Lecture 7

Lecture date: Sept. 19 *Scribe: Partha S. Dey*

In the previous lecture, using first moment argument on number of SAT assignments of a random k-SAT formula we proved the following theorem.

Theorem 1 *Let $P_N(k, \alpha)$ be the probability that a random formula from $SAT_N(k, M = \alpha N)$ ensemble is SAT. If $\log 2 + \alpha \log(1 - 2^{-k}) < 0$, then*

$$P_N(k, \alpha) \to 0 \text{ as } N \to \infty.$$

In today's lecture, instead of considering the set of all SAT assignments(that may have large cardinality for certain instantiation of the random formula), we consider a subset of it by taking SAT assignments that satisfy a local maximality criteria.

1 Locally Maximal Satisfying (LMS) Assignments

Definition 2 *For a SAT formula ψ on N variables, an assignment $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$ is called a* Locally Maximal Satisfying *(LMS) assignment for ψ if*

1. *$\mathbf{x}$ satisfies ψ,* i.e. *$\psi(\mathbf{x}) = 1$ and*
2. *any assignment obtained from $\mathbf{x}$ by changing exactly one 0 value to 1 does not satisfy ψ.*

The following claim immediately follows from the definition.

Claim 1 *If ψ is satisfiable then there exists a LMS for ψ.*

Proof: Consider the set of all assignments that satisfy ψ. This set is finite and non-empty. Take an assignment in this set with smallest number of 0's. That is clearly a LMS assignment for ψ. □

Definition 3 *For a SAT formula ψ, define,*

$$U(\psi) = \text{ Total number of LMS assignments for } \psi.$$

Lemma 4 *Let ψ be a random formula from the $SAT_N(k, M)$ ensemble, where $M = \alpha N$ for some $\alpha > 0$. Then an upper bound for the expected value of the random variable $U(\psi)$ is given by,*

$$\mathbf{E}[U(\psi)] \leq (1+q)^N (1-2^{-k})^M,$$

where $q = 1-(1-p)^M$, and $p = k/[N(2^k-1)]$.

Before going to the proof of lemma 4, we state a corollary which gives a better upper bound on the satisfiability threshold than the one given in theorem 1.

Corollary 5 *For any $k \geq 2$, let α^* be the unique positive solution of the equation:*

$$f(\alpha) \equiv \alpha \log(1-2^{-k}) + \log\left[1 - \exp\left(\frac{-k\alpha}{2^k-1}\right)\right] = 0. \tag{1}$$

Then $\lim_{N\to\infty} P_N(k,\alpha) = 0$ if $\alpha > \alpha^*$.

Exercise 6 *(1 point) Prove that equation (1) has a unique positive solution (Note that 0 is always a solution of (1)).*

Proof: Assuming that the positive solution α^* of $f(\alpha) = 0$ is unique, it is easy to check that $f(\alpha) < 0$ for $\alpha > \alpha^*$. Now note that,

$$(1-p)^M = \left(1 - \frac{k}{N(2^k-1)}\right)^{\alpha N} = \exp\left[-\frac{\alpha k}{2^k-1}\right] + o(1).$$

Hence, using lemma 4 and the fact that ψ is SAT iff $U(\psi) \geq 1$, we have

$$\begin{aligned}
P_N(k,\alpha) = \mathbf{P}(\psi \text{ is SAT}) &= \mathbf{P}(U(\psi) \geq 1) \\
&\leq \mathbf{E}[U(\psi)] \\
&\leq (1-2^{-k})^M (1+q)^N \\
&= \left[(1-2^{-k})^\alpha \left(2 - \exp\left[-\frac{\alpha k}{2^k-1}\right] + o(1)\right)\right]^N \\
&= \left[e^{f(\alpha)} + o(1)\right]^N \\
&\stackrel{N\to\infty}{\longrightarrow} 0, \text{ if } \alpha > \alpha^*.
\end{aligned}$$

□

Let us consider an assignment $\mathbf{x}$ where exactly L variables are set to 0 and the remaining $N-L$ variables are set to 1. Without loss of generality, assume $x_1 = x_2 = \cdots = x_L = 0$, $x_{L+1} = \cdots = x_N = 1$.

Claim 2 *The probability that a random clause constrains the variable x_1, given that the clause is satisfied by the assignment $\mathbf{x}$ is $p = k/[N(2^k-1)]$.*

Proof: Total number of k-clauses satisfied by $\mathbf{x}$ is $(2^k-1)\binom{N}{k}$. Among these, the number of $(k-1)$-clauses not satisfied by $(x_2, x_3, \ldots, x_N)$ is $\binom{N-1}{k-1}$.

