CS 683: Advanced Design and Analysis of Algorithms 1/30/2008

Lecture Notes 5

Professor: John Hopcroft *Scribes: Kathleen King & Brook Li*

1 Thresholds for Monotone Properties

Theorem. Every monotone property Q in $N(n,p)$ has a threshold.

Proof. First consider what must be done to prove this theorem. Then take any ε such that $0 < \varepsilon < \frac{1}{2}$, with ε independent of n. Let $p_\varepsilon(n)$ be a function such that $N(n, p_\varepsilon(n))$ has property Q with probability k. Similarly define $p_{1-\varepsilon}(n)$. We need to show that there exists a constant c such that $p_{1-\varepsilon}(n) \leq np_\varepsilon(n)$.

Now, for our given ε, select m such that $(1-\varepsilon)^m \leq \varepsilon$. Consider the union of m independent copies of $N(n, p_\varepsilon(n))$. We claim that this union gives us $N(n,q)$ for some q. [Note: Taking the union is like a flipping a coin m times for each integer and including the integer if the coin is heads on any one of those flips. The same result could be gained from flipping a differently weighted coin once.] We can calculate q, which is the probability that at least one of m coin flips comes up as heads. The probability of always getting tails is $(1-p_\varepsilon)^m$, so we can take the binomial expansion to get

$$\begin{aligned} q &= 1-(1-p_\varepsilon)^m \\ &= 1-[1-mp_\varepsilon+...] \qquad \text{where the "..." is positive} \\ &= mp_\varepsilon - ... \\ &\leq mp_\varepsilon. \end{aligned}$$

So we have

$$\Pr[N(n, mp_\varepsilon) \in Q] \geq \Pr[N(n,q) \in Q]. \tag{1}$$

Since $N(n,q)$ is the union of m independent copies of $N(n, p_\varepsilon(n))$, we can use properties of unions to get the following:

$$\begin{aligned} \Pr[N(n,q) \notin Q] &\leq \Pr[\forall_m N(n, p_\varepsilon(m)) \notin Q] \\ &\leq (1-\varepsilon)^m \\ &\leq \varepsilon \qquad \text{(by the definition of } m). \end{aligned}$$

So we have

$$\Pr[N(n,q) \in Q] \geq 1-\varepsilon. \tag{2}$$

Then, combining Equations 1 and 2, we get

$$\Pr[N(n, mp_\varepsilon) \in Q] \geq \Pr[N(n,q) \in Q \geq 1-\varepsilon.$$

Since Q is a monotone property, we then know that $mp_\varepsilon \geq p_{1-\varepsilon} \geq p_\varepsilon$. So p_ε and $p_{1-\varepsilon}$ are asymptotically identical. □

Note: One possible homework problem would be to develop a more intuitive proof of this theorem.

2 Thresholds for CNF Satisfiability Problems

A boolean formula is in CNF, or conjunctive normal form, if it is composed solely of a conjunction of clauses, each of which is a disjunction of literals. A boolean formula is in k-CNF if it is in CNF with k literals per clause. The form of a 3-CNF formula is shown below:

$$f(x_1, x_2, ..., x_n) = (x_1 + x_2 + x_3)(\overline{x_1} + x_4 + x_5)...$$

An ongoing challenge is to develop efficient algorithms to find combinations of values of variables that satisfy boolean formulae in CNF.

For a randomly generated formula in CNF, as the number of literals per clause is kept constant, the probability that a formula is satisfiable decreases as the number of clauses increases, with a threshold where the probability decreases very quickly.

There appears to be a second threshold, occuring at a fewer number of clauses than the threshold where the probability of satisfiability decreases quickly. For formulae with a fewer number of clauses than occurs at this apparent second threshold, algorithms to solve the satisfiability problem in polynomial time are known. For formulae on the other side of this threshold, no efficient (polynomial time) algorithm to solve the satisfiability problem is known. The idea is shown in figure 1.

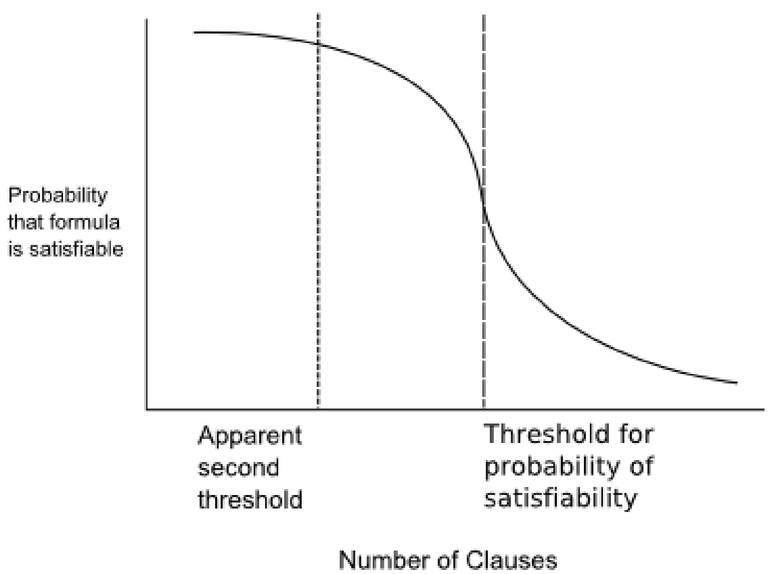

Figure 1: Diagram showing idea of two thresholds for random formulas in k-CNF

The expected number of occurences of a particular literal in a random formula in k-CNF is

$$\frac{3c}{2n}$$

where c is the number of clauses and n is the number of distinct variables. Note that the number of distinct literals is equal to $2n$: for each variable the variable itself and its complement are both literals.