Locally satisfiable formulas

Daniel Král'*

Abstract

A CNF formula ψ is k-satisfiable if each k clauses of ψ can be satisfied simultaneously. Let π_k be the largest real number such that for each k-satisfiable formula ψ with variables x_i, there are probabilities p_i with the following property: If each variable x_i is chosen randomly and independently to be true with the probability p_i, then each clause of ψ is satisfied with the probability at least π_k.

We determine the numbers π_k and design a linear-time algorithm which given a formula ψ either outputs that ψ is not k-satisfiable or finds probabilities p_i such that each clause of ψ is satisfied with the probability at least π_k. Our approach yields a robust linear-time deterministic algorithm which finds for a k-satisfiable formula a truth assignment satisfying at least the fraction of π_k of the clauses.

A related parameter is r_k which is the largest ratio such that for each k-satisfiable CNF formula with m clauses, there is a truth assignment which satisfies at least $r_k m$ of its clauses. It was known that $\pi_k = r_k$ for $k = 1, 2, 3$. We compute the ratio r_4 and show $\pi_4 \neq r_4$. We also design a linear-time algorithm which finds a truth assignment satisfying at least the fraction r_4 of the clauses for 4-satisfiable formulas.

1 Introduction

CNF formulas have a prominent position in computer science because of their essential role in many hardness reductions and constructions, see [1, 2, 6, 11]. We study an extremal problem which relates local and global satisfiability of CNF formulas. A CNF formula is *k-satisfiable* if any k clauses of it can be

*Institute for Theoretical Computer Science, Charles University, Malostranské náměstí 25, 118 00 Prague 1, Czech Republic. E-mail: `kral@kam.mff.cuni.cz`.
Institute for Theoretical Computer Science (ITI) is supported by Ministry of Education of Czech Republic as project LN00A056.

simultaneously satisfied. The notion of k-satisfiable formulas was introduced by Lieberherr and Specker [7, 8]; a separate section (20.6) is devoted to this concept in a recent monograph on extremal combinatorics by Jukna [5].

One of the problems which we address is the following: Let ψ be a k-satisfiable formula with the variables $x_1, \ldots, x_n$. What is the largest number $\pi(\psi)$ for which there are probabilities $p_1, \ldots, p_n$ such that if each x_i is chosen randomly and independently to be true with the probability p_i, then each clause of ψ is satisfied with the probability at least $\pi(\psi)$? Observe that $\pi(\psi) = 1$ iff ψ is satisfiable. Probabilities p_i may be also understood to be a fractional truth assignment. Let π_k be the largest number such that $\pi(\psi) \geq \pi_k$ for each k-satisfiable formula ψ. In this paper, we compute the numbers π_k for all $k \geq 1$ and discover a surprising connection between them and the voting paradox studied by Usiskin [10].

Based on our structural results, we design a linear-time algorithm which for a k-satisfiable formula with m clauses finds probabilities and constructs a truth assignment such that at least $\pi_k \cdot m$ clauses are satisfied. The ratios of satisfied clauses of our algorithm dominate the ratios of the previous algorithm (using a similar technique) for the problem by Trevisan [9]. By the definition of π_k, our algorithm is optimal in sense that no algorithm based on fractional truth assignments can guarantee a larger fraction of satisfied clauses in the class of k-satisfiable formulas. In addition our algorithm is robust, i.e., its input may be any formula and the algorithm either constructs a truth assignment or outputs that the input formula is not k-satisfiable.

The numbers π_i are actually related in this paper to the voting paradox studied by Usiskin [10] which is described in the next: Let X_i for $i = 1, \ldots, k$ be independent real-valued random variables. We define $\mu(X_1, \ldots, X_k)$ to be the following quantity:

$$\min\{\mathbb{P}(X_1 > X_2), \mathbb{P}(X_2 > X_3), \ldots, \mathbb{P}(X_{k-1} > X_k), \mathbb{P}(X_k > X_1)\}$$

The k-th Usiskin number b_k is the largest real number such that there exist random variables $X_1, \ldots, X_k$ with $\mu(X_1, \ldots, X_k) \geq b_k$. The values of b_k were determined by Usiskin [10]. In addition, he showed that $\lim_{k\to\infty} b_k = 3/4$. We prove that $b_{k+1} = \pi_k$ for all $k \geq 1$. Hence, the values of π_k are determined for all k (they were unknown for $k \geq 4$) and $\lim_{k\to\infty} \pi_k = 3/4$ (however, this was previously proved by Trevisan [9]).

So far, another modification of the problem (which is also considered in this paper) was mainly studied: Let r_k be the largest real number such that each k-satisfiable CNF formula with m (not necessarily distinct) clauses has a truth assignment which satisfies at least $r_k m$ clauses. Clearly, $r_1 = 1/2$ and $\pi_k \leq r_k$.

The latter inequality is because the expected number of satisfied clauses for optimal probabilities p_i is at least $\pi_k m$ and hence there is a truth assignment which satisfies at least $\pi_k m$ clauses.

We briefly survey previously known results on the values of r_k; the reader is also welcomed to see the section 20.6 in the monograph [5]. The study of the problem was started by Lieberherr and Specker. They showed $r_2 = \frac{\sqrt{5}-1}{2} \approx 0.6180$ [7] and they consequently established $r_3 = 2/3$ [8]. Later, Yannakakis [12] simplified proofs of the lower bounds for r_2 and r_3 using a probabilistic argument. Huang and Lieberherr [4] studied the asymptotic behavior of r_k and proved $\lim\limits_{k\to\infty} r_k \le 3/4$. The asymptotic behavior of r_k was completely determined by Trevisan [9] by showing $\lim\limits_{k\to\infty} r_k = 3/4$. We compute the value of r_4. Namely, we prove:

$$r_4 = \frac{3}{5 + \sqrt[3]{\frac{3\sqrt{69}-11}{2}} - \sqrt[3]{\frac{3\sqrt{69}+11}{2}}} \approx 0.6992$$

In addition, our proof gives a linear-time algorithm for 4-satisfiable formulas which finds an assignment satisfying at least the fraction r_4 of the clauses.

The just introduced problem is actually studied for CNF formulas where the non-negative weights are assigned to the clauses. This is the same problem because the weights of the clauses can be simulated (with a negligible error) by repeating the clauses several times in the formula (allowing an exponential blow-up in the size). If ψ is a formula, let $w(\psi)$ be the sum of the weights of all its clauses and $w_0(\psi)$ be the maximum weight of the clauses of ψ which can be simultaneously satisfied. Let Ψ_k be the set of all k-satisfiable CNF formulas. Using the just defined notation, we can write:

$$r_k = \inf_{\psi\in\Psi_k} \frac{w_0(\psi)}{w(\psi)}$$

In this paper, we compute the value of this infimum for $k = 4$ which substantially differs from the cases $k < 4$: In the cases of $k = 1, 2, 3$, Yannakakis found a probability distribution for each k-satisfiable formula ψ such that each clause was satisfied with probability at least r_k (and hence $\pi_k = r_k$ for $k < 4$). Thus the expected weight of the satisfied clauses with respect to such a distribution is at least $r_k w(\psi)$ regardless the actual choice of the weight function. Such a probability distribution may be considered as universal for a formula ψ for all possible choices of weights of the clauses. However such a distribution does not exist in general for 4-satisfiable formulas (hence $\pi_4 \neq r_4$) as shown in Proposition 1.

The paper is organized as follows. We recall certain equalities and inequalities which hold for Usiskin's numbers b_k in Section 2. Then, we show that $b_{k+1} \le \pi_k$ in Section 3. We complete the proof of the equality $b_{k+1} = \pi_k$ in Section 4 by establishing the opposite inequality $b_{k+1} \ge \pi_k$. Next, we establish several identities for the claimed value of r_4 in Section 5, in particular the equality $r_4 = 1/(2 - p_0^2)$ where p_0 is a root of the equation $p_0^3 + p_0^2 - 1 = 0$. The matching lower and upper bounds on r_4 are proved in Section 6 and Section 7, respectively. Finally, we utilize the obtained results to design our algorithms in Section 8.