Hence, the required probability is,

$$p = \frac{\binom{N-1}{k-1}}{(2^k-1)\binom{N}{k}} = \frac{k}{N(2^k-1)}.$$

$\square$

Claim 3 *The probability that the variable x_1 is constrained by at least one of the M clauses, given that all these clauses are satisfied by $\mathbf{x}$ is $q = 1-(1-p)^M$.*

Proof: Proof follows from claim 2 and the independence of the M clauses. $\square$

Claim 4 *Let $\mathcal{A}_i$ be the event that x_i is constrained by at least one of the M clauses. Then,*

$$\mathbf{P}\left[\bigcap_{i=1}^{L} \mathcal{A}_i \mid \textit{all clauses are satisfied by } \mathbf{x}\right] \leq \prod_{i=1}^{L} \mathbf{P}[\mathcal{A}_i | \textit{ all clauses are satisfied by } \mathbf{x}] = q^L.$$

Exercise 7 *(1 or 2 points inversely proportional to proof length) Prove claim 4.*

Now we go to the proof of lemma 4.

Proof: For $L = 1, 2, \ldots, N$, define $\mathbf{z}_L = (\underbrace{0, 0, \ldots, 0}_{L}, \underbrace{1, 1, \ldots, 1}_{N-L})$.

$$\begin{aligned}
\mathbf{E}\left[U(\psi)\right] &= \sum_{\mathbf{x}\in\{0,1\}^N} \mathbf{P}\left[\psi \text{ is SAT by } \mathbf{x}\right] \cdot \mathbf{P}\left[\mathbf{x} \text{ is LMS for } \psi \mid \psi \text{ is SAT by } \mathbf{x}\right] \\
&= \sum_{L=0}^{N} (1-2^{-k})^M \binom{N}{L} \mathbf{P}\left[\mathbf{z}_L \text{ is LMS for } \psi \mid \psi \text{ is SAT by } \mathbf{z}_L\right] \\
&\leq (1-2^{-k})^M \sum_{L=0}^{N} \binom{N}{L} q^L \\
&= (1-2^{-k})^M (1+q)^N.
\end{aligned}$$

$\square$

2 SAT lower bounds

Two main strategies have been used to derive lower bounds of the satisfiability threshold: Algorithmic approach and Second Moment approach.

The first approach consists in analyzing explicit heuristic algorithms for finding SAT assignments. The idea is to prove that a particular algorithm finds a SAT assignment with finite probability as $N \to \infty$ when α is smaller than some value. One of the simplest such bounds is obtained by considering *unit clause propagation algorithm*. Before going to that we present a trivial lower bound of the satisfiability threshold.

2.1 Trivial Lower bound

Exercise 8 *(0.1 point) Show that if $k_1 < k_2$ then we have, $P_N(k_1, \alpha) \leq P_N(k_2, \alpha)$.*

Using this with the fact that $\lim_{N\to\infty} P_N(2, \alpha) = 1 \; \forall \; \alpha < 1$, we have,

Corollary 9 *For all $k \geq 2$ if $\alpha < 1$, then $P_N(k, \alpha) \to 1$ as $N \to \infty$.*

2.2 Unit Clause Propagation Algorithm

Input: A SAT formula ψ with varying size clauses.

Output: *"A solution exists"* or *"Cannot determine whether a solution exists"*.

Algorithm:

- If there are any unit clauses, pick a unit clause uniformly at random, satisfy it and simplify ψ to obtain ψ'.
- Otherwise pick a variable uniformly at random, assign it a value 0/1 uniformly at random and simplify ψ to obtain ψ'.
- Run the algorithm on ψ'.
- If there is an empty clause, output "cannot determine whether a solution exists".
- otherwise output "a solution exists".

Theorem 10 *For the $SAT_N(k, \alpha N)$ ensemble if*

$$\alpha < \frac{1}{2}\left(\frac{k-1}{k-2}\right)^{k-2} \frac{2^k}{k},$$

then for some $\epsilon_\alpha(k) > 0$ *the probability that the algorithm outputs* "a solution exists" *is greater that* $\epsilon_\alpha(k)$ *in the limit* $N \to \infty$.

Note that this gives a lower bound on the satisfiability threshold, by the following corollary of Friedgut's theorem.

Corollary 11 *Let* $k \geq 2$. *If* α *is such that* $\liminf_{N\to\infty} P_N(k,\alpha) > 0$ *then for any* $\delta \in (0,\alpha)$,

$$\lim_{N\to\infty} P_N(k, \alpha - \delta) = 1.$$

Idea of the proof: After t-steps the formula will contain k-clauses, $(k-1)$-clauses, ... and 1-clauses. Denote by $\mathcal{C}_j(t)$ the set of j-clauses, $j = 1, 2, \ldots, k$ and by $c_j(t) = |\mathcal{C}_j(t)|$ its size. The main steps are as follows.

- Show that at each step t, each clause in $\mathcal{C}_j(t)$ is uniformly distributed.
- Analyze expected change in $c_j(t), j = k, k-1, \ldots, 1$ over t.
- Show as $N \to \infty$ at fixed $s = t/N$, the variables $c_j(t)/N, j \geq 2$ concentrate around their means and they converge to smooth functions that satisfy some differential equations.
- Solve the equations and show that number of 1-clauses remain small.
- Show that the probability for an empty clause to appear is bounded away from 1 if α is smaller than some number.