

Satisfiability Threshold of the Skewed Random k -SAT

Danila A. Sinopalnikov

Department of Mathematics and Mechanics,
St.Petersburg State University, Russia
dasinopalnikov@yahoo.com

Abstract. We consider the satisfiability phase transition in skewed random k -SAT distributions. It is known that the random k -SAT model, in which the instance is a set of m k -clauses selected uniformly from the set of all k -clauses over n variables, has a satisfiability phase transition at a certain clause density. The essential feature of the random k -SAT is that positive and negative literals occur with equal probability in a random formula. How does the phase transition behavior change as the relative probability of positive and negative literals changes?

In this paper we focus on a distribution in which positive and negative literals occur with different probability. We present empirical evidence for the satisfiability phase transition for this distribution. We also prove an upper bound on the satisfiability threshold and a linear lower bound on the number of literals in satisfying partial assignments of skewed random k -SAT formulas.

1 Introduction

The problem to decide whether a given propositional formula has a satisfying truth assignment (SAT) is one of the first for which NP -completeness was proven. Nowadays it attracts much attention, since many hard combinatorial problems in areas including planning [11, 12] and finite mathematics [17] can be naturally encoded and studied as SAT instances.

While the SAT hardness is determined by the difficulty of solving an instance of the problem in the worst case, the scientific interest is also focused on randomly chosen SAT instances in attempt to determine the typical-case complexity. The choice of the probabilistic distribution is critical for the significance of such a study. In particular, it was proven that in some probabilistic spaces a random formula is easy-to-decide with high probability [6, 8, 14]. To date, most of the research in the field is concentrated on the random k -SAT model $RD(n, k, m)$, which appears to be more robust in this respect.

Let X be a set of n boolean variables; a proper k -clause is a disjunction of k distinct and non-contradictory literals corresponding to variables in X . Under the random k -SAT distribution $RD(n, k, m)$, a random formula $F_k(n, m)$ is built by selecting uniformly, independently and with replacement m clauses from the set of all proper k -clauses over n variables.

Numerous empirical results suggested that $RD(n, k, m)$ exhibits a phase transition behavior as the clause density $\delta = m/n$ changes [2, 13, 15, 16]. When the number of variables tends to infinity and the clause density remains constant, the random formula $F_k(n, \delta n)$ is almost surely satisfiable for low clause densities while for higher clause densities it is almost surely unsatisfiable. The satisfiability threshold conjecture asserts that for every $k \geq 2$ there exists δ_k such that

$$\begin{aligned} \delta_k &= \sup\{\delta \mid \lim_{n \rightarrow \infty} P(F_k(n, \delta n) \text{ is satisfiable}) = 1\} = \\ &= \inf\{\delta \mid \lim_{n \rightarrow \infty} P(F_k(n, \delta n) \text{ is satisfiable}) = 0\} . \end{aligned} \tag{1}$$

This conjecture was settled for $k = 2$ with $\delta_2 = 1$ by Chvátal and Reed [3], Goerdt [7] and Fernandez de la Vega [4]. More recently the asymptotic form of the conjecture with $\delta_k = 2^k \ln 2(1 + o(1))$ was established by Achlioptas and Peres [1].

The essential feature of the random k -SAT is that positive and negative literals occur in a formula with equal probability. In this paper we consider satisfiability of random formulas from the skewed random k -SAT distribution, in which positive and negative literals occur with different probability. To the best of our knowledge, there has not been much work on this generalization

of random k -SAT. The paper answers the question whether the satisfiability phase transition manifests in the skewed distributions and presents a proof of an upper bound on the threshold location for skewed random k -SAT. We expect that this study will provide further insight into the nature of the phase transition phenomenon in the boolean satisfiability problem.

