A sharp threshold for a random constraint satisfaction problem

Abraham Flaxman [1]

Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh, PA 15213, USA

Abstract

We consider random instances I of a constraint satisfaction problem generalizing k-SAT: given n boolean variables, m ordered k-tuples of literals, and q "bad" clause assignments, find an assignment which does not set any of the k-tuples to a bad clause assignment. We consider the case where $k = \Omega(\log n)$, and generate instance I by including every k-tuple of literals independently with probability p. Appropriate choice of the bad clause assignments results in random instances of k-SAT and not-all-equal k-SAT. For constant q, a second moment method calculation yields the sharp threshold

$$\lim_{n\to\infty} \Pr[I \text{ is satisfiable}] = \begin{cases} 1, & \text{if } p \le (1-\epsilon)\frac{\ln 2}{qn^{k-1}}; \\ 0, & \text{if } p \ge (1+\epsilon)\frac{\ln 2}{qn^{k-1}}. \end{cases}$$

Key words: Constraint Satisfaction, Threshold Phenomena

1 Introduction

We study the following constraint satisfaction problem (CSP):

Input:

- A set of boolean variables $V = \{x_1, \ldots, x_n\}$
- A set of clauses, $C = \{C_1, \ldots, C_m\}$, where $C_i = (s_{i_1} x_{i_1}, \ldots, s_{i_k} x_{i_k})$, for $s_{i_j} \in \{-1, 1\}$

Email address: abie@cmu.edu (Abraham Flaxman).
[1] Supported in part by NSF VIGRE Grant DMS-9819950

- A set of "bad" clause assignments $Q \subseteq \{-1,1\}^k$ with $|Q| = q$.

Question: Does there exists an assignment $\psi\colon\ V \to \{-1,1\}$ such that for all C_i, we have $(s_{i_1}\psi(x_{i_1}),\ldots,s_{i_k}\psi(x_{i_k})) \notin Q$?

An instance $I = (V, C, Q)$ is called *satisfiable* if such an assignment exists. If no such assignment exists, I is called *unsatisfiable*.

This note focuses on instances generated by including every k-tuple of literals independently at random with probability $p = p(n)$, while allowing arbitrary sets Q of bad clause assignments (provided that $q = |Q|$ is constant). By considering particular sets of bad clause assignments, CSP specializes to two well known problems, k-SAT and not-all-equal k-SAT.

- *k-SAT is a special case of CSP:* we take $Q = \{-1^k\}$, i.e. there is one way for a clause to go bad, the setting which makes every literal in the clause false. Random k-SAT has been well studied, and a sharp threshold is known for $k = 2$ [6,10,12,14,15,17,21] and $k - \log n \to \infty$ [16]. For other values of k, in particular $k = 3$, a sharp threshold function is known to exist [13], but it is unknown what the function is. Upper and lower bounds are given in [1,4,7–9,11,15,18,20]
- *not-all-equal k-SAT is a special case of CSP:* we take $Q = \{-1^k, 1^k\}$. The satisfiability threshold for random not-all-equal-SAT is studied for $k = 3$ in [2] and a sharp threshold is known when k is sufficiently large [3].

In this note, we make the clause size $k = k(n)$ a function satisfying $k \geq D_\epsilon \log_2 n$, where D_ϵ is sufficiently large (for $\epsilon \leq \frac{1}{9}, D \geq 5\frac{1}{\epsilon}\ln\frac{q}{\epsilon}$ is enough). Then for any p and for a family of bad clause assignments $\{Q_i\}$ with $|Q_n| = q$, we define $I = I_{n,p}$ to be $(\{x_1,\ldots,x_n\}, C_{n,p}, Q_n)$, where $C_{n,p}$ is generated by including each k-tuple of literals independently at random with probability p.

Theorem 1 *For any natural number q and any $\epsilon > 0$ there exists D_ϵ such that for $k \geq D_\epsilon \log n$ and any family of bad clause assignments $\{Q_i\}$ with $|Q_n| = q$ we have*

$$\lim_{n\to\infty} \Pr[I_{n,p} \textit{ is satisfiable}] = \begin{cases} 1, & \textit{if } p \leq (1-\epsilon)\frac{\ln 2}{qn^{k-1}}; \\ 0, & \textit{if } p \geq (1+\epsilon)\frac{\ln 2}{qn^{k-1}}. \end{cases}$$

The consideration of "moderately growing clauses" is inspired by the work of Frieze and Wormald [16]. It appears that threshold results which require great labor for constant clause size become much easier when clause size is a sufficiently large function of n. In the following, the minimum necessary clause size $D_\epsilon \log n$ will be larger than $\log n$, so Theorem 1 holds for a smaller range of k than the threshold of [16]. However, Theorem 1 does not require as delicate a calculation as [16], and proves thresholds for other interesting specializations in one go.