2 Usiskin's numbers

First, we define recursively functions f_k for all non-negative integers k:

$$f_k(x) = \begin{cases} 1 & \text{for } k = 0 \text{ and} \\ 1 - \frac{1-x}{f_{k-1}(x)} & \text{otherwise.} \end{cases} \tag{1}$$

Note that the function f_1 is actually an identity, i.e., $f_1(x) = x$. Observe that the recursive definition of the functions f_k can be reversed, in particular, the following holds for all $k \ge 0$:

$$f_k(x) = \frac{1-x}{1 - f_{k+1}(x)} \tag{2}$$

Usiskin showed that the number b_k, $k \ge 2$, is the only solution x with $x \in \langle 1/2, 3/4)$ of the following equation [10]:

$$f_k(x) = 0 \tag{3}$$

which satisfies in addition that $f_i(x) > 0$ for $i = 1, \ldots, k-1$. It can be shown that $b_k < 3/4$ for all k and $\lim_{k\to\infty} b_k = 3/4$. Observe that the equality of (3) can be rewritten to the following form:

$$1 - b_k = f_{k-1}(b_k) \tag{4}$$

Based on (3), it is possible to construct an equation of degree $\lceil k/2 \rceil$ whose root is the number b_k [10]. Then, some of the Usiskin's numbers can be expressed explicitly, e.g., $b_2 = 1/2$, $b_3 = (\sqrt{5} - 1)/2$, $b_5 = \sqrt{2}/2$ and $b_{10} = \sqrt{3} - 1$.

We now state several lemmas on the values $f_i(b_k)$:

Lemma 1 *Let $k \ge 2$ be an integer. Then:*

$$1 = f_0(b_k) > f_1(b_k) > \ldots > f_{k-1}(b_k) > f_k(b_k) = 0$$

Proof: By the definition of the function f_0, we have $f_0(b_k) = 1$. On the other hand, the equality of (3) implies that $f_k(b_k) = 0$. We now show that $f_i(b_k) > f_{i+1}(b_k)$ for each $i = 0, \ldots, k-1$:

$$\begin{aligned} f_i(b_k) &> f_{i+1}(b_k) \\ f_i(b_k) &> 1 - \frac{1-b_k}{f_i(b_k)} \\ f_i(b_k)^2 - f_i(b_k) + 1/4 &> 1/4 - (1-b_k) \\ (f_i(b_k) - 1/2)^2 &> b_k - 3/4 \end{aligned}$$

The last inequality holds because its left-hand side is non-negative and its right-hand side is negative (recall that $b_k < 3/4$).

■

Lemma 2 *Let i, j and k be non-negative integers such that $i + j \leq k - 1$ and $k \geq 2$. Then:*

$$f_i(b_k) \cdot f_j(b_k) \geq 1 - b_k$$

Moreover, the equality holds only if $i + j = k - 1$.

Proof: By Lemma 1, it is enough to prove that $f_i(b_k) \cdot f_j(b_k) = 1 - b_k$ for $i + j = k - 1$. The proof proceeds by induction on i. If $i = 0$, then $f_i(b_k) = f_0(b_k) = 1$ and $f_j(b_k) = f_{k-1}(b_k) = 1 - b_k$ (the latter follows from the equality of (4)). Hence, let us assume that $i \geq 1$ and we need to show that $f_i(b_k) \cdot f_{k-1-i}(b_k) = 1 - b_k$. By the equations of (1) and (2), it is enough to prove:

$$\begin{aligned} \left(1 - \frac{1-b_k}{f_{i-1}(b_k)}\right) \cdot \frac{1-b_k}{1 - f_{k-i}(b_k)} &= 1 - b_k \\ 1 - \frac{1-b_k}{f_{i-1}(b_k)} &= 1 - f_{k-i}(b_k) \\ f_{k-i}(b_k) &= \frac{1-b_k}{f_{i-1}(b_k)} \\ f_{i-1}(b_k) \cdot f_{k-i}(b_k) &= 1 - b_k \end{aligned}$$

The last equality holds by the induction hypothesis.

■

Lemma 3 *Let i and k be non-negative integers such that $k \geq 2$. Then:*

$$f_i(b_k) + f_{k-i}(b_k) = 1$$

Proof: The proof proceeds by induction i. If $i = 0$, then $f_i(b_k) = f_0(b_k) = 1$ and $f_j(b_k) = f_k(b_k) = 0$ (the latter is from the definition of b_k based on the equality of (3)). Let us suppose now that $i > 0$:

$$\begin{aligned}
f_i(b_k) + f_{k-i}(b_k) &= 1 \\
1 - \frac{1 - b_k}{f_{i-1}(b_k)} + \frac{1 - b_k}{1 - f_{k-i+1}(b_k)} &= 1 \\
\frac{1 - b_k}{f_{k-i+1}(b_k)} &= \frac{1 - b_k}{1 - f_{i-1}(b_k)} \\
1 - f_{i-1}(b_k) &= f_{k-i+1}(b_k) \\
f_{i-1}(b_k) + f_{k-i+1}(b_k) &= 1
\end{aligned}$$

Now, the last equality follows from the induction hypothesis.

■

Lemma 4 *Let i, j and k be positive integers such that $i \leq j$ and $i + j \leq k$. Then:*

$$f_i(b_k) \cdot f_{j-1}(b_k) \leq f_{i-1}(b_k) \cdot f_j(b_k)$$

Proof: If $i = j$, then the claim trivially holds. Hence, assume that $i < j$ in the rest. In particular, $f_i(b_k) > f_j(b_k)$ by Lemma 1. Let us first rewrite the inequality from the statement to a little different form:

$$\begin{aligned}
f_i(b_k) \cdot f_{j-1}(b_k) &\leq f_{i-1}(b_k) \cdot f_j(b_k) \\
f_i(b_k) \cdot \frac{1 - b_k}{1 - f_j(b_k)} &\leq \frac{1 - b_k}{1 - f_i(b_k)} \cdot f_j(b_k) \\
f_i(b_k) - f_i(b_k)^2 &\leq f_j(b_k) - f_j(b_k)^2 \\
f_i(b_k) - f_j(b_k) &\leq f_i(b_k)^2 - f_j(b_k)^2 \\
1 &\leq f_i(b_k) + f_j(b_k) \\
f_i(b_k) + f_{k-i}(b_k) &\leq f_i(b_k) + f_j(b_k)
\end{aligned}$$

The equality $1 = f_i(b_k) + f_{k-i}(b_k)$ holds by Lemma 3. The last inequality above follows from Lemma 1 and the fact that $k - i \geq j$.

■

Lemma 5 *Let i, j and k be non-negative integers such that $i+j \leq k-1$ and $k \geq 2$. Then:*

$$f_i(b_k) \cdot f_j(b_k) \leq f_{i+j}(b_k)$$

Proof: We can assume that $i \leq j$. If $i = 0$, the inequality follows from the fact that $f_0(b_k) = 1$. Otherwise, Lemma 4 implies that:

$$f_i(b_k) \cdot f_j(b_k) \leq f_{i-1}(b_k) f_{j+1}(b_k) \leq \ldots \leq f_0(b_k) f_{i+j}(b_k) = f_{i+j}(b_k)$$

■

3 The Lower Bound on π_k

We first introduce some notation used in this section and Section 8. The true value is denoted by 1 and the false value by 0. Following this notation, x^1 denotes the literal x and x^0 the literal $\neg x$. Let ψ now be a formula and x a variable contained in it. Then, $\mathrm{cl}^{\varepsilon}_{\psi}(x)$ is the cardinality of the smallest set C of clauses of ψ such that there is no truth assignment which satisfies all the clauses of C and which assigns $1-\varepsilon$ to the variable x. If there is not such set C, then $\mathrm{cl}^{\varepsilon}_{\psi}(x) = \infty$.

We state and prove two simple lemmas on the values of $\mathrm{cl}^{\varepsilon}_{\psi}(x)$:

Lemma 6 *Let ψ be a k-satisfiable formula. Then, the following holds for each variable x contained in ψ:*

$$\mathrm{cl}^{0}_{\psi}(x) + \mathrm{cl}^{1}_{\psi}(x) \geq k+1$$

Moreover, if ψ cannot be satisfied, then each $\mathrm{cl}^{\varepsilon}_{\psi}(x)$ is finite.

Proof: If ψ cannot be satisfied, then a set C of all the clauses of ψ satisfies the properties required by the definition of $\mathrm{cl}^{\varepsilon}_{\psi}(x)$. Then, each of $\mathrm{cl}^{\varepsilon}_{\psi}(x)$ is finite.