We also investigate the minimal number of literals in a satisfying partial assignment of a random formula. This study is motivated by the fact that if a random k -CNF formula is satisfiable then it has an exponential number of satisfying assignments with high probability [10]. On the other hand, it is known that k -CNF formulas with many satisfying assignments have short satisfying partial assignments [9]. This might imply that a random formula with clause density far below the satisfiability threshold is likely to have short satisfying partial assignments. In this paper we elaborate on this intuition and prove a linear lower bound on the number of literals in a satisfying partial assignment of a random formula for skewed and plain random k -SAT distributions.

The paper is organized as follows. Section 2 contains basic definitions. In Section 3 we present the results of our empirical evaluation of the phase transition behavior of the skewed distribution. We formulate and prove the main results in Section 4. Section 5 concludes the paper.

2 Basic Definitions

Let X be a set of n boolean variables. A *literal* is a variable (positive literal) or its negation (negative literal). A variable and its negation are *contradictory* literals. A k -clause is an ordered collection of k literals. A *clause density* of a formula F is the ratio of number of clauses in F to the number of variables. A *complementary* formula for F is a formula obtained from F by replacing all literals with their negations.

A *partial assignment* σ is an arbitrary set of non-contradictory literals. The *size* of a partial assignment is the number of literals in it. A *complete* assignment on n variables is an assignment of size n . A *complementary* partial assignment for σ is an assignment obtained from σ by replacing of all literals with their negations. A partial assignment σ is *satisfying* for a formula F if in each clause of F there is at least one literal from σ .

Let $n, k, m \in \mathbb{N}$, $p \in (0, 1)$. $SD(n, k, m, p)$ denotes the skewed random k -SAT distribution, where a random formula is obtained by building m k -clauses as follows: for each clause we select independently, uniformly and with replacement k variables from the set of n boolean variables; then for each selected variable we take a positive literal with probability p and a negative literal otherwise and add it to the clause. $A(n, m)$ denotes the set of all partial assignments of size m over n variables.

For a propositional formula F and a partial assignment σ we use the following notations:

$F \in SAT$ means F is satisfiable,

$\bar{F}, \bar{\sigma}$ denote the complementary formula and assignment respectively,

$\sigma \in S(F)$ means σ satisfies F ,

$pl(\sigma)$ denotes the number of positive literals in σ ,

$minsat(F)$ denotes the minimum size of a satisfying partial assignment of F .

We omit floors and ceilings when it doesn't result in confusion.

For $\lambda \in [0, 1]$ denote $H(\lambda) = \lambda^\lambda(1 - \lambda)^{1-\lambda}$. In the paper we will use the following standard bounds

$$\binom{n}{\lambda n} \leq H(\lambda)^{-n}, \quad (2)$$

$$(1 + x)^\alpha \geq 1 + \alpha x \text{ for } \alpha > 1, x > 0, \quad (3)$$

$$(1 - \alpha)^\beta \leq e^{-\alpha\beta} \text{ for } \alpha \in [0, 1], \beta > 0. \quad (4)$$

3 The Phase Transition Behavior of the Skewed Random k -SAT

In this section we present empirical evidence that random k -SAT instances undergo a satisfiability phase transition even if positive and negative literals occur with different probabilities. We took skewed random 3-SAT distributions at four levels of skewness: $p = 0.5$ (the plain random 3-SAT); 0.4; 0.3; 0.2.

Figure 1 shows the sample probability that a random clause is unsatisfiable at the particular clause density and skewness. There are four curves, one for each level of skewness. Along the horizontal axis is the clause density. Each sample contains 500 random 3-SAT instances on 200 variables.

We observe that all four distributions exhibit the satisfiability phase transition. The transition manifests at a greater density and becomes less distinct as the skewness of the distribution increases and the number of variables remains constant.

