STAT 206A: Gibbs Measures | Elchanan Mossel

Lecture 6: Random k-SAT problems

Lecture date: Sept 14 | *Scribe: Guillaume Obozinski*

Definition 1 ($\mathbf{SAT}_N(k, \alpha)$) *is a random k-SAT formula ψ on N variables where each possible clause is chosen independently with probability $\frac{\alpha N}{2^k \binom{N}{k}}$*

Theorem 2 (Friedgut) *Suppose $k \geq 2$. Let $P_N(k, \alpha) = \mathbf{P}(SAT_N(k, \alpha)$ is SAT).*

$$\exists \alpha_c(N), \quad \forall \epsilon > 0, \quad \begin{cases} P_N(k, \alpha_c(N) + \epsilon) \to 0 \\ P_N(k, \alpha_c(N) - \epsilon) \to 1 \end{cases}$$

1 Random 2-SAT and 2-XOR-SAT

Theorem 3

$$\lim_N P_N(k = 2, \alpha) = \begin{cases} 1, & \alpha < 1 \\ 0, & \alpha > 1 \end{cases}$$

Proof: *We won't prove this. The probability of the formula being unsatisfiable may be expressed in terms of the directed graph associated with the formula as: $\mathbf{P}(\exists i | x_i \to \bar{x}_i, \bar{x}_i \to x_i)$. The proof of the theorem follows by analyzing this probability.* □

We consider the simpler:

Definition 4 ($\mathbf{XOR\text{-}SAT}_N(2, \alpha)$) *is a random XOR-SAT problem where, given the choice of a pair of variables, the two possible constraints are $x_i \oplus x_j = 0$ and $x_i \oplus x_j = 1$ and each of the clauses are drawn independently with probability $\frac{\alpha N}{2\binom{N}{2}}$. We denote by $\lim P_N^X(k = 2, \alpha)$ the probability that a random XOR-SAT$_N(2, \alpha)$ formula is satisfiable.*

Theorem 5

$$\lim P_N^X(k = 2, \alpha) = \begin{cases} > 0, & \alpha < \frac{1}{2} \\ 0, & \alpha > \frac{1}{2} \end{cases}$$

Proof: *We will prove the theorem by considering the random graph on the vertex set $x_1, \ldots, x_n$, where an edge is present between x_i and x_j if the equation $x_i \oplus x_j = c_{i,j}$ is part of the $2 - XOR - SAT$ formula.*

Consider first the case $\alpha > 1/2$. In this case from the theory of random graph we know that with high probability there is a $\Omega(n)$ number of vertices belonging to cycles of length $O(\log n)$. This implies in particular, that w.h.p. there are $\Omega(n/\log n)$ disjoint cycles. The probability of each cycle being satisfiable is $1/2$ by Claim 6. Therefore, the probability that the formula is satisfiable goes to 0 as $N \to \infty$.

Now consider $\alpha < 1/2$. Note first that $\forall \alpha$ $P_N(k=2,\alpha) \nrightarrow 1$ since

$$\mathbf{P}(\exists (i,j), x_i \oplus x_j = 0 \textit{ and } x_i \oplus x_j = 1 \textit{ are both in } \psi) \to e^{-\frac{\alpha^2}{2}}$$

.
However, it still holds that for $\alpha < 1$, the formula is satisfiable with positive probability as we now prove. By Claim 6 ψ is UNSAT iff there is a cycle with an odd number of $x_i \oplus \bar{x}_i = 1$ constraints and since by Claim 7, for $\alpha < 1/2$, with positive probability none of these cycles have any $x_i \oplus \bar{x}_i = 1$ constraint. These two claims together imply that ψ has to be SAT with positive probability. □

Claim 6 *ψ is UNSAT iff there exists a cycle $(i_1, i_2, ..., i_J)$ s.t.*

$$\begin{cases} \forall j, \quad x_{i_j} \oplus x_{i_{j+1}} = a_{i_j, i_{j+1}} \\ \bigoplus_{j=1}^{J} a_{i_j, i_{j+1}} = 1 \end{cases}$$

