta(1/n^2)$. Therefore, a reasonable guess for the threshold value would be $\Theta(n^{-2/3})$.

4. (a) Each assignment satisfies a random instance (with m) clauses with probability 2^{-m}. Since there are 2^n assignments, the expected number of satisfying assignments is 2^{n-m}.

   (b) By Markov inequality, we have

$$\begin{aligned}\Pr[\mathcal{C}(n,m) \text{ satisfiable}] &= \Pr[\#\text{satisfying assignment for } \mathcal{C}(n,m) \geq 1] \\ &\leq \mathbf{E}[\#\text{satisfying assignment for } \mathcal{C}(n,m)] = 2^{n-m} = 2^{-\epsilon n} = o(1).\end{aligned}$$

   (c) When the two variables are both from the k variables (where a and b are differ), or both from the rest $n-k$ variables, the probability that both a and b satisfy the random clause is $1/2$. In other cases, a and b cannot simultaneously satisfy the random clause. Therefore, the overall probability is

$$\frac{1}{2} \cdot \frac{\binom{k}{2} + \binom{n-k}{2}}{\binom{n}{2}} = \frac{k(k-1) + (n-k)(n-k-1)}{2n(n-1)}.$$

   (d) Let N_i be the indicator variable that the i-th assignment satisfies the random assignment (for $1 \leq i \leq 2^n$). We have

$$\begin{aligned}\mathbf{E}[N^2] &= \sum_i \sum_j \mathbf{E}[N_i N_j] \\ &= \sum_i \sum_{k=0}^{n} \sum_{j:\text{Assignment } j \text{ and Assignment } i \text{ differ at } k \text{ places}} \left(\frac{k(k-1) + (n-k)(n-k-1)}{2n(n-1)}\right)^m \\ &= \sum_i \sum_{k=0}^{n} \binom{n}{k} \cdot \left(\frac{k(k-1) + (n-k)(n-k-1)}{2n(n-1)}\right)^m \\ &= \sum_{k=0}^{n} 2^n \binom{n}{k} \cdot \left(\frac{k(k-1) + (n-k)(n-k-1)}{2n(n-1)}\right)^m.\end{aligned}$$

   (e) (Proof sketch.) Given an NAE instance, we construct a corresponding *constraint graph* by creating a vertex for each variable, and connecting two vertices by an undirected edge when the two corresponding variables are involved in a same NAE clause. Now our proof goes by two steps.

   Step 1. We are going to prove that when $m = (1-\epsilon)n$. the constraint graph has only $O(1)$ cycles with probability at least $1/2$. This can be done by a first moment method.

   Step 2. We prove the claim that if the constraint graph has only t cycles, then over the random choices of clause types (i.e. whether to put negation on the related literals), the NAE instance is satisfiable with probability $2^{-O(t)}$.

   Therefore, with probability $2^{-O(1)}$, a random NAE instance is satisfiable (when $m = (1-\epsilon)n$).

5. (a) Consider the process that one starts from v, and each time visits the only next vertex of the current vertex (since each vertex has out-degree exactly 1), until a vertex is revisited.

$r(v) = k$ if and only if the process above runs for $k-1$ times before a vertex is revisited. Since all the choices of out-going edges are independent, we have

$$\Pr[r(v) = k] = \prod_{i=1}^{k-1}\left(1 - \frac{i}{n}\right) \cdot \frac{k}{n}.$$

(b) Note that for any t,

$$\begin{aligned}
\Pr[r(v) \geq t] &= \sum_{k=t}^{n}\prod_{i=1}^{k-1}\left(1 - \frac{i}{n}\right) \cdot \frac{k}{n} \\
&= \left(\prod_{i=1}^{t-1}\left(1 - \frac{i}{n}\right)\right)\left(\frac{t}{n} + \left(1 - \frac{t}{n}\right)\left(\frac{t+1}{n} + \left(1 - \frac{t+1}{n}\right)\left(\frac{t+2}{n} + \cdots\right)\right)\right) \\
&= \prod_{i=1}^{t-1}\left(1 - \frac{i}{n}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\Pr[r(v) \leq \sqrt{n}/10] = 1 - \Pr[r(v) \geq \sqrt{n}/10 + 1] \\
=&1 - \prod_{i=1}^{\sqrt{n}/10}\left(1 - \frac{i}{n}\right) \leq 1 - \left(1 - \sum_{i=1}^{\sqrt{n}/10}\frac{i}{n}\right) = \frac{(\sqrt{n}/10)(\sqrt{n}/10 - 1)}{2n} \leq \frac{1}{3},
\end{aligned}$$

and

$$\Pr[r(v) \geq 10\sqrt{n}] = \prod_{i=1}^{10\sqrt{n}-1}\left(1 - \frac{i}{n}\right) \leq \prod_{i=\sqrt{n}}^{10\sqrt{n}-1}\left(1 - \frac{i}{n}\right) \leq \left(1 - \frac{\sqrt{n}}{n}\right)^{9\sqrt{n}} \leq \frac{1}{3}.$$