Assume that there is a variable x such that $\mathrm{cl}^{0}_{\psi}(x) + \mathrm{cl}^{1}_{\psi}(x) \leq k$. Let C^{ε} be a set of the size $\mathrm{cl}^{\varepsilon}_{\psi}(x)$ of the clauses which can be satisfied only when the value of the variable x is ε. Then, the clauses of $C^0 \cup C^1$ cannot be simultaneously satisfied, but $|C^0 \cup C^1| \leq k$.

■

Lemma 7 *Let ψ be a k-satisfiable formula and let $(x_1^{\varepsilon_1} \vee \ldots \vee x_n^{\varepsilon_n})$ be a clause of ψ. Then, the following holds:*

$$\mathrm{cl}_\psi^{1-\varepsilon_1}(x_1) + \ldots + \mathrm{cl}_\psi^{1-\varepsilon_n}(x_n) \geq k$$

Proof: Assume that $\mathrm{cl}_\psi^{1-\varepsilon_1}(x_1) + \ldots + \mathrm{cl}_\psi^{1-\varepsilon_n}(x_n) < k$, in particular, all the numbers $\mathrm{cl}_\psi^{1-\varepsilon_i}(x_i)$ are finite. Let C_i be a set of the size $\mathrm{cl}_\psi^{1-\varepsilon_i}(x_i)$ of clauses which can be satisfied only when the value of x_i is $1-\varepsilon_i$. Let C be now the set containing the clause $(x_1^{\varepsilon_1} \vee \ldots \vee x_n^{\varepsilon_n})$ and all the clauses of the sets C_i. If there is a truth assignment which satisfies all the clauses of C, then the value of x_i must be $1-\varepsilon_i$ (because of the clauses of C_i). But, then the clause $(x_1^{\varepsilon_1} \vee \ldots \vee x_n^{\varepsilon_n})$ is not satisfied. Hence, there is no truth assignment which satisfies all the clauses of the set C, but $|C| \leq k$.

$\blacksquare$

Theorem 1 *Let ψ be a k-satisfiable formula. Then, $b_{k+1} \leq \pi(\psi)$. Hence, $b_{k+1} \leq \pi_k$.*

Proof: Let $x_1, \ldots, x_n$ be variables of the formula ψ. If the formula ψ can be satisfied, then there are probabilities $p_1, \ldots, p_n$ such that each clause is satisfied with the probability equal to one (consider the probabilities p_i derived from a satisfying truth assignment). Assume in the rest that the formula ψ is unsatisfiable. In particular, all the numbers $\mathrm{cl}_\psi^\varepsilon(x_i)$ are finite by Lemma 6.

We now define the probabilities p_i:

$$p_i = \begin{cases} f_{\mathrm{cl}_\psi^1(x_i)}(b_{k+1}) & \text{if } \mathrm{cl}_\psi^1(x_i) \leq k/2, \\ f_{k-\mathrm{cl}_\psi^0(x_i)}(b_{k+1}) & \text{if } \mathrm{cl}_\psi^0(x_i) \leq k/2, \\ 1/2 & \text{otherwise.} \end{cases}$$

The probabilities p_i are well-defined because the inequalities $\mathrm{cl}_\psi^0(x_i) \leq k/2$ and $\mathrm{cl}_\psi^1(x_i) \leq k/2$ cannot hold both by Lemma 6. A variable x_i is called *neutral* if $p_i = 1/2$, i.e., both $\mathrm{cl}_\psi^0(x_i)$ and $\mathrm{cl}_\psi^1(x_i)$ are at least $k/2$.

Choose each x_i randomly and independently to be true with the probability p_i. By Lemmas 1 and 3, if $\mathrm{cl}_\psi^\varepsilon(x_i) \leq k/2$, then x_i is chosen to be ε with the probability greater than $1/2$. We show that each clause of ψ is satisfied with the probability at least b_{k+1}. Let us fix a clause Γ of the formula ψ. If Γ contains at least two neutral variables, then it is satisfied with the probability at least $3/4 \geq b_{k+1}$. Hence, we can assume that Γ contains at most one neutral variable.

We first consider the case that the clause Γ contains a neutral variable. If Γ also contains a literal x_i^ε for x_i with $\mathrm{cl}_\psi^\varepsilon(x_i) \le k/2$, then the clause Γ is satisfied with the probability at least $3/4 > b_{k+1}$. Hence, it is enough to consider only the case that the clause Γ is of the form:

$$\Gamma = (x_{i_0}^{\varepsilon_0} \vee x_{i_1}^{\varepsilon_1} \vee \ldots x_{i_l}^{\varepsilon_l})$$

where $p_{i_0} = 1/2$ and $\mathrm{cl}_\psi^{1-\varepsilon_j}(x_{i_j}) < k/2$ for $j = 1, \ldots, l$. By Lemma 3 and the definition of p_i, $x_{i_j}^{\varepsilon_j}$ is chosen to be $1 - \varepsilon_j$ with the probability equal to:

$$f_{\mathrm{cl}_\psi^{1-\varepsilon_j}(x_j)}(b_{k+1})$$

Since $\mathrm{cl}_\psi^{\varepsilon_0}(x_{i_0}) \ge (k+1)/2$, we have that $\mathrm{cl}_\psi^{1-\varepsilon_1}(x_{i_1}) + \ldots + \mathrm{cl}_\psi^{1-\varepsilon_l}(x_{i_l}) \ge (k+1)/2 - 1$ by the definition of $\mathrm{cl}_\psi^{\varepsilon_0}(x_{i_0})$. Lemmas 1 and 5 now bound the probability that each variable $x_{i_j}^{\varepsilon_j}$ is chosen to be $1 - \varepsilon_j$ for all $j = 1, \ldots, l$ as follows:

$$\prod_{j=1}^{l} f_{\mathrm{cl}_\psi^{1-\varepsilon_j}(x_j)}(b_{k+1}) \le f_{\lceil (k+1)/2 \rceil - 1}(b_{k+1})$$

Observe that $f_{\lceil (k+1)/2 \rceil}(b_{k+1}) \le 1/2$ by Lemmas 1 and 3. Thus, we can majorize using Lemmas 2 and 3 the upper bound $1/2 \cdot f_{\lceil (k+1)/2 \rceil - 1}(b_{k+1})$ on the probability that the clause Γ is not satisfied as follows:

$$1/2 \cdot f_{\lceil (k+1)/2 \rceil - 1}(b_{k+1}) \le (1 - f_{\lceil (k+1)/2 \rceil}(b_{k+1})) \cdot f_{\lceil (k+1)/2 \rceil - 1}(b_{k+1}) =$$

$$f_{k+1-\lceil (k+1)/2 \rceil}(b_{k+1}) \cdot f_{\lceil (k+1)/2 \rceil - 1}(b_{k+1}) = 1 - b_{k+1}$$

The final case is that the clause Γ contains no neutral variables. Let us assume that the clause Γ is of the following form:

$$\Gamma = (x_{i_1}^{\varepsilon_1} \vee \ldots x_{i_l}^{\varepsilon_l})$$

If there are distinct j and j' such that $\mathrm{cl}_\psi^{\varepsilon_j}(x_{i_j}) \le k/2$ and $\mathrm{cl}_\psi^{\varepsilon_{j'}}(x_{i_{j'}}) \le k/2$, then each of x_j and $x_{j'}$ is chosen to be ε_j and $\varepsilon_{j'}$, respectively, with the probability at least $1/2$. Hence, Γ is satisfied with the probability at least $3/4 > b_{k+1}$.