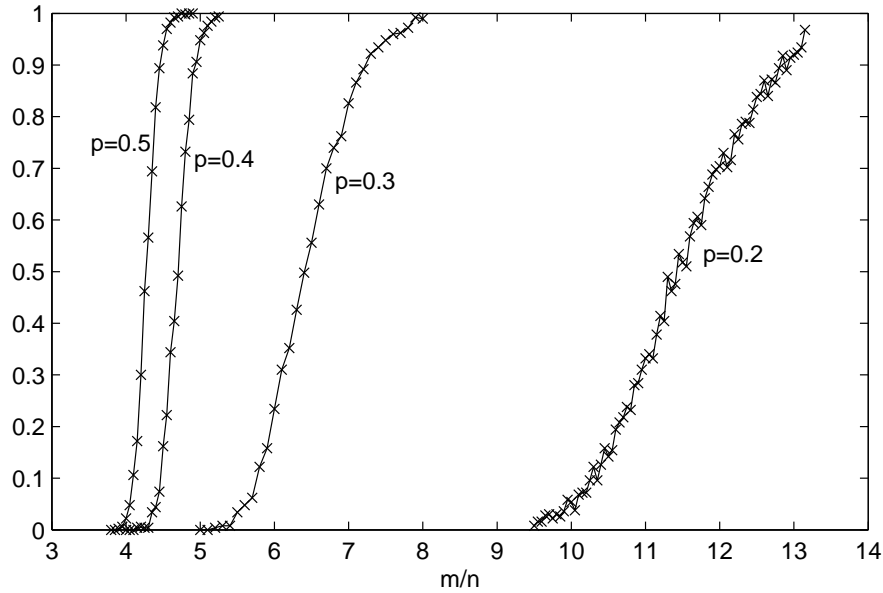


Fig. 1. The probability that a skewed random 3-SAT formula is unsatisfiable, as a function of the clause density m/n (p - the skewness of the distribution)

4 A Lower Bound on the Size of Satisfying Partial Assignments of a Skewed Random k -SAT Formula

Given a propositional formula over n variables, we are interested in finding a satisfying partial assignment of the minimum size, that corresponds to the largest cluster of satisfying complete assignments. In this section we prove a linear lower bound on the size of a satisfying partial assignment of a random formula from a skewed random k -SAT distribution.

Let's consider a fixed ratio λ between the size of a partial assignment and the total number of variables, a skewed random k -SAT distribution $SD(n, k, \delta n, \pi)$ with fixed k and p and a random formula F from that distribution. The main theorem below provides an equation that defines a clause density δ such that for all larger density values the probability that a formula F has a satisfying partial assignment of size λn tends to zero as the number of variables n tends to infinity.

For $k \in \mathbb{N}$, $\lambda \in (0, 1]$, $q \in (1 - \lambda, 1 - \lambda/2]$ we consider the following equation

$$\frac{H(\lambda)}{(1 - q^k)^\delta} = \left(1 + \exp \left(-\frac{k\delta q^k}{1 - q^k} \cdot \frac{2(1 - q) - \lambda}{q} \right) \right)^\lambda. \quad (5)$$

Lemma 1 Let $k \in \mathbb{N}$, $\lambda \in (0, 1]$, $q \in (1 - \lambda, 1 - \lambda/2]$.

Then there exists a single δ_0 that satisfies (5), $\delta_0 > 0$ and for all δ

$$\delta > \delta_0 \Leftrightarrow \frac{H(\lambda)}{(1 - q^k)^\delta} > \left(1 + \exp \left(-\frac{k\delta q^k}{1 - q^k} \cdot \frac{2(1 - q) - \lambda}{q} \right) \right)^\lambda. \quad (6)$$

Proof. Take arbitrary $k \in N$, $\lambda \in (0, 1]$, $q \in (1 - \lambda, 1 - \lambda/2]$.

Consider two functions

$$f(\delta) = \frac{H(\lambda)}{(1 - q^k)^\delta}, \tag{7}$$

$$g(\delta) = \left(1 + \exp\left(-\frac{k\delta q^k}{1 - q^k} \cdot \frac{2(1 - q) - \lambda}{q}\right) \right)^\lambda. \tag{8}$$

$f(\delta)$ is a continuous, strictly increasing function of δ , $\lim_{\delta \rightarrow +\infty} f(\delta) = +\infty$, $f(0) = H(\lambda) \leq 1$. $g(\delta)$ is a continuous, decreasing function of δ , $g(0) = 2^\lambda > 1$. This implies the statement of the lemma. \square

Definition 1 Let $k \in N$, $\lambda \in (0, 1]$, $q \in (1 - \lambda, 1 - \lambda/2]$.