Xu and Li obtained similar results using similar techniques for a different type of constraint satisfaction problem in [22]. They consider instances which have clauses of a fixed size k, allow variables to take values from a domain with $d = n^\alpha$ values, and have a different bad set for each clause chosen randomly, to prohibit $\Theta(d^k)$ candidate assignments. (In contrast, we have clauses of size $k = \Omega(\log n)$, a boolean domain of size $d = 2$, and a bad set prohibiting a *constant* number candidate assignments, which is the same set for each clause, and chosen non-randomly.)

The remainder of this note will prove Theorem 1. In Section 2 we will show unsatisfiability above the threshold by the first moment method. In Section 3 we will show satisfiability below the threshold by the second moment method.

In this note $\log x$ means $\log_2 x$. We use $\ln x$ for the natural logarithm, and $\log_\alpha x$ for the base-α logarithm.

2 Upper bound

We first show $I = I_{n,p}$ is unsatisfiable above the threshold. The proof is by the first moment method.

Claim 1 *Let* $p_0 = \frac{\ln 2}{qn^{k-1}}$*. Then for any* $p \geq (1+\epsilon)p_0$*, for any* Q *with* $|Q| = q$*, we have*

$$\lim_{n\to\infty} \Pr[I_{n,p} \textit{ is satisfiable}] = 0.$$

Proof For a particular assignment ϕ, there are qn^k clauses which violate some constraint of Q with respect to ϕ. So the probability that ϕ satisfies I is the probability that none of these clauses occur,

$$\Pr[\phi \text{ satisfies } I] = (1-p)^{qn^k}.$$

Let X denote the expected number of assignments satisfying I.

$$E[X] = 2^n(1-p)^{qn^k}.$$

For $p \geq \frac{\ln 2}{qn^{k-1}}(1+\epsilon)$ we have

$$E[X] \leq 2^n \exp(-n(1+\epsilon)\ln 2) = 2^{-\epsilon n}.$$

Therefore

$$\Pr[X \neq 0] \leq E[X] \leq 2^{-\epsilon n}.$$

□

3 Lower bound

We next show $I = I_{n,p}$ is satisfiable below the threshold. The proof is by the second moment method.

Claim 2 *Let $p_0 = \frac{\ln 2}{qn^{k-1}}$. Then for any $p \le (1-\epsilon)p_0$, for any Q with $|Q| = q$, we have*

$$\lim_{n\to\infty} \Pr[I_{n,p} \textit{ is satisfiable}] = 1.$$

Proof As above, let X denote the number of assignments satisfying I. We begin by calculating the second moment of X. Let $Q_i = \{\{b, b'\} \in Q \times Q\colon\ dist(b, b') = i\}$, where $dist(b, b')$ is the Hamming distance between b and b' (in other words, Q_i is the set of pairs of bad assignments which differ in i places). Let $q_i = |Q_i|$. Note that $q_0 = q$ and $q_k \le q/2$.

$$\begin{aligned}
E[X^2] &= \sum_{\phi} \Pr[\phi \text{ satisfies } I] \sum_{\phi'} \Pr[\phi' \text{ satisfies } I \mid \phi \text{ satisfies } I] \\
&= \sum_{\phi} \Pr[\phi \text{ satisfies } I] \sum_{s=0}^{n} \binom{n}{s} \Pr\left[\phi' \text{ satisfies } I \,\middle|\, {\scriptstyle \phi \text{ satisfies } I \atop dist(\phi,\phi')=n-s}\right] \\
&= \sum_{\phi} (1-p)^{qn^k} \sum_{s=0}^{n} \binom{n}{s} (1-p)^{qn^k - \sum_{i=0}^{k} q_i s^{k-i}(n-s)^i} \\
&= 2^n (1-p)^{qn^k} \sum_{s=0}^{n} \binom{n}{s} (1-p)^{qn^k - \sum_{i=0}^{k} q_i s^{k-i}(n-s)^i}.
\end{aligned}$$

where the probabilities in the second to last line follow since there are qn^k candidate clauses which are bad for assignment ϕ, qn^k which are bad for assignment ϕ', and $\sum_{i=0}^{k} q_i s^{k-i}(n-s)^i$ which are bad for both ϕ and ϕ'.

