

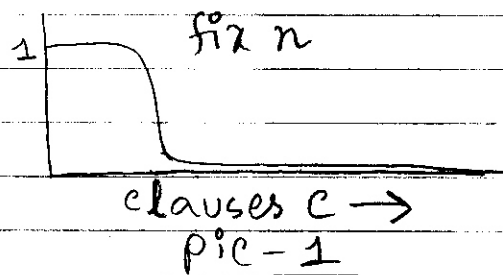
Lecture-8

David Lin & Savdia Ogyshsee

1
Feb-8, 06

Phase Transition for CNF-Satisfiability

Generating random 3-CNF formula f with n variables and we ask the question what is probability that f is satisfiable as a function of number of clauses.



we fix n , and then we have some number of clauses for formula, if have just 1 clause, let's say we have 100 variables and 1 clause, that formula is clearly satisfiable. Let's pick a variable that appears in the clause and set it to true. Clearly no matter what these literals are, there is ^{just} 1 clause that we can find that makes the formula true. If we put 2 clauses, probably still can; but if we put enough clauses there, and randomly put these literals in, it may be possible to satisfy the first hundred but highly unlikely to satisfy the hundred and first.

So, probability is 1 when it starts out. And somewhere goes to zero for random assignments (pic 1). And there is a sharp threshold. It is still an open problem to know where this threshold is.

We start with n variables and c clauses.

The word literal is used here to refer to the symbol (either complemented or non-complimented).

f has $2n$ literals. Literals are things like $x, \bar{x}, y, \bar{y}, \dots$. Let's suppose $c = \alpha n$. [some constant α].

We will find out how often each literal appears on average. So, for 3-CNF each clause has 3 literals; thus f has a total of $3\alpha n$ literals appearing.

So, Average number of times a given literal appears is $\frac{3\alpha n}{2n}$ which is $\frac{3\alpha}{2}$.

Let's fix the constant α .

Example:- suppose $\alpha = 5$; then each literal appears on average $\frac{3 \times 5}{2} = 7.5$ times.

Which means if we set 2 a literal to value true then we would expect to satisfy 7.5 clauses. [α is a relatively small number compared to n].

But if we set n literals to true, we would not expect $7.5n$ clauses to be satisfiable. We will try to calculate a probabilistic argument to get a bound on where the threshold is.

Now, let's switch to a k -CNF formula with n variables and c clauses. We want to claim that there is a threshold $c = r_k n$. We'll try to estimate what r_k is. We'll start out with what the upperbound is.

Upperbound on r_k :- For a given truth

assignment to variables what is probability that a random clause is satisfiable. There are k literals in the clause.

$x_1 + x_2 + \dots + x_k$. What's the probability that one of these is false? $\Rightarrow \frac{1}{2}$ [either we'll get a value 0 or 1].

Then, probability that all false is $(\frac{1}{2})^k$ which is k of them.

And, probability that 1 or more is true is $(1 - (\frac{1}{2})^k)$. This is probability that a given clause is satisfied.

We want all clauses to be satisfied.

Probability all clauses satisfiable is

$$\left(1 - \left(\frac{1}{2^k}\right)\right)^{cn}$$

Let's say there are n clauses. So, $(1 - (\frac{1}{2^k}))^n$ is the probability that all clauses are satisfiable for 1 true assignment. There are 2^n truth assignments. Want to know what is the expected number of these satisfying all clauses?

Expected number of truth assignment satisfying all formula = $2^n (1 - \frac{1}{2^k})^n$.

Question is why did we multiply the number of truth assignments * the probability that a truth assignment is correct to get the expected number?

Create an indicator variable for each truth assignment. We'll sum over the expected value of each truth assignment being correct to get the expected value of all truth assignments being correct. And by linearity of expectation we derived the expected number of truth assignment is $2^n (1 - \frac{1}{2^k})^n$ satisfying all clauses.

If $n = 2^k \ln 2$ then the expected number of satisfying assignments is

$$\begin{aligned} 2^n \left(1 - \frac{1}{2^k}\right)^{n 2^k \ln 2} &\cong 2^n \left(\frac{1}{e}\right)^{n \ln 2} \\ &= 2^n e^{-\ln 2 \cdot n} = 2^n 2^{-n} = 1 \end{aligned}$$

This tells if $n > 2^k \ln 2$ then the number of satisfying assignments goes to 0. Upperbound of the threshold is $2^k \ln 2$. If we have more than $2^k \ln 2$ clauses it won't be satisfiable.

Lower Bound Chao Franco 1990 - Given algorithm for finding satisfying assignment if $n < \frac{2^k}{k}$. if $\exists$ clause with just one literal, then select that literal and set it to true. Otherwise select literal at random and set it to true.

That says the threshold is somewhere between $\frac{2^k}{k}$ and $2^k \ln 2$.

for $k=3$, these are 2.667 and 5.5452
So, if we have a 3 CNF formula with n variables and if you put more than $5.5452n$ clauses, then it is not likely to be satisfied. If we put fewer than $2.6667n$ clauses, it will be satisfied and somewhere in between there is a tight bound where the random formula will switch from being satisfiable to non satisfiable.

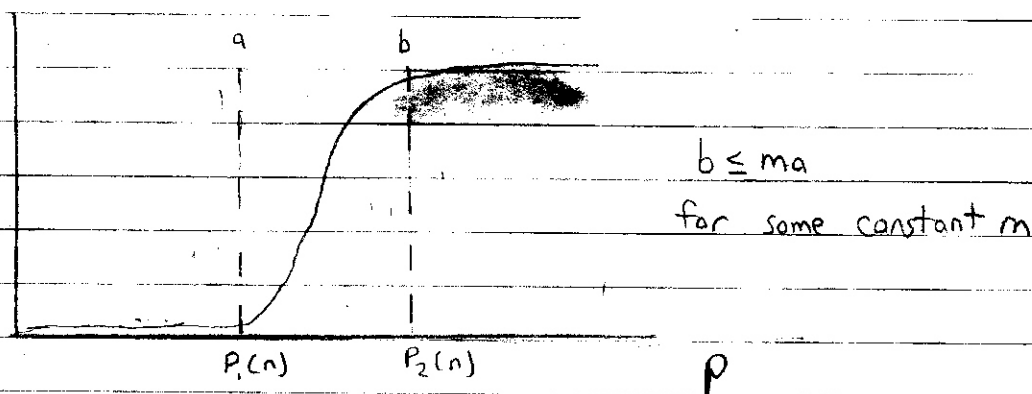
Increasing Property and Threshold

We'll prove a theorem that if we have any property which is an increasing property (described later) then that property will have a sharp threshold.

Increasing property:- Q is an increasing property of G if when a graph G_1 has the property and any graph obtained from G_1 by adding edges has the property.

Example:- $\rightarrow$ Connectivity of graphs $\rightarrow$ once we put enough edges for graph to be connected, adding more edges would not cause it to become unconnected.

$\rightarrow$ Disappearance of isolated vertex.



What is threshold? A property Q has a threshold $p(n)$ if for $p_1(n)$ where $p_1(n)$ is $o(p(n))$ the graph almost surely does not have property and for $p_2(n)$ such that the threshold is $o(p_2(n))$ then the graph ^{almost} surely has a property.