Let us consider now the case that there is exactly one j such that $\mathrm{cl}_\psi^{\varepsilon_j}(x_{i_j}) \le k/2$. We can assume that such j is equal to 1. The variable x_{i_1} is chosen to be ε_1 with the probability $f_{\mathrm{cl}_\psi^{\varepsilon_1}(x_{i_1})}(b_{k+1})$ due to the definition of p_{i_1}. By the assumption that no variable of Γ is neutral, we have $\mathrm{cl}_\psi^{1-\varepsilon_{j'}}(x_{i_{j'}}) \le k/2$ for $j' \ne 1$. The definition of $\mathrm{cl}_\psi^{\varepsilon_1}(x_{i_1})$ now implies that:

$$\mathrm{cl}_\psi^{1-\varepsilon_2}(x_{i_2}) + \ldots + \mathrm{cl}_\psi^{1-\varepsilon_l}(x_{i_l}) \ge \mathrm{cl}_\psi^{\varepsilon_1}(x_{i_1}) - 1$$

The variable $x_{i_j}^{\varepsilon_j}$ for $j = 2, \ldots, l$ is chosen to be $1 - \varepsilon_j$ with the probability:

$$f_{\mathrm{cl}_\psi^{1-\varepsilon_j}(x_j)}(b_{k+1})$$

By Lemmas 1 and 5, we can now bound the probability that each variable $x_{i_j}^{\varepsilon_j}$ is chosen to be $1 - \varepsilon_j$ for all $j = 2, \ldots, l$:

$$\prod_{j=2}^{l} f_{\mathrm{cl}_\psi^{1-\varepsilon_j}(x_j)}(b_{k+1}) \leq f_{\mathrm{cl}_\psi^{\varepsilon_1}(x_{i_1})-1}(b_{k+1})$$

Hence, the clause Γ is satisfied with the probability at least:

$$1 - (1 - f_{\mathrm{cl}_\psi^{\varepsilon_1}(x_{i_1})-1}(b_{k+1})) \cdot f_{\mathrm{cl}_\psi^{\varepsilon_1}(x_{i_1})-1}(b_{k+1}) =$$

$$1 - f_{k+1-\mathrm{cl}_\psi^{\varepsilon_1}(x_{i_1})-1}(b_{k+1}) \cdot f_{\mathrm{cl}_\psi^{\varepsilon_1}(x_{i_1})-1}(b_{k+1}) = b_{k+1}$$

The first equality follows from Lemma 3 and the last one from Lemma 2.

Finally, if there is no j such that $\mathrm{cl}_\psi^{\varepsilon_j}(x_{i_j}) \leq k/2$, we have $\mathrm{cl}_\psi^{1-\varepsilon_j}(x_{i_j}) \leq k/2$ for all j because no variable of Γ is neutral. The variable $x_{i_j}^{\varepsilon_j}$ is chosen to be $1 - \varepsilon_j$ with the probability equal to:

$$f_{\mathrm{cl}_\psi^{1-\varepsilon_j}(x_j)}(b_{k+1})$$

By Lemma 7, we have that $\mathrm{cl}_\psi^{1-\varepsilon_1}(x_{i_1}) + \ldots + \mathrm{cl}_\psi^{1-\varepsilon_l}(x_{i_l}) \geq k$. By Lemmas 1 and 5, we can now bound the probability that each variable $x_{i_j}^{\varepsilon_j}$ is set to value $1 - \varepsilon_j$ for all $j = 1, \ldots, l$:

$$\prod_{j=1}^{l} f_{\mathrm{cl}_\psi^{1-\varepsilon_j}(x_j)}(b_{k+1}) \leq f_k(b_{k+1}) = 1 - b_{k+1}$$

Thus, the clause Γ is satisfied with the probability at least b_{k+1}.

■

4 The Upper Bound on π_k

Theorem 2 *Let ψ be the following formula:*

$$\psi = (x_1) \wedge (\neg x_1 \vee x_2) \wedge (\neg x_2 \vee x_3) \wedge \ldots \wedge (\neg x_{k-1} \vee x_k) \wedge (\neg x_k)$$

The formula ψ is k-satisfiable and $b_{k+1} \geq \pi(\psi)$. Hence, $b_{k+1} \geq \pi_k$ for all $k \geq 1$.

Proof: Let k be a fixed integer. We first show that the formula ψ is k-satisfiable, i.e., that any formula ψ' obtained from ψ by removing a single clause can be satisfied. If the missing clause is (x_1) or $(\neg x_k)$, then set all variables to be false or true, respectively. Otherwise, the missing clause is $(\neg x_i \vee x_{i+1})$ for some i, $1 \leq i \leq k-1$. Then, set all the variables $x_{i'}$ for $i' < i$ to be true and all the others to be false.

Let $p_1, \ldots, p_k$ be fixed probabilities. We prove that if each variable x_i is chosen randomly and independently to be true with the probability p_i, then ψ contains a clause which is satisfied with the probability at most b_{k+1}. Assume that all the clauses are satisfied with the probability at least b_{k+1}. We show that then each clause is satisfied with the probability exactly b_{k+1}.

We prove by induction on i that $p_i \geq f_i(b_{k+1})$ and that if $p_i = f_i(b_{k+1})$, then $p_{i'} = f_{i'}(b_{k+1})$ for all $i' < i$. If $i = 1$, then $p_1 \geq f_1(b_{k+1}) = b_{k+1}$ because the clause (x_1) is satisfied with the probability at least b_{k+1}. Let $i \geq 2$. Since the clause $(\neg x_{i-1} \vee x_i)$ is satisfied with the probability at least b_{k+1}, we have:

$$1 - p_{i-1}(1-p_i) \geq b_{k+1}$$

By the induction hypothesis and the equality of (1), we get:

$$\begin{aligned}
1 - f_{i-1}(b_{k+1})(1-p_i) &\geq b_{k+1} \\
1 - b_{k+1} &\geq f_{i-1}(b_{k+1})(1-p_i) \\
\frac{1-b_{k+1}}{f_{i-1}(b_{k+1})} &\geq 1 - p_i \\
p_i &\geq 1 - \frac{1-b_{k+1}}{f_{i-1}(b_{k+1})} \\
p_i &\geq f_i(b_{k+1})
\end{aligned}$$

In addition, if $p_i = f_i(b_{k+1})$, then all the inequalities are strict, in particular, $p_{i-1} = f_{i-1}(b_{k+1})$. Hence, $p_{i'} = f_{i'}(b_{k+1})$ for all $i' < i$.

So, $p_k \geq f_k(b_{k+1})$. Since the clause $(\neg x_k)$ is satisfied with the probability at least b_{k+1}, we have $1 - f_k(b_{k+1}) \geq 1 - p_k \geq b_{k+1}$. Thus, $p_k = f_k(b_{k+1})$. This implies that $p_i = f_i(b_{k+1})$ for all i and each clause is satisfied with the probability exactly b_{k+1} by Lemma 2.

■

5 Properties of r_4

In this section, we establish several identities which are satisfied by the claimed value of r_4. Let q_0 be the unique real solution of the following equation:

$$q_0^3 = (1 - q_0)^2 \tag{5}$$

Using Cardano's formula, we have:

$$q_0 = \sqrt[3]{\frac{3\sqrt{69}+11}{54}} - \sqrt[3]{\frac{3\sqrt{69}-11}{54}} + \frac{1}{3} \approx 0.5698$$

Let us define $p_0 = \sqrt{q_0} \approx 0.7547$. It is easy to check from (5) that p_0 and q_0 satisfy the following equation:

$$p_0^4 = q_0^2 = p_0(1 - q_0) \tag{6}$$

The first and the last part of equation (6) together with the definition $p_0 = \sqrt{q_0}$ gives:

$$p_0^3 + p_0^2 - 1 = 0 \tag{7}$$

The claimed value of r_4 can be determined from the values of p_0 and q_0:

$$r_4 = \frac{1}{1+p_0^3} = \frac{1}{2-p_0^2} = \frac{1}{2-q_0} = \frac{3}{5 + \sqrt[3]{\frac{3\sqrt{69}-11}{2}} - \sqrt[3]{\frac{3\sqrt{69}+11}{2}}} \approx 0.6992 \tag{8}$$

6 The Lower Bound on r_4

We prove that r_4 is at least the value given by (8) in the next theorem:

Theorem 3 *Let ψ be a 4-satisfiable CNF formula. Then:*

$$\frac{w_0(\psi)}{w(\psi)} \geq \frac{1}{2-p_0^2}$$

In particular:

$$r_4 \geq \frac{1}{2-p_0^2}$$

Proof: Consider a fixed 4-satisfiable formula ψ. It is possible to assume that no clause of ψ contains simultaneously a positive and a negative occurrence of the same variable (such a clause is always satisfied and removing it from ψ can only decrease the ratio $w_0(\psi)/w(\psi)$). If ψ does not contain a clause of size one, then set each variable randomly and independently to be true with the probability 1/2. The expected weight of the satisfied clauses is at least $3w(\psi)/4$ (each clause is satisfied with the probability at least 3/4) and hence there is a truth assignment which satisfies clauses of the total weight at least $3w(\psi)/4 \geq w(\psi)/(2-p_0^2)$.

