Regular $k$-SAT: A model for Random $k$-Satisfiability

September 17, 2004

Abstract

We define a model for generating random $k$-SAT formulas (Regular $k$-SAT) in which each literal appears in almost the same number of clauses. We report experimental results that show that Regular $k$-SAT instances are much harder than the usual random $k$-SAT problems. We find that the new model exhibits a phase transition around 3.6, similar to the results found in the usual model. We show that the analysis of the greedy algorithm proposed in Kaporis et al [15] can be extended for the new model. We prove that for $r \leq r_s = 2.46$ the random formulas are whp satisfiable. We prove that for $r \geq 11$ whp all random Regular 3-SAT formulas are unsatisfiable. We finally extend the results of Chvatal and Szemeredi [6] to prove that if the random formula is unsat it needs and exponential refutation proof.

1 Introduction

The introduction of new methods for generating random hard instances is an important factor in the recent development of new search algorithms for satisfiability testing (SAT) (see SAT competition [19]). Randomly generated SAT problems give important insights into typical case complexity.

The most popular model for generating random SAT problems is the uniform $k$-SAT model, formed by selecting uniformly and independently $m$ clauses from the set of all $2^k \binom{n}{k}$ $k$-clauses on a given set of $n$ variables. The uniform $k$-SAT model have been extensively studied. Such randomly generated instances exhibit a “phase transition” as a function of the ratio $\alpha$ of clauses to variables [20]. Uniform $k$-SAT problems with a small $\alpha$ value almost all have one or more satisfying assignment, whereas problems with a large $\alpha$ value have too many constraints and become unsatisfiable.

Experimental results showing the “phase transition” phenomenon motivated theoretical interest in understanding uniform $k$-SAT. The main open question on uniform $k$-SAT is the existence of a sharp threshold as the ratio of clauses to variables is increased. More precisely, there exist constants $\alpha_k$ such that a random formula with ratio $\alpha < \alpha_k$ is satisfiable w.h.p., whereas a random formula with $\alpha > \alpha_k$ is unsatisfiable w.h.p. For $k = 2$, Chvatal and Reed [5], Goerdt [14] and Fernandez de la Vega [11] independently proved the existence at $\alpha_2 = 1$ of the sharp threshold. For $k \geq 3$, much less is known. Friedgut [12] proved the existence of a sharp threshold around some critical sequence of values. More precisely, Friedgut showed that there exists a function $\alpha_k(n)$ such that when the number of clauses is around $\alpha_k(n)n$ the satisfiability of the formula drops abruptly from near 1 to near 0. Moreover, these results do not provide information about the value of $\alpha_k(n)$ and the dependence on $n$.
For uniform 3-SAT there has been a number of results on lower bounds on the threshold $\alpha_3$, see [2] for a survey, the best known result is from [16, 10] where it is proven that a random uniform instance for 3-SAT is satisfiable w.h.p if $\alpha < 3.52$. The best known result for upper bounds states that for $\alpha > 4.571$ random uniform 3-SAT formulas are unsatisfiable w.h.p [17], for a survey of upper bounds see [9].

In this paper we give experimental and theoretical results for a different model for random satisfiability. In the regular $k$-SAT (Reg $k$-SAT) model all the literals have the same number of occurrences in the formula. Given $\alpha$ the expected ratio of clauses to variables and $n$ the number of variables, let $\rho = \frac{\alpha n}{2}$, the expected number of occurrences of each literal in the formula. We generate our instances such that each literal appears $\lfloor \rho \rfloor$ or $\lceil \rho \rceil$ times in the formula and such that $\rho$ is the expected value.

To understand why this model is interesting we show in figure 1 a comparison of the hardness between uniform 3-SAT and Reg 3-SAT. The hardest problems with 300 variables for uniform 3-SAT needed $1e + 04$ branches while for the same number of variables Reg 3-SAT needed over $1e + 06$ branches. Another interesting feature of Regular 3-SAT model is that it exhibits a “phase transition” similar to the uniform 3-SAT. Around the ratio $\alpha = 3.6$ Reg 3-SAT instances change from being satisfiable to being unsatisfiable (see figure 2) and around the same ratio is the peak of complexity seen in figure 1.