Proof: *If such a cycle exists then summing shows that ψ is UNSAT. Conversely ψ is UNSAT if the linear system of equations is over-constrained, ie gaussian elimination must eliminate all the variables leaving an equation impossible to satisfy, eliminating all variables is only possible with a cycle and then the sum of the RHS must be 1.* □

Claim 7 *Suppose $\alpha < 1/2$ then There exist $C_1(\alpha), C_2(\alpha)$ so that with probability greater than 0.9*

- *The total number of cycles is less than $C_1(\alpha)$*
- *All cycles are of length less than $C_2(\alpha)$*
- *With positive probability independent of N, no cycle has any $(x_i \oplus \bar{x}_i = 1)$ constraint.*

Proof:

- *There are at most $\frac{N(N-1)...(N-k+1)}{k}$ possible cycles of length k*
- *Each is present with probability $\left(2\frac{\alpha N}{2\binom{N}{2}}\right)^k = \left(\frac{\alpha}{N-1}\right)^k$.*

- *So* $\mathbf{E}(\sharp\textit{cycles of length } k) \le \alpha^k$
- *Thus* $\mathbf{E}(\sharp\textit{cycles}) \le \frac{1}{1-\alpha}$
- *If* $C_1(\alpha)$ *is chosen s.t* $\sum_{k>C_1(\alpha)} \alpha^k < 0.1$, *then*
  $\mathbf{P}(\exists \textit{ a cycle of length} > C) \le \mathbf{E}(\sharp\textit{cycles of length} > C) \le \sum_{k>C} \alpha^k < 0.1$
- *By Markov's inequality, with* $C_2(\alpha) = \frac{10}{1-\alpha}$, $\mathbf{P}(\sharp\textit{cycles} > C_2(\alpha)) < \frac{\mathbf{E}(\sharp\textit{cycles})}{C_2(\alpha)} < 0.1$
- *Given that an edge* (x_i, x_j) *is present the probability that the only corresponding clause present is* $(x_i \oplus x_j = 0)$ *is at least* $\frac{1}{4}$. *This follows since w.h.p. there is no clause chosen* 3 *or more times.*
  So with probability at least $4^{-C_1(\alpha)C_2(\alpha)}$ *all cycles are without* $(x_i \oplus \bar{x}_i = 1)$ *constraint.*

□

2 Random 3-SAT

Algorithm 1 The k-SAT random walk algorithm

choose x_0 uniformly at random
while $\psi(x^t) = 0$ **do**
  choose an UNSAT clause at random
  choose randomly one of the literals z_i
  flip x_i^t to obtain x_i^{t+1}
end while

Claim 8 *Assume* $\exists y,\ \psi(y) = 1$ *and* $\psi(x^t) = 0$ *then*

$$\mathbf{P}(\, d_H(x^{t+1}, y) = d_H(x^t, y) - 1\,) \ge \frac{1}{k}$$

Exercise 9 *Prove that the expected time to hit* y *is* $\mathcal{O}\left(\frac{2(k+1)^N}{k}\right)$

Claim 10 $SAT_N(k, M = \alpha M)$ *is UNSAT with high probability if* $\alpha > \frac{-\log 2}{\log(1-2^k)}$
Proof: *Let* $z[\psi] = \sharp$ *of SAT assignments.*

$$\mathbf{E}[Z] = 2^N \mathbf{P}(\mathbf{0} \textit{ satisfies } \psi) = 2^N (1 - 2^{-k})^{\alpha N} = \exp(N(\log 2 + \alpha \log(1 - 2^{-k})))$$

.

Note that here again we assumed that the number of clauses if fixed. Since the number of clauses is concentrated around its mean, the claim also holds for $SAT_N(k, \alpha)$. □

Exercise 11 *Show that the first moment method applied to* $SAT_N(k, \alpha)$ *gives the bound* $\alpha > 2^k \log 2$