Assume in the rest that ψ contains clauses of size one. It is also possible to assume that all the occurrences of variables in clauses of size one are positive: If this is not true and ψ contains a clause $(\neg x)$, then change all positive occurrences of x to negative and vice versa. Observe that no variable can appear both positively and negatively in clauses of size one because ψ is 4-satisfiable. Now, let X be the set of variables which occur in the clauses of size one. In addition, assume that the sum of the weights of clauses containing a literal[1] x for $x \in X$is equal to one (this can be assured by multiplying all the weights of all the clauses by the same suitable constant which clearly preserves the ratio $w_0(\psi)/w(\psi)$).

We can also assume that there is no clause $(\neg x \vee \neg y)$ for $x \in X$ and any variable y. If there is such a clause, change all negative occurrences of y to positive and vice versa (note that $y \notin X$ because ψ is 4-satisfiable). If this does not help, then ψ contains a clause $(\neg x \vee y)$ and a clause $(\neg x' \vee \neg y)$ for some $x, x' \in X$. But then, the clauses (x), (x'), $(\neg x \vee y)$ and $(\neg x' \vee \neg y)$ of ψ contradict the 4-satisfiability of the formula ψ.

Let Y be the set of variables y which occur in the clauses of type $(\neg x \vee y)$ and which are not contained in X, i.e., Y is the set of all the variables which are not contained in X and which are contained in a clause of size two together with a literal of the form $\neg x$ for $x \in X$. Finally, let Z be the set of the remaining variables contained in ψ, i.e., the variables of ψ included neither to X nor to Y.

We divide clauses of the formula ψ into nine disjoint sets (see Table 1):

- The set A_1 consists of the clauses containing a literal x for $x \in X$.
- The set A_2 consists of the clauses containing only literals of the form $\neg x$ for $x \in X$.
- The set A_3 consists of the clauses which contain a literal $\neg x$ for $x \in X$ and a literal y for $y \in Y$ and which are not in A_1.

[1] We strictly distinguish between a literal and a variable when speaking about containment in clauses, e.g., a clause $(\neg x \vee y)$ contains a variable x but it does not contain a literal x.

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
x	$\surd$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
$\neg x$	?	$\surd$	$\surd$	$\surd$	$\surd$	$\times$	$\times$	$\times$	$\times$
y	?	$\times$	$\surd$	$\times$	$\times$	$\surd$	$\surd$	$\times$	$\times$
$\neg y$	?	$\times$	?	?	$\surd$	$\surd$	$\times$	$\surd$	$\times$
$z, \neg z$	?	$\times$	?	$\surd$	$\times$	?	?	?	$\surd$

Table 1: Distribution of clauses of a 4-satisfiable formula into nine sets $A_1, \ldots, A_9$. The sign $\surd$ denotes that a literal of the particular type must be present in a clause, the sign $\times$ that it cannot be present and the sign ? that it can be present (but it does not need); x represents a variable from the set X, y from Y and z from Z.

- The set A_4 consists of the clauses which contain a literal $\neg x$ for $x \in X$ and a variable from Z and which are contained neither in A_1 nor in A_3.
- The set A_5 consists of the clauses containing only literals of the form $\neg x$ for $x \in X$ and literals of the form $\neg y$ for $y \in Y$.
- The set A_6 consists of the clauses containing no variable from X but which do contain a literal of the form y and a literal of the form $\neg y'$ for $y, y' \in Y$.
- The set A_7 consists of the clauses containing only literals of the type y for $y \in Y$ and literals with the variables of Z.
- The set A_8 consists of the clauses containing only literals of the type $\neg y$ for $y \in Y$ and literals with the variables of Z.
- The set A_9 consists of the clauses containing only variables from the set Z.

It is easy to check that each clause belongs to exactly one of the nine sets defined above. Note that by the assumption on the weights of clauses containing a literal x for $x \in X$, the sum of the weights of the clauses of A_1 is equal to 1. Note also that the sets $A_2, \ldots, A_9$ contain only clauses of size two and more. Let σ be the sum of the weights of clauses contained in the sets A_2, A_3, A_5 and A_8 and let δ be the sum of the weights of clauses of the set A_7.

Assume first that σ is small, namely $\sigma \leq 1/p_0^3$. Let us consider the following probability distribution: Set a variable of X to be true with the probability p_0, a variable of Y with the probability q_0 and a variable from Z with the probability

$1/2$. We analyze the probability that a clause of ψ is satisfied for each of the nine possible clause types:

- A clause of A_1 is satisfied with a probability at least p_0 because it contains a literal x for $x \in X$.
- A clause of A_2 is satisfied with a probability at least $1 - p_0^4$ because it consists of at least four (because of 4-satisfiability) literals of the type $\neg x$ for $x \in X$.
- A clause of A_3 is satisfied with a probability at least $1 - p_0(1 - q_0) = 1 - p_0^4$ because it contains both a literal $\neg x$ for $x \in X$ and a literal y for $y \in Y$.
- A clause of A_4 must be of size at least three (by the definition of the set Y) and hence it is satisfied with a probability at least $\min\{1 - p_0^2/2, 1 - p_0 q_0/2, 1 - p_0/4\} = 1 - p_0^2/2 \geq 1/(2 - p_0^2)$.
- A clause of A_5 must be of size at least three (again by the definition of the set Y) and hence it is satisfied with a probability at least $\min\{1 - p_0^2 q_0, 1 - p_0 q_0^2\} = 1 - p_0^4$.
- A clause of A_6 is satisfied with a probability at least $1 - q_0(1 - q_0) \geq 3/4 \geq 1/(2 - p_0^2)$.
- A clause of A_7 is satisfied with a probability at least $\min\{1-(1-q_0)/2, 1-(1 - q_0)^2\} = (1 + q_0)/2 \geq 3/4 \geq 1/(2 - p_0^2)$.
- A clause of A_8 is satisfied with a probability at least $\min\{1-q_0^2, 1-q_0/2\} = 1 - p_0^4$.
- A clause of A_9 is satisfied with a probability at least $3/4 \geq 1/(2 - p_0^2)$.

We prove, under the assumption $\sigma \leq 1/p_0^3$ (which was made earlier) that the expected weight of the satisfied clauses of ψ is at least $w(\psi)/(2-p_0^2)$. Since the expected weight of the satisfied clauses of A_4, A_6, A_7 and A_9 is at least $1/(2-p_0^2)$ of the sum of their weights, it is enough to prove that the expected weight of the satisfied clauses of A_1, A_2, A_3, A_5 and A_8 is at least $1/(2 - p_0^2)$ of the sum of their weights. The expected weight of the satisfied clauses of A_1 is at least $1 \cdot p_0$ and of the clauses of A_2, A_3, A_5 and A_8 is at least $\sigma \cdot (1 - p_0^4)$. Hence, it is enough to show the following inequality:

$$\frac{1 \cdot p_0 + \sigma \cdot (1 - p_0^4)}{1 + \sigma} \geq \frac{1}{2 - p_0^2}$$

But the previous inequality can be easily verified using the assumption $\sigma \leq 1/p_0^3$ and the equality (8):

$$\frac{p_0 + \sigma(1-p_0^4)}{1+\sigma} \geq \frac{p_0 + (1-p_0^4)/p_0^3}{1+1/p_0^3} = \frac{p_0^4 + 1 - p_0^4}{p_0^3+1} = \frac{1}{2-p_0^2}$$

Assume now that $\sigma > 1/p_0^3$. We consider two probability distributions and we show that the expected weight of the satisfied clauses is large enough for at least one of them. The first one is the distribution defined above, i.e., variables from X are set to be true with the probability p_0, from Y with the probability q_0 and from Z with the probability $1/2$. Again, the expected weight of the satisfied clauses of A_4, A_6 and A_9 is at least $1/(2-p_0^2)$ of the sum of their weights. The expected weight of the satisfied clauses in the rest of ψ is at least $1 \cdot p_0 + \sigma \cdot (1-p_0^4) + \delta \cdot (1+p_0^2)/2$. If $p_0 + \sigma(1-p_0^4) + \delta(1+p_0^2)/2 \geq (1+\sigma+\delta)/(2-p_0^2)$, then the expected weight of the satisfied clauses of the whole formula ψ is at least $w(\psi)/(2-p_0^2)$. This happens if the following inequalities hold:

$$\begin{aligned}
p_0 + \sigma(1-p_0^4) + \delta(1+p_0^2)/2 &\geq (1+\sigma+\delta)/(2-p_0^2) \\
\delta\left(\frac{1+p_0^2}{2} - \frac{1}{2-p_0^2}\right) &\geq \sigma\left(\frac{1}{2-p_0^2} - 1 + p_0^4\right) - p_0 + \frac{1}{2-p_0^2} \\
\delta\left(\frac{2-p_0^2+2p_0^2-p_0^4}{2} - 1\right) &\geq \sigma\left(-1+p_0^2+2p_0^4-p_0^6\right) - 2p_0 + p_0^3 + 1 \\
\delta &\geq \frac{\sigma\left(-1+p_0^2+2p_0^4-p_0^6\right) - 2p_0 + p_0^3 + 1}{p_0^2(1-p_0^2)/2} \qquad (9)
\end{aligned}$$