Achlioptas et al. [1] introduced a generator of satisfiability formulas based on Latin squares that creates only satisfiable problems. More recently, that model was modified for a more “balanced” version [18] that increases the difficulty of the problems. As in the comparison uniform 3-SAT versus Reg 3-SAT in the case of these generators the effect of balancing increase the hardness of the problem. Another example of this phenomenon appears in coloring random graphs. When considering the Erdős–Rényi model $G(n, p = \frac{\rho}{n})$ versus the regular graphs $G(n, r)$ with the same average degree $r$, regular graphs are much harder to color than graph in $G(n, p)$. Intuitively, with balancing we eliminate the irregularities that algorithms try to use in order to solve a given instance.

Bayardo and Schrag [3] gave experimental results on a model very similar to the one we
present here. They report results for experiments uniform random 3-SAT and their regular model. At a relatively low value of the number of variables, the peak in difficulty for their model is almost an order of magnitude higher than that of uniform 3-SAT.

We begin with a precise definition of our model and use the results of Cooper et al. [8] to derive the sharp threshold for Reg 2-SAT. In section 3 we use a counting argument to prove that for \( \alpha > \) ... a Reg 3-SAT formula is unsat w.h.p. In section 4 we analyze a greedy algorithm on Reg 3-SAT formulas to prove that for \( \alpha < 2.46 \) the algorithms find w.h.p a satisfying assignment and therefore a Reg 3-SAT formula with that density is satisfiable w.h.p. Finally...

## 2 The model

Let \( n \) be the number of variables, \( m = \alpha n \) the number of clauses in a \( k \)-CNF formula \( F \). Let \( \rho = \kappa n / 2 \) be the average degree of each literal. We say that a literal \( x \) has degree \( l \) if \( x \) appears \( l \) times in the formula. The degree sequence associated with a formula \( F \) is the sequence \( \{d_1, d_2, \ldots, d_n, d_{n\ldots} \} \) where \( d_x \) is the number of clauses in which the literal \( x \) occurs.

We define our model in a similar way people have defined random graphs with prescribed literals degree. Given a sequence of non negative real numbers \( \{p_l \} \geq 0 \) where \( \sum_{l=0}^{\infty} p_l = 1 \) and \( p_l \) is the probability of a literal having degree \( l \). We get the sequence of literals degrees \( \{d_1, d_2, \ldots, d_n, d_{n\ldots} \} \) independently from that distribution and conditioning on the event that the sum of all degrees is a multiple of \( k \).

In our case sequence of probabilities is very simple \( p_{|\rho|} = p \) and \( p_{|\rho| + 1} = 1 - p \) where \( p \) is defined so that the expected number of clauses is \( m = \alpha n \), i.e \( p = |\rho| + 1 - \rho \), and all the other values for \( p \) are zero. Then we consider we consider uniformly all possible formulas with that degree sequence.

### 2.1 Reg 2-SAT

Let \( d = \{d_1, d_2, \ldots, d_n, d_{n\ldots} \} \) a degree sequence corresponding to a 2-SAT formula. If \( d_x < \Delta \) we say that \( d \) is \( \Delta \) proper.

The analysis of the threshold for the Reg 2-SAT model comes as a consequence of the following theorem.

**Theorem 1.** [8] Let \( 0 < \varepsilon < 1 \) and \( n \to \infty \). Let \( d \) be any \( \Delta \)-proper degree sequence over \( n \) variables, with \( \Delta = n^{1/11} \), and let \( F \) be a uniform random simple formula with degree sequence \( d \), then

If \( D < (1 - \varepsilon)m \) then \( P(F \text{ is satisfiable}) \to 1 \)

If \( D > (1 - \varepsilon)m \) then \( P(F \text{ is satisfiable}) \to 0 \)

where \( m \) is the number of clauses and \( D = \sum_{i=1}^{n} d_i d_{-i} \)

**Corollary 1.** The Reg 2-SAT formulas have a threshold at \( \alpha = 1 \).

**Proof.** We are going to prove that w.h.p degree sequences generated with our Reg 2-SAT model have the property that \( D / m \to \alpha \). Using theorem 1 we can conclude that \( \alpha = 1 \) is the value of the threshold.