We now observe that the ratio $E[X^2]/E[X]^2$ is the expected value of a different random variable:

$$\begin{aligned}
\frac{E[X^2]}{E[X]^2} &= \sum_{s=0}^{n} \binom{n}{s} 2^{-n} (1-p)^{-\sum_{i=0}^{k} q_i s^{k-i}(n-s)^i} \\
&= E\left[(1-p)^{-\sum_{i=0}^{k} q_i S^{k-i}(n-S)^i}\right] \\
&= E\left[\left(1 + \frac{p}{1-p}\right)^{\sum_{i=0}^{k} q_i S^{k-i}(n-S)^i}\right],
\end{aligned}$$

where $S \sim B(n, 1/2)$.

Letting $Y = \left(1+\frac{p}{1-p}\right)^{\sum_{i=0}^{k} q_i S^{k-i}(n-S)^i}$, we bound $E[Y]$ in 3 parts using conditional expectations:

$$E[Y] \le \sum_{i=1}^{3} E[Y \mid \eta_{i-1} \le |n/2 - S| \le \eta_i] \Pr[\eta_{i-1} \le |n/2 - S| \le \eta_i],$$

where

$$\eta_0 = 0 \quad \eta_1 = \epsilon\frac{n}{2} \quad \eta_2 = \frac{n}{2}\left(1 - \frac{\epsilon}{\log n}\right) \quad \eta_3 = \frac{n}{2}.$$

In the following, we will rely on the fact that $\sum_{i=0}^{k} q_i = q(q+1)/2 < q^2$.

First Term: Provided $k \ge 2\log_\alpha n$ where $\alpha = \frac{2}{1+\epsilon}$, we have

$$\begin{aligned}
&E\left[Y \,\middle|\, \eta_0 \le |n/2 - S| \le \eta_1\right] \Pr\left[\eta_0 \le |n/2 - S| \le \eta_1\right] \\
&\quad \le \left(1 + \frac{p}{1-p}\right)^{q^2\left(\frac{1}{2}n(1+\epsilon)\right)^k} \\
&\quad \le \exp\left(n\frac{q\ln 2(1-\epsilon)}{1-p}\left(\frac{1+\epsilon}{2}\right)^k\right) \\
&\quad = 1 + o(1).
\end{aligned}$$

Second Term: By the standard Chernoff bound, $\Pr\left[\eta_1 \le |n/2 - S| \le \eta_2\right] \le 2e^{-n\epsilon^2/3}$. So provided $k \ge \left(\frac{2}{\epsilon}\ln\frac{3q}{\epsilon^2}\right)\log n$ we have

$$\begin{aligned}
&E[Y \mid \eta_1 \le |n/2 - S| \le \eta_2] \Pr[\eta_1 \le |n/2 - S| \le \eta_2] \\
&\quad \le \left(1 + \frac{p}{1-p}\right)^{q^2\left(n(1-\frac{\epsilon}{2\log n})\right)^k} e^{-n\epsilon^2/3} \\
&\quad \le \exp\left(n\frac{q\ln 2(1-\epsilon)}{1-p}\left(1 - \frac{\epsilon}{2\log n}\right)^k - n\epsilon^2/3\right) \\
&\quad = o(1).
\end{aligned}$$

Third Term: Note that $q_k \le q$. So for $\eta_2 \le |n/2 - S| \le \eta_3$ we have

$$\sum_{i=0}^{k} q_i S^{k-i}(n-S)^i \le qn^k + q\left(\frac{n}{\log n}\right)^k + \sum_{i=1}^{k-1} q_i S^{k-i}(n-S)^i \le (q + q^2/\log n)n^k,$$

and

$$
\begin{aligned}
&E\left[Y \,\middle|\, \eta_2 \le |n/2 - S| \le \eta_3\right] \Pr[\eta_2 \le |n/2 - S| \le \eta_3] \\
&\le \left(1 + \frac{p}{1-p}\right)^{n^k\left(q + q^2/\log n\right)} 2\binom{n}{n/2+\eta_2} 2^{-(n/2+\eta_2)} \\
&\le 2e^{n\frac{\ln 2(1-\epsilon)}{1-p}(1+q/\log n)} n^{n\frac{\epsilon}{2\log n}} 2^{-n(1-\frac{\epsilon}{2\log n})} \\
&= 2^{1+n(1-\epsilon)(1+o(1))+n\frac{\epsilon}{2}-n(1-\frac{\epsilon}{2\log n})} \\
&= 2^{-\frac{\epsilon}{2}n(1-o(1))} \\
&= o(1).
\end{aligned}
$$

Putting the parts together and using the second moment inequality, we have

$$
\Pr[X \neq 0] \ge \frac{E[X]^2}{E[X]^2} \ge 1 - o(1).
$$

□

4 Acknowledgments

Many thanks go to Alan Frieze for suggesting this problem and also for helpful comments and discussion.