Then $\Delta_s(q, k, \lambda)$ denotes the root of the equation (5)

Theorem 1 Let $\delta > 0$, $\lambda \in (0, 1]$, $\pi \in (0, 1)$, $p = \min(\pi, 1 - \pi)$, $k \in N$.

Let $F_k(n, \delta, \pi)$ be a random formula from $SD(n, k, \delta n, \pi)$.

Then

$$\delta > \Delta_s(1 - \lambda(1 - p), k, \lambda) \Rightarrow \lim_{n \rightarrow \infty} P(\text{minsat}(F_k(n, \delta, \pi)) \leq \lambda n) = 0. \tag{9}$$

Proof. Take arbitrary $k \in N$, $\lambda \in (0, 1]$, $m = \lambda n$. A partial assignment $\sigma \in A(n, m)$ satisfies a formula F if and only if the complementary assignment $\bar{\sigma}$ satisfies the complementary formula \bar{F} . So it is sufficient to prove the theorem for $\pi \in (0, 1/2]$, $p = \pi$. Let $q = 1 - \lambda(1 - p)$.

Let X_m denote the number of partial assignments of size m that satisfy $F_k(n, \delta, \pi)$. It is easy to see that

$$P(\text{minsat}(F_k(n, \delta, \pi)) \leq \lambda n) = P(X_m \geq 1) \leq E[X_m]. \tag{10}$$

Now we are going to obtain a suitable upper bound for $E[X_m]$.

$$E[X_m] = \sum_{\sigma \in A(n, m)} P(\sigma \in S(F_k(n, \delta, \pi))). \tag{11}$$

Due to the symmetry of the distribution, the probability that σ satisfies a random formula depends only on $pl(\sigma)$, so

$$E[X_m] = \sum_{z=0}^m \sum_{\substack{\sigma \in A(n, m) \\ pl(\sigma)=z}} P(\sigma \in S(F_k(n, \delta, \pi)) | pl(\sigma) = z). \tag{12}$$

Now let's compute the probability that a partial assignment with a fixed number of positive literals satisfies a random formula. Let C and l denote random formulas from $SD(n, k, 1, \pi)$ and $SD(n, 1, 1, \pi)$ respectively (C and l can be viewed as a random clause and a random literal from $F_k(n, \delta, \pi)$). Then, since clauses in $F_k(n, \delta, \pi)$ are independent,

$$P(\sigma \in S(F_k(n, \delta, \pi)) | pl(\sigma) = z) = (1 - P(\sigma \notin S(C) | pl(\sigma) = z))^{\delta n}. \tag{13}$$

Let $var(C)$ and $var(\sigma)$ denote the set of variables in C and σ respectively, $overlap(C, \sigma) = |var(C) \cap var(\sigma)|$ - the number of variables shared by C and σ . Note that $overlap(C, \sigma) \leq k$ and thus

$$\begin{aligned} &P(\sigma \notin S(C) | pl(\sigma) = z) = \\ &= \sum_{j=0}^k P(\sigma \notin S(C) | overlap(C, \sigma) = j \ \& \ pl(\sigma) = z) P(overlap(C, \sigma) = j). \end{aligned} \tag{14}$$

Literals in a random clause are independent, so

$$P(overlap(C, \sigma) = j) = \binom{k}{j} \left(\frac{m}{n}\right)^j \left(\frac{n-m}{n}\right)^{k-j}, \tag{15}$$

$$\begin{aligned} &P(\sigma \notin S(C) | overlap(C, \sigma) = j \ \& \ pl(\sigma) = z) = \\ &= P(\sigma \notin S(l) | pl(\sigma) = z \ \& \ var(l) \in var(\sigma))^j. \end{aligned} \tag{16}$$