Let \( D = \sum_{i=1}^{n} d_i d_{-i} \) a random variable. Note the the expected value \( E(D) \) of \( D \) is \( \alpha^2 n \) and \( E(m) = \alpha n \). Note that \( 2m = \sum_{i=1}^{n} d_i + d_{-i} \) the sequence of variables \( d_i, x \in \{-n, \cdots -1, 1, \cdots n\} \)
are independent idetically distributed random variables. The variance of the variables $D$ and $m$ are easy to compute and we can see that there exist constants $c, c'$ such that $Var(D) = cn$ and $Var(m) = c'n$.

Using Chebyshev’s inequality one has that $P(|D - c^2 n| \geq n^{1/2+\delta}) \to 0$ as $n$ goes to infinity for any $\delta > 0$. A similar property follows for the variable $m$, $P(|m - cm| \geq n^{1/2+\delta}) \to 0$ as $n$ goes to infinity. Therefore the property follows and then the claim.

3 Upper bound on the threshold

Let suppose first that every literal has the same integer value degree $\rho$. In order to estimate the probability that a random formula is satisfiable, we bound that probability by the expected number of solutions, i.e.

$$Pr(F \text{ is sat}) \leq E(\# \text{ solutions } F) = 2^n Pr(x \text{ is a solution})$$

Last equality follows from the fact that as the formula is symmetric in all its literals all assignments $x \in \{0, 1\}^n$ have the same probability of been a solution.

**Lemma 1.**

$$Pr(0, 0, \ldots, 0 \text{ satisfies } F) \leq \frac{7^{3pn}}{8}$$

where $\beta < 2/3$ a constant.

**Proof.** Note that $z = (0, 0, \ldots, 0)$ is a solution if and only if $F$ does not contain any clause with all positive literals. Let $m = 2pn/3$ and $F = \bigwedge_{i=1}^m C_i$.

Let $A_i$ the event that $z$ satisfies the clause $C_i$, i.e. the event that there is at least one negative literal in $C_i$. Note that

$$Pr(\bigwedge_{i=1}^m C_i \text{ satisfies } z) \leq Pr(\bigwedge_{i=1}^{m'} C_i \text{ satisfies } z) = Pr(\bigcap_{i=1}^{m'} A_i) = \prod_{i=1}^{m'} Pr(A_i \cap \{\neg A_{j} \mid j < i \})$$

for any $m' < m$.

Also note that $Pr(A_1) = 1 - \left(\frac{p_1}{\rho n} \right)$ and $Pr(A_2 \mid A_1) = 1 - \left(\frac{p_2}{\rho n} \right)$ where $p_1$ is the number of positive literals in the $C_1$ clause. Moreover $Pr(A_{i+1} \mid \bigwedge_{j=1}^i A_j) = 1 - \left(\frac{p_i}{\rho n} \right)$ where $p_i$ is the sum of all positive literals in the clauses $A_1, \ldots, A_i$.

We claim the for $i$ sufficiently large $p_i < 3k/2$, we are going to prove this at the end.

Then using the claim and given that $2rn - 3m' > 0$ , note that $\left(\frac{p_i}{\rho n} \right) \geq \frac{1}{8}$ and

$$Pr(\bigwedge_{i=1}^m C_i \text{ satisfies } z) \leq \prod_{i=1}^{m'+1} (1 - \left(\frac{p_i}{2\rho n} \right)) \leq \frac{7^{3pn}}{8}$$

Note now that if $p_{m'} < 3m' / 2$ as in the claim then for $m' = (1 - \epsilon) \frac{2}{7} pn$ for any $\epsilon > 0$, we can use previous formula to end our proof.
proof of the claim : ...