References

[1] D. Achlioptas, Setting two variables at a time yields new lower bound for random 3-SAT, *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing* (2000) 28-37.

[2] D. Achlioptas, A.D. Chtcherba, G. Istrate, and C. Moore, The Phase Transition in 1-in-k SAT and NAE 3-SAT, *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms* (2001) 721-722.

[3] D. Achlioptas and C. Moore, The two-colorability of random hypergraphs, *6th International Workshop on Randomization and Approximation Techniques in Computer Science* (2002).

[4] D. Achlioptas and G. Sorkin, Optimal myopic algorithms for random 3-SAT, *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computing* (2000) 590-600.

[5] E. Ben-Sasson and Y. Bilu, A gap in average proof complexity, *Electronic Colloquium on Computational Complexity* 9 (2002) #R3.

[6] B. Bollobás, C. Borgs, J. Chayes, J.H. Kim, and D.B. Wilson, The scaling window of the 2-SAT transition. Manuscript, 1999.

[7] A.Z. Broder, A.M. Frieze, and E. Upfal, On the satisfiability and maximum satisfiability of random 3-CNF formulas, *Proceedings of the 4th Annual ACM-SIAM Symposium on Discrete Algorithms* (1993) 322-330.

[8] M.T. Chao and J. Franco, Probabilistic analysis of 2 heuristics for the 3-satisfiability problem, *SIAM Journal of Computing* 15 (1986) 1106-1118.

[9] M.T. Chao and J. Franco, Probabilistic analysis of a generalization of the unit-clause literal selection heuristic for the k satisfiability problem, *Information Science* 51 (1990) 289-314.

[10] V. Chvátal and B. Reed, Mick gets some (the odds are on his side). *Proceedings of the 33rd Annual IEEE Symposium on Foundations of Computer Science* (1992) 620-627.

[11] O. Dubois, Y. Boufkhad, and J. Mandler, Typical random 3-SAT formulae and the satisfiability threshold, *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms* (2000) 126-127.

[12] W. Fernandez de la Vega, On Random 2-sat, (1992), manuscript.

[13] E. Friedgut, Necessary and sufficient conditions for sharp thresholds of graph properties, and the k-SAT problem. *J. Amer. Math. Soc. 12* (1999) 1017-1054.

[14] A.M. Frieze and G. Sorkin, A note on random 2-SAT with prescribed literal degrees, *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms* (2002).

[15] A.M. Frieze and S. Suen, Analysis of two simple heuristics on a random instance of k-SAT. *J. Algorithms 20*, 2 (1996), 312-355.

[16] A.M. Frieze and N.C. Wormald, Random k-SAT: A tight threshold for moderately growing k, *Proceedings of the 5th International Symposium on Theory and Applications of Satisfiability Testing* (2002) 1-6.

[17] A. Goerdt, A threshold for unsatisfiability, *Journal of Computer and System Science* 33 (1996) 469-486.

[18] S. Janson, Y. Satmatiou, and M. Vamvakari, Bounding the unsatisfiability threshold of random 3-SAT, *Random Structures and Algorithms* 17 (2000) 103-116.

[19] A.C. Kaporis, L.M. Kirousis, and E.G. Lalas, The probabilistic analysis of a greedy satisfiability algorithm, *Proceedings of the 5th International Symposium on Theory and Applications of Satisfiability Testing* (2002) 362-376.

[20] L.M. Kirousis, E. Kranakis, D. Krizanc, and Y.C. Stamatiou, Approximating the unsatisfiability threshold of random formulas, *Random Structures and Algorithms* 17 (2000) 103-116.

[21] Y. Verhoeven, Random 2-SAT and unsatisfiability, *Information Processing Letters* 72 (1999), 119-123.

[22] K. Xu and W. Li, Exact phase transitions in random constraint satisfaction problems, *Journal of Artificial Intelligence Research* 12 (2000) 93-103.