We apply (7) to (9):

$$\delta \geq \frac{-\sigma p_0^3(1-2p_0+p_0^3) + (1-2p_0+p_0^3)}{p_0^2(1-p_0^2)/2} = \frac{(1-\sigma p_0^3)(p_0^3-2p_0+1)}{p_0^2(1-p_0^2)/2} =$$

$$\frac{(1-\sigma p_0^3)(1-p_0-p_0^2)}{p_0^2(1+p_0)/2} = \frac{(1-\sigma p_0^3)(p_0^3-p_0)}{p_0^2(1+p_0)/2} = \frac{(\sigma p_0^3-1)(1-p_0)}{p_0/2} \qquad (10)$$

The other probability distribution which we consider is simpler: All the variables of X and Y are set to be false and the variables of Z are set to be true with the probability equal to $1/2$. The expected weight of the satisfied clauses of A_4, A_6 and A_9 is at least $3/4$ of the sum of their weights (all the clauses of A_4 and A_6 are satisfied and a clause of A_9 is satisfied with the probability at least $3/4$). Among the remaining clauses all the clauses of A_2, A_3, A_5 and A_8 are satisfied, hence the expected weight of the satisfied clauses of A_1, A_2, A_3, A_5,

A_7 and A_8 is at least σ. Hence, if $\sigma \geq (1+\sigma+\delta)/(2-p_0^2)$, the expected weight of the satisfied clauses of the whole formula ψ is at least $w(\psi)/(2-p_0^2)$. Thus, this distribution provides a sufficient expected weight of the satisfied clauses if the following inequality hold (the last equality is derived using (7)):

$$\delta \leq \sigma(2-p_0^2)-(1+\sigma)=\sigma(1-p_0^2)-1=\sigma p_0^3-1 \tag{11}$$

We have shown that for the first probability distribution, the expected weight of the satisfied clauses is sufficient, i.e., it is at least $w(\psi)/(2-p_0^2)$, if (10) holds, and for the second distribution, it is sufficient if (11) holds. We now complete the proof by showing that at least one of the inequalities (10) and (11) holds providing $\sigma > 1/p_0^3$:

$$\begin{aligned} \frac{(\sigma p_0^3-1)(1-p_0)}{p_0/2} &\leq \sigma p_0^3-1 \\ 1-p_0 &\leq p_0/2 \\ 1 &\leq 3p_0/2 \end{aligned}$$

Since $p_0 \geq 2/3$, the last inequality clearly holds.

■

We show that unlike in the case of k-satisfiable formulas for $k=1,2,3$ there is no universal probability distribution for 4-satisfiable formulas in general:

Proposition 1 *There is a 4-satisfiable formula ψ and $\epsilon > 0$ such that for each probability distribution over its variables (where each of the variables is set independently of the others), weights of the clauses of ψ can be chosen in such a way that the expected weight of the satisfied clauses of ψ is smaller than $1/(2-p_0^2)-\epsilon$.*

Proof: Consider the following 4-satisfiable CNF formula:

$$\psi = (x) \wedge (x') \wedge (\neg x \vee y) \wedge (\neg x' \vee y') \wedge (\neg y \vee \neg y')$$

Let p, p', q and q', respectively, be the probability that the variable x, x', y and y', respectively, is set to be true. It can be easily verified that $\min\{p, p', 1-p+pq, 1-p'+p'q', 1-qq'\} < 1/(2-p_0^2)-\epsilon$ for a fixed $\epsilon > 0$ regardless the actual choice of p, p', q and q'. Assume, e.g., that $1-p+pq < 1/(2-p_0^2)-\epsilon$. If the weights of all the clauses except for the clause $(\neg x \vee y)$ to be negligible, then the expected weight of the satisfied clauses is at most $(1/(2-p_0^2)-\epsilon)w(\psi)$. If the minimum is attained by another term, we proceed analogously.

■

7 The Upper Bound on r_4

We first present a construction of a certain 4-satisfiable formula. Next, we show that any truth assignment of it behaves in some sense almost like a random assignment and then we prove the upper bound by a careful choice of parameters in the construction of the formula.

The formula $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ is defined as follows: It has n variables $x_i, 1 \leq i \leq n$, and n^k variables $y_{\vec{a}}$ where $\vec{a}$ ranges over all ordered k-tuples of numbers $1, \ldots, n$ for $k = \lfloor n^{1/3} \rfloor$. We say that two ordered k-tuples $\vec{a}$ and $\vec{b}$ *have a common entry* if there is i which is an entry both of $\vec{a}$ and $\vec{b}$; we do not require that i has the same position in $\vec{a}$ and $\vec{b}$, e.g., i may be the first entry of $\vec{a}$ and the last one of $\vec{b}$. The clauses of the formula $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ are the following:

- n clauses (x_i) for $1 \leq i \leq n$ each of weight $1/n$.
- kn^k clauses $(\neg x_i \vee y_{\vec{a}})$ for all pairs i and $\vec{a}$ such that i is contained in $\vec{a}$. If i is contained in $\vec{a}$ several times, then the formula contains as many clauses $(\neg x_i \vee y_{\vec{a}})$ as is the number of occurrences of i in $\vec{a}$. Each of these kn^k clauses has the weight equal to $\alpha/(kn^k)$, i.e., the sum of their weights is equal to α.
- clauses $(\neg y_{\vec{a}} \vee \neg y_{\vec{b}})$ for all ordered pairs of k-tuples $\vec{a}$ and $\vec{b}$ which do not have a common entry. Note that each clause of this type is included to the formula twice because the pairs of $\vec{a}$ and $\vec{b}$ are considered to be ordered. All the clauses of this type have the same weight chosen in such a way that the sum of their weights is equal to β.
- $\binom{n}{4}$ clauses $(\neg x_{i_1} \vee \neg x_{i_2} \vee \neg x_{i_3} \vee \neg x_{i_4})$ for all unordered quadruples i_1, i_2, i_3 and i_4 of different numbers between 1 and n. All these clauses have the same weight equal to $y/\binom{n}{4}$, i.e., the sum of their weights is equal to γ.

We first show that the formula $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ is 4-satisfiable:

Lemma 8 *The* $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ *is* 4*-satisfiable for all* n*,* α*,* β *and* γ*.*

Proof: Let W be a minimal set of clauses of $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ which cannot be simultaneously satisfied. Assume for the sake of contradiction that $|W| \leq 4$. By the minimality of W, any variable either does not appear in the clauses of W at all or it has both a positive and a negative occurrence in the clauses of W. Since $|W| \leq 4$, W cannot contain a clause of size four. Moreover, W must contain at least one clause of the type $(\neg y_{\vec{a}} \vee \neg y_{\vec{b}})$ (otherwise, it contains only the clauses of types (x_i) and $(\neg x_i \vee y_{\vec{a}})$ which can be satisfied by setting all the variables to be true). If W contains a clause $(\neg y_{\vec{a}} \vee \neg y_{\vec{b}})$, it must contain also clauses $(\neg x_i \vee y_{\vec{a}})$ for some i contained in $\vec{a}$ and $(\neg x_j \vee y_{\vec{b}})$ for some j contained in $\vec{b}$. Since the clause $(\neg y_{\vec{a}} \vee \neg y_{\vec{b}})$ is present in $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$, we have $i \neq j$. But then, W must contain also the clauses (x_i) and (x_j) which implies that $|W| > 4$.