\[\square\]

**Lemma 2.** For \( \rho \geq 8 \), \( \Pr(F \text{ is sat}) \to 0 \) as \( n \to \infty \).

**Proof.** Using lemma 1 \( \Pr(F \text{ is sat}) \leq 2^{n\rho^2/\beta n} = e^{-n(\beta \rho \log n - \epsilon \rho^2)} \) that goes to 0 when \( \rho \geq 8 \). \( \square \)

4 The Algorithm

For a fixed ratio \( \alpha \) and the corresponding \( \rho = 3\alpha/2 \), let \( h \) the smallest integer greater than \( \rho \). Let \( \mathcal{X}_j \) for \( j = 0 \ldots h \) be the current collection of literals of degree \( j \). Consider the following algorithm.

**Greedy algorithm**

\begin{verbatim}
  begin
  let \( j = h \)
  while unset literals exists
    while \( \mathcal{X}_j \neq \emptyset \)
      set an arbitrary literal from \( \mathcal{X}_j \) to TRUE
      and its negation to FALSE and Del\&Shrink
      while unit clauses exits
        set an arbitrary unit clause to true and
        its negation to FALSE and Del\&Shrink
    end
    \( j = j - 1 \)
  end
end
\end{verbatim}

4.1 Formulas

Let \( l \) scaled number of current unset literals, \( c_3, c_2 \) scaled number of 3-clauses and 2-clauses respectively and \( x_s \) scaled number of literals of degree \( s \), \( s = 1, \ldots, 4 \). The equations for round
\[ \begin{align*}
\frac{dl}{dt} &= -2 - 4 \frac{c_2}{l - 2c_2} \\
\frac{dc_3}{dt} &= - \frac{3jc_3}{p} - \frac{3c_3}{l} + \left( - \frac{3jc_3}{p} - \frac{3c_3}{l} \right) \frac{2c_2}{l - 2c_2} \\
\frac{dc_2}{dt} &= \frac{3c_3 - 2c_2}{l} - \frac{2jc_2}{p} + \left( \frac{3c_3 - 2c_2}{l} - \frac{2jc_2}{p} \right) \frac{2c_2}{l - 2c_2} \\
\frac{dx_4}{dt} &= - (6c_3 + 2c_2) \frac{4x_4}{p^2} j - \frac{x_4}{l} - \delta_{4,j} \\
&\quad - (6c_3 + 2c_2) \frac{4x_4}{p^2} j + \frac{x_4}{l} + \frac{4x_4}{p} \frac{2c_2}{l - 2c_2} \\
\frac{dx_s}{dt} &= (6c_3 + 2c_2) \left( \frac{s+1}{p^2} x_{s+1} - \frac{(s) x_s}{l} \right) + x_4 - \delta_{s,j} \\
&\quad + (6c_3 + 2c_2) \left( \frac{s+1}{p^2} x_{s+1} - \frac{(s) x_s}{l} \right) \frac{2c_2}{l - 2c_2} \\
\text{for } s = 1, 2, 3
\end{align*} \]

with initial conditions \( l = 2, c_3 = c, c_2 = 0, x_4 = 2p, x_3 = 2(1-p), x_2 = 0, x_1 = 0 \)

5 Experimental Results

In this section we discuss the hardness of the Regular \( k \)-SAT model with respect to the usual random model for \( k \)-SAT. In our experiments we use instances for 3-SAT from both the regular and the usual model. We run the solver satz , one of the best solver for random instances, and we compare the results for both models.

We run our experiments for 150 and 300 number of variables for the regular model and 300 instances for the usual model. Each experiment was run for 100 instances for many value of the parameter \( \alpha = \frac{m}{n} \).

Figure 2 shows the median number branches used by satz to solve the instances. Figure 2 shows the sat-unsat phase transition like the one for the usual model.
Figure 2: Phase transition in 3-SAT. Probability that a 3-SAT problem has at least one satisfying assignment as a function of the ratio. We consider problems with 150 and 200 variables.
References