Conditioning on the event of l being positive, we compute

$$P(\sigma \notin S(l) | pl(\sigma) = z \ \& \ \text{var}(l) \in \text{var}(\sigma)) = p + (1 - 2p)z/m. \tag{17}$$

Plugging this into (14), we can return to (12)

$$E[X_m] = \sum_{z=0}^m \sum_{\substack{\sigma \in A(n,m) \\ pl(\sigma)=z}} \left(1 - \left(1 - \frac{m}{n} \cdot \left(1 - p - \frac{z}{m}(1 - 2p) \right) \right)^k \right)^{\delta n}. \tag{18}$$

Using (3), the expectation can be bounded from above as follows

$$E[X_m] \leq (1 - q^k)^{\delta n} \binom{n}{\lambda n} \times \sum_{z=0}^{\lambda n} \binom{\lambda n}{z} \left(1 - \frac{q^k}{1 - q^k} \cdot k \cdot \frac{z}{n} \cdot \frac{2(1 - q) - \lambda}{q} \right)^{\delta n}. \tag{19}$$

Now we sequentially apply (4) and (2) to obtain a simpler bound

$$E[X_m] \leq H(\lambda)^{-n} (1 - q^k)^{\delta n} \cdot \left(1 + \exp \left(-\frac{k\delta q^k}{1 - q^k} \cdot \frac{2(1 - q) - \lambda}{q} \right) \right)^{\lambda n}. \tag{20}$$

We take $p \in (0, 1/2]$, so $q \in (1 - \lambda, 1 - \lambda/2]$ and we can use Lemma 1 to state that for all $\delta > \Delta_s(1 - \lambda(1 - p), k, \lambda)$ $E[X_m] = o(1)$ as $n \rightarrow \infty$. This statement together with (10) proves the theorem. \square

For $\lambda = 1$ we get an upper bound on the satisfiability threshold for a skewed random k -SAT formula: for any $\delta, \pi \in (0, 1)$ and $p = \min(\pi, 1 - \pi)$

$$\delta > \Delta_s(p, k, 1) \Rightarrow \lim_{n \rightarrow \infty} P(F_k(n, \delta, \pi) \in SAT) = 0. \tag{21}$$

For $p = 1/2$ we get a linear lower bound on the size of satisfying partial assignments for a plain random k -SAT formula: for any δ and $\lambda \in (0, 1]$

$$\delta > \frac{\ln H(\lambda) - \lambda \ln 2}{\ln \left(1 - (1 - \lambda/2)^k \right)} \Rightarrow P(\text{minsat}(F_k(n, \delta)) \leq \lambda n) = o(1) \text{ as } n \rightarrow \infty. \tag{22}$$

Setting $\lambda = 1$ we obtain the known upper bound for the random k -SAT satisfiability threshold (see [5])

$$\Delta_s(1/2, k, 1) = -\frac{\ln 2}{\ln(1 - 1/2^k)}. \tag{23}$$

Table 1 provides approximate values of $\Delta_s(1 - \lambda(1 - p), k, \lambda)$ for $k = 3$ and $k = 4$ respectively.

Table 1. Approximate values of $\Delta_s(1 - \lambda(1 - p), 3, \lambda)$.

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
p=0.2	0.256	0.670	1.266	2.108	3.314	5.088	7.789	12.000	18.263	22.385
p=0.3	0.238	0.605	1.110	1.783	2.675	3.855	5.413	7.419	9.705	10.204
p=0.4	0.220	0.545	0.976	1.526	2.216	3.068	4.095	5.269	6.406	6.346
p=0.5	0.203	0.489	0.860	1.325	1.897	2.592	3.414	4.335	5.215	5.191

5 Conclusion

In this paper we considered a skewed random k -SAT distribution and investigated the phase transition behavior in this model. Empirical evidence for the satisfiability phase transition was

presented. Further experiments suggest that even for a highly skewed random k -SAT distribution the phase transition becomes sharp as the number of variables increases.