■

Next, we prove that any truth assignment behaves on $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ like a random assignment. Observe that $p + \alpha(1 - p + pq) + \beta(1 - q^2) + \gamma(1 - p^4)$ is the expected weight of the satisfied clauses of $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ when the variables x_i are set to be true with the probability p and the variables $y_{\vec{a}}$ with the probability q.

Lemma 9 *The weight of the satisfied clauses of* $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ *is equal to* $p + \alpha(1 - p + pq + O(n^{-1/12})) + \beta(1 - q^2 + O(n^{-1/3})) + \gamma(1 - p^4 + O(n^{-1}))$ *for all* n, α, β *and* γ *where* p *is equal to the fraction of variables* x_i *set to true and* q *is the fraction of variables* $y_{\vec{a}}$ *set to true. The constants in the functions* $O(n^{-1/12})$, $O(n^{-1/3})$ *and* $O(n^{-1})$ *are independent of* n, α, β, γ, p *and* q *and the actual terms estimated by them can be both negative and positive.*

Proof: Let $k = \lfloor n^{1/3} \rfloor$ be the size of the tuples $\vec{a}$ as in the construction of $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$. We deal with each of the four types of the clauses contained in $\mathrm{SAT}_4(n, \alpha, \beta, \gamma)$ separately:

- (x_i) for $1 \leq i \leq n$
  There are exactly pn clauses x_i, $1 \leq i \leq n$ satisfied. Hence the weight of the satisfied clauses of this type is exactly p.

- $(\neg x_i \vee y_{\vec{a}})$ for i contained in $\vec{a}$
  Let $\tau(\vec{a})$ be the number of entries of $\vec{a}$ which correspond to the variables x_i set to be true by the assignment (counting multiplicity if i is contained several times in $\vec{a}$). Chernoff bounds [3] can be used to bound the number of ordered k-tuples $\vec{a}$ for which $\tau(\vec{a})$ differs significantly from the "average value" kp:

$$\begin{aligned} |\{\vec{a}, \tau(\vec{a}) \geq (1+\lambda)pk\}| &\leq e^{-\lambda^2 pk/3} n^k \\ |\{\vec{a}, \tau(\vec{a}) \geq (1-\lambda)pk\}| &\leq e^{-\lambda^2 pk/2} n^k \end{aligned}$$

  Set $\lambda = n^{1/4}/pk \geq n^{-1/12}$. Observe that except for at most $e^{-\Theta(n^{1/6})} n^k$ tuples $\vec{a}$, all tuples $\vec{a}$ satisfy that the number of their entries i which correspond to the variables x_i set to true is within the difference of $n^{1/4}$ from pk. If $y_{\vec{a}}$ is false and the number of such i's is within the difference of $n^{1/4}$ from pk, then there are satisfied $(1-p+O(n^{1/4-1/3}))k = (1-p+O(n^{-1/12}))k$ clauses out of all the k clauses $(\neg x_i \vee y_{\vec{a}})$ containing the literal $y_{\vec{a}}$. Hence the ratio of the satisfied clauses of the type $(\neg x_i \vee y_{\vec{a}})$ is equal to $q+(1-q)(1-p+O(n^{-1/12}))+e^{-\Theta(n^{1/6})} = 1-p+pq+O(n^{-1/12})$.

- $(\neg y_{\vec{a}} \vee \neg y_{\vec{b}})$ for $\vec{a}$ and $\vec{b}$ with no common entry
  Consider all ordered pairs of literals $\neg y_{\vec{a}}$ and $\neg y_{\vec{b}}$ (including those with $\vec{a} = \vec{b}$); exactly $1-q^2$ of them contain a satisfied literal. There are n^{2k} such pairs. However only some of them correspond to real clauses. The number of the clauses of the type $(\neg y_{\vec{a}} \vee \neg y_{\vec{b}})$ is at least $n^k(n-k)^k$. Hence the error made by approximating the ratio of the satisfied clauses by $1-q^2$ is of the order $1-\frac{(n-k)^k}{n^k} = O(k^2/n) = O(n^{-1/3})$.

- $(\neg x_{i_1} \vee \neg x_{i_2} \vee \neg x_{i_3} \vee \neg x_{i_4})$
  Consider all n^4 ordered quadruples of (not necessarily distinct) literals $\neg x_i$; exactly $(1-p^4)n^4$ of them contain a satisfied literal. The number of ordered quadruples corresponding to the clauses of the formula is $4!\binom{n}{4}$ (4! quadruples correspond to each clause). Hence at most $n^4 - n(n-1)(n-2)(n-3) = \Theta(n^3)$ of the quadruples do not correspond to the clauses and the fraction of the satisfied clauses of this type is $1-p^4+O(n^{-1})$.

$\blacksquare$

Theorem 4 *For each $\epsilon > 0$, there exists a 4-satisfiable formula ψ such that $w_0(\psi) \leq w(\psi)/(2-p_0^2)+\epsilon$.*

Proof: Set α, β and γ to be the unique solution of the following system of equations:

$$\alpha p_0^4 + \beta p_0^4 + \gamma p_0^4 - p_0 = 0 \quad (12)$$
$$1 - \alpha + \alpha q_0 - 4\gamma p_0^3 = 0 \quad (13)$$
$$\alpha p_0 - 2\beta q_0 = 0 \quad (14)$$

The values of α, β and γ, respectively, are approximately equal to 1.234, 0.819 and 0.272, respectively.

Consider a formula $\mathrm{SAT}_4(n,\alpha,\beta,\gamma)$ for a sufficiently large n, i.e., such that the error term $\alpha O(n^{-1/12})+\beta O(n^{-1/3})+\gamma O(n^{-1})$ from Lemma 9 is smaller than the considered ϵ. The formula $\mathrm{SAT}_4(n,\alpha,\beta,\gamma)$ is 4-satisfiable by Lemma 8. The value $w_0(\mathrm{SAT}_4(n,\alpha,\beta,\gamma))$ is equal up to an additive error of ϵ to the maximum of the function $p+\alpha(1-p+pq)+\beta(1-q^2)+\gamma(1-p^4)$ for $0 \le p,q \le 1$ by Lemma 9 and the choice of n.

The first partial derivatives according to p and q are the following:

$$\begin{aligned}\frac{\partial}{\partial p}\left(p+\alpha(1-p+pq)+\beta(1-q^2)+\gamma(1-p^4)\right) &= 1-\alpha+\alpha q-4\gamma p^3\\ \frac{\partial}{\partial q}\left(p+\alpha(1-p+pq)+\beta(1-q^2)+\gamma(1-p^4)\right) &= \alpha p-2\beta q\end{aligned}$$

Hence, the function $p+\alpha(1-p+pq)+\beta(1-q^2)+\gamma(1-p^4)$ attains its maximum for one of the following pairs of values (p,q): $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, $(1,\alpha/2\beta)$ and (p_0,q_0). The last pair is by the choice of α, β and γ to satisfy (13) and (14). It is easy to check that the values of the function $p+\alpha(1-p+pq)+\beta(1-q^2)+\gamma(1-p^4)$ for $(p,q)=(0,0)$ and for $(p,q)=(p_0,q_0)$ dominate all the others. By (6) and (12), the value of $p+\alpha(1-p+pq)+\beta(1-q^2)+\gamma(1-p^4)$ is equal to $\alpha+\beta+\gamma$ both for $(p,q)=(0,0)$ and $(p,q)=(p_0,q_0)$.

We now estimate the ratio $w_0(\mathrm{SAT}_4(n,\alpha,\beta,\gamma))/w(\mathrm{SAT}_4(n,\alpha,\beta,\gamma))$ using the fact that the maximum of the function $p+\alpha(1-p+pq)+\beta(1-q^2)+\gamma(1-p^4)$ is equal to $\alpha+\beta+\gamma$ and it is within the error of ϵ from $w_0(\mathrm{SAT}_4(n,\alpha,\beta,\gamma))$. We have also $w(\mathrm{SAT}_4(n,\alpha,\beta,\gamma)) = 1+\alpha+\beta+\gamma$ due to the construction of $\mathrm{SAT}_4(n,\alpha,\beta,\gamma)$. The desired estimate on the ratio now follows using the equalities (8) and (12):

$$\frac{w_0(\mathrm{SAT}_4(n,\alpha,\beta,\gamma))}{w(\mathrm{SAT}_4(n,\alpha,\beta,\gamma))} \le \frac{\alpha+\beta+\gamma}{1+\alpha+\beta+\gamma}+\epsilon = \frac{1}{1+p_0^3}+\epsilon = \frac{1}{2-p_0^2}+\epsilon$$

■

8 The Algorithms

One can check that the numbers $\mathrm{cl}_\psi^\varepsilon(x)$ can be replaced in the proof of Theorem 1 by any numbers $\xi_\psi^\varepsilon(x)$ which have the following three properties:

1. $\xi_\psi^0(x)+\xi_\psi^1(x) \ge k+1$ for each variable x of ψ,
2. $\xi_\psi^{1-\varepsilon_1}(x_1)+\ldots+\xi_\psi^{1-\varepsilon_l}(x_l) \ge k$ for each clause $(x^{\varepsilon_1} \vee \ldots \vee x^{\varepsilon_l})$ of ψ and

3. $\xi_\psi^{\varepsilon_0}(x_0) \le 1 + \xi_\psi^{1-\varepsilon_1}(x_1) + \ldots + \xi_\psi^{1-\varepsilon_l}(x_l)$ for each clause $(x^{\varepsilon_0} \vee \ldots \vee x^{\varepsilon_l})$.

Moreover, all the numbers $\xi_\psi^\varepsilon(x)$ need not to be finite (this assumption only simplified several arguments in the proof). The numbers $\mathrm{cl}_\psi^\varepsilon(x)$ have the first two of the above properties by Lemmas 6 and 7 and the third one followed from their definition. Instead of $\mathrm{cl}_\psi^\varepsilon(x)$, our algorithms use the numbers $\mathrm{im}_\psi^\varepsilon(x)$ which are defined to be the (largest) numbers which have the following properties:

- If ψ contains a clause (x^ε), then $\mathrm{im}_\psi^\varepsilon(x) = 1$.
- If ψ contains a clause $(x^{\varepsilon_0} \vee \ldots \vee x^{\varepsilon_l})$, then $\mathrm{im}_\psi^{\varepsilon_0}(x_0) \le 1 + \mathrm{im}_\psi^{1-\varepsilon_1}(x_1) + \ldots + \mathrm{im}_\psi^{1-\varepsilon_l}(x_l)$.
- If there is no restriction on the value of $\mathrm{im}_\psi^\varepsilon(x)$, set $\mathrm{im}_\psi^\varepsilon(x)$ to ∞.

The third property on the numbers $\xi_\psi^\varepsilon(x)$ is satisfied by the definition of $\mathrm{im}_\psi^\varepsilon(x)$. Observe that if $\mathrm{im}_\psi^\varepsilon(x)$ is finite, then there is a set C of clauses of the size $\mathrm{im}_\psi^\varepsilon(x)$ such that the clauses of C can be satisfied only if the value of x is ε. Thus, $\mathrm{cl}_\psi^\varepsilon(x) \le \mathrm{im}_\psi^\varepsilon(x)$. We can conclude that if ψ is k-satisfiable, then the numbers $\mathrm{im}_\psi^\varepsilon(x)$ have all the three properties needed in the proof of Theorem 1. Note the numbers $\mathrm{im}_\psi^\varepsilon(x)$ may also have the properties for a formula which is not k-satisfiable.

Trevisan [9] considered similar numbers but he required instead of $\xi_\psi^{\varepsilon_0}(x_0) \le 1 + \xi_\psi^{1-\varepsilon_1}(x_1) + \ldots + \xi_\psi^{1-\varepsilon_l}(x_l)$ that $\xi_\psi^{\varepsilon_0}(x_0) \le 1 + \max\{\xi_\psi^{1-\varepsilon_1}(x_1), \ldots, \xi_\psi^{1-\varepsilon_l}(x_l)\}$. Hence Trevisan's numbers correspond to depths of certain "derivations" that x is ε and our numbers $\mathrm{im}_\psi^\varepsilon(x)$ to their sizes. This causes that the ratios of the algorithm obtained by Trevisan are worse, but on the other hand, our analysis is less straightforward.

The linear time algorithm of Figure 1 computes the numbers $\mathrm{im}_\psi^\varepsilon(x)$. Once the numbers $\mathrm{im}_\psi^\varepsilon(x)$ are computed, one can check (in linear time) whether the numbers $\mathrm{im}_\psi^\varepsilon(x)$ have the above mentioned three properties. If this is not the case, then the input formula is not k-satisfiable. Otherwise, the probabilities p_i can be defined as in the proof of Theorem 1. In particular, each clause of ψ with respect to the probabilities p_i is satisfied with the probability at least b_{k+1}:

Theorem 5 *There is a linear time algorithm which for a given formula ψ with variables x_i either outputs that ψ is not k-satisfiable or finds probabilities p_i such that each clauses is satisfied with the probability at least π_k if a variable x_i is chosen to be true with the probability p_i.*

```
Input:     a formula ψ
Output:    the numbers im^ε_ψ(x)

unmark all literals x^ε
set all im^ε_ψ(x) to ∞
for each clause (x^ε) set im^ε_ψ(x) to 1
while there is an unmarked literal x^ε with im^{1−ε}_ψ(x) < ∞ do
  choose an unmarked x_1^{ε_1} with im^{1−ε_1}_ψ(x_1) as small as possible
  mark the literal x_1^{ε_1}
  for all clauses (x_0^{ε_0} ∨ ... ∨ x_l^{ε_l}) with only unmarked literal x_0^{ε_0}
    set im^{ε_0}_ψ(x_0) to min{im^{ε_0}_ψ(x_0), 1 + im^{1−ε_1}_ψ(x_1) + ... + im^{1−ε_l}_ψ(x_l)}
  endfor
endwhile
```

Figure 1: A linear time algorithm computing the numbers $\mathrm{im}^{\varepsilon}_{\psi}(x)$.

Using the algorithm from Theorem 5, we can find a truth assignment which satisfies at least the fraction of b_{k+1} of the clauses using a derandomization method proposed by Yannakakis [12]:

Theorem 6 *There is a deterministic linear time algorithm which for a given formula ψ with m clauses either outputs that ψ is not k-satisfiable or finds a truth assignment such that at least $b_{k+1} \cdot m$ clauses are satisfied.*

Similarly, the same method can be applied to the result presented in Section 6:

Theorem 7 *There is a deterministic linear time algorithm which for a given 4-satisfiable formula ψ with m clauses either outputs that ψ is not k-satisfiable or finds a truth assignment such that at least $r_4 \cdot m$ clauses are satisfied.*

Acknowledgement

The author would to thank Gerhard Woeginger for attracting his attention to the problem and for pointing out the paper [10].

References

[1] M. R. Garey, D. S. Johnson: Computers and intractability, A guide to the theory of NP-completeness, Freeman, San Francisco, Cal. (1979).

[2] J. Håstad: Some optimal inapproximability results, Proc. 28th ACM Symposium on Theory of Computing (1997), 1–10.

[3] T. Hagerup, Ch. Rüb: A guided tour Chernoff bounds, Inform. Process. Letters 33 (1989), 305–308.

[4] M. A. Huang, K. Lieberherr: Implications of forbidden structures for extremal algorithmic problems, Theoretical Computer Science 40 (1985), 195–210.

[5] S. Jukna: Extremal combinatorics with applications in computer science, Springer, Heidelberg (2001).

[6] H. Karloff, U. Zwick: A 7/8-approximation algorithm for MAX 3SAT?, Proc. 38th IEEE Symposium on Foundations of Computer Science (1997), 406–415.

[7] K. Lieberherr, E. Specker: Complexity of partial satisfaction, Journal of the ACM 28(2) (1981), 411–422.

[8] K. Lieberherr, E. Specker: Complexity of partial satisfaction II, Technical Report 293, Dept. of EECS, Princeton University (1982).

[9] L. Trevisan: On local versus global satisfiability, to appear in SIAM Journal on Discrete Mathematics. A preliminary version is available as ECCC report TR97-12.

[10] Z. Usiskin: Max-min probabilities in the voting paradox, Ann. Math. Stat. 35 (1963), 857–862.

[11] I. Wegener: The complexity of Boolean functions, John Wiley and Sons, Stuttgart (1987).

[12] M. Yannakakis: On the approximation of maximum satisfiability, Journal of Algorithms 17 (1994), 475–502. A preliminary version appeared in Proc. 3rd Annual ACM/SIGACT-SIAM Symposium on Discrete Algorithms (1992), 1–9.