We proved an upper bound on the satisfiability threshold and a lower bound on the number of literals in satisfying partial assignments for a skewed random k -SAT formula. For the considered skewed distribution there is still a large gap between the observed threshold location and the proved upper bound, so better bounds are still to be obtained. Lower bounds on the threshold and upper bounds on the minimum number of literals in a satisfying partial assignment of a skewed random k -SAT formula are needed to complete the picture.

Another interesting direction is to evaluate the computational hardness of skewed random k -SAT formulas with respect to the skewness of the distribution for a fixed clause density. The possible candidates for the maximum hardness are the non-skewed distribution and the skewed distribution that undergoes the satisfiability phase transition at this clause density.

6 Acknowledgments

The author would like to thank Edward Hirsch for bringing this problem to his attention and for valuable comments.

References

1. D. Achlioptas and Y. Peres. The threshold for random k -SAT is $2^k(\ln 2 + o(1))$. Submitted for publication.
2. P. Cheeseman, B. Kanefsky and W. Taylor. Where the really hard problems are. In *12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, volume 1, pages 331-337. Morgan Kaufman, 1991.
3. V. Chvátal and B. Reed. Mick gets some (the odds are on his side). In *33th Annual Symposium on Foundation of Computer Science (Pittsburg, PA, 1992)*, pages 620-627. IEEE Comput. Soc. Press, Los Alamitos, CA, 1992.
4. W. Fernandez de la Vega. On random 2-SAT. Manuscript, 1992.
5. J. Franco and M. Paull. Probabilistic analysis of the Davis-Putnam procedure for solving satisfiability. *Discrete Applied Mathematics*, 5, pages 77-87, 1983.
6. J. Franco and R. Swaminathan. Average case results for satisfiability algorithms under the random clause width model. *Annals of Mathematics and Artificial Intelligence* 20(1-4), pages 357-391, 1997.
7. A. Goerdt. A threshold for unsatisfiability. *J. Comput. System Sci.*, 53(3), pages 469-486, 1996.
8. A. Goldberg. Average case complexity of the satisfiability problem. In *Proc. of 4th Workshop on Automated Deduction*, 1979.
9. E. A. Hirsch. A Fast Deterministic Algorithm for Formulas That Have Many Satisfying Assignments. *Logic Journal of the IGPL*, Vol.6, No.1, Oxford University Press, pages 59-71, 1998.
10. A. Kamath, R. Motwani, K. Palem and P. Spirakis. Tail bounds for occupancy and the satisfiability threshold conjecture. *Random structures and algorithms* 7(1), pages 59-80, 1995.
11. H. Kautz and B. Selman. Planning as satisfiability. In *Proc. of ECAI'92*, volume 2, pages 359-363. John Wiley & Sons, 1996.
12. H. Kautz and B. Selman. Pushing the envelope: planning, propositional logic and stochastic search. In *Proc. of AAAI'96*, volume 2, pages 1194-1201. MIT Press, 1996.
13. S. Kirkpatrick and B. Selman. Critical behavior in the satisfiability of random boolean expressions. *Science* 264, pages 1297-1301, 1994.
14. Elias Koutsoupias and Christos H. Papadimitriou. On the greedy algorithm for satisfiability. *IPL*, 43(1):53-55, 1992.
15. T. Larrabee and Y. Tsuji. Evidence for satisfiability threshold for random 3CNF formulas. In *Proc. of the AAAI Symposium on Artificial Intelligence and NP-hard problems*, 112. 1993.
16. D. Mitchell, B. Selman and H. Levesque. Hard and easy distributions of SAT problems. In *Proc. of 10th National Conference on Artificial Intelligence*, pages 459-465. AAAI Press, Menlo Park, CA, 1992.
17. H. Zhang, M. Bonacina and J. Hsiang. PSATO: a distributed propositional prover and its application to quasigroup problems. *Journal of Symbolic Computation*, 21(4), pages 543-560, 1996.