# Lower and Upper Bounds for Random Mimimum Satisfiability Problem

Ping Huang[1(✉)] and Kaile Su[2]

[1] Key Laboratory of High Confidence Software Technologies, Ministry of Education,
School of Electronic Engineering and Computer Science,
Institute of Software, Peking University, Beijing 100871, China
`huangping@pku.edu.cn`
[2] Institute for Integrated and Intelligent Systems,
Griffith University, Nathan, Australia
`k.su@griffith.edu.au`

**Abstract.** Given a Boolean formula in conjunctive normal form with $n$ variables and $m = rn$ clauses, if there exists a truth assignment satisfying $(1 - 2^{-k} - q(1 - 2^{-k}))m$ clauses, call the formula $q$-satisfiable. The Minimum Satisfiability Problem (MinSAT) is a special case of $q$-satisfiable, which asks for an assignment to minimize the number of satisfied clauses. When each clause contains $k$ literals, it is called Min$k$SAT. If each clause is independently and randomly selected from all possible clauses over the $n$ variables, it is called random MinSAT. In this paper, we give upper and lower bounds of $r$ (the ratio of clauses to variables) for random $k$-CNF formula with $q$-satisfiable. The upper bound is proved by the first moment argument, while the proof of lower bound is the second moment with weighted scheme. Interestingly, our experimental results about MinSAT demonstrate that the lower and upper bounds are very tight. Moreover, these results give a partial explanation for the excellent performance of MinSatz, the state-of-the-art MinSAT solver, from the perspective of pruning effects. Finally, we give a conjecture about the relationship between the minimum number and the maximum number of satisfied clauses on random SAT instances.

**Keywords:** MinSAT · Upper bounds · Lower bounds

## 1 Introduction

Given a Boolean formula in conjunctive normal form(CNF), the satisfiability problem (SAT), which is a prototype of many NP-complete problems, asks for the existence of a satisfying assignment to the formula. In the past decades, SAT has been one of the most active and prolific research areas. Many problems, such as planning [21] and Pseudo-Boolean Constraints, can be translated into SAT [10]. Recently, the success of SAT research has led to exploring its optimization formalisms, such as the maximum satisfiability problem (MaxSAT) [4–6,8,12,15, 19,23] and the minimum satisfiability problem (MinSAT) [14,16,17]. MaxSAT

asks for a Boolean assignment to maximize the number of satisfied clauses, while MinSAT asks for a Boolean assignment to minimize the number of satisfied clauses.

MaxSAT is considered as one of the fundamental combinatorial optimization problems, with close ties to important problems like max cut or max clique, and with applications in scheduling, routing, etc. For MaxSAT, there is a long tradition of theoretical works, e.g. [6,12,15,19]. Moreover, Coppersmith et al. consider the phase transitions of random Max$k$SAT ($k \geq 2$) problem, where each clause contains two literals and is selected independently and randomly [9]. They demonstrate that, with increasing of $r$, i.e. the ratio between number of clauses and number of variables, the expected number of unsatisfied clauses under an optimal assignment quickly changes from $\Theta(1/n)$ to $\Theta(n)$. Furthermore, they provide the upper and lower bounds of the maximum number of satisfied clauses for MaxSAT. Xu et al. improve the upper bound of Max2SAT by the first moment argument via correcting error items [24]. Achlioptas et al. first studied the $p$-satisfiable problem while there exists a truth assignment satisfying a fraction of $1 - 2^{-k} + p2^{-k}$ of all clauses. They introduced weighting second moment method to prove the upper and lower bound of $r$(clauses/variables) [2]. Zhou et al. improve the lower bound by giving a different weight to the truth assignment if exactly one of $k$ literals in a clause is satisfied [26].

The research of MaxSAT leads to increasing interest in its counterpart, MinSAT, which is introduced by Kohli, Krishnamurti and Mirchandani in [14]. They show that MinSAT is NP-complete, even when the formula is a 2-CNF formula, i.e. each clause of which contains at most two literals, or a Horn formula, i.e. each clause of which contains at most one positive variable. They also analyze the performances of deterministic greedy and probabilistic greedy heuristics for MinSAT. A reduction from MinSAT to the minimum vertex cover (MinVC) problem is given in [20], to improve the approximation ratio of MinSAT to 2. A simple randomized 1.1037-approximation algorithm for Min2SAT, and a 1.2136-approximation algorithm for Min3SAT, are given by Avidor and Zwick in [1]. The first exact algorithm for MinSAT is by encoding MinSAT to MaxSAT and solving it with a MaxSAT solver [16]. A branch and bound algorithm for solving MinSAT is proposed by Li et al. in [17]. Their experiments show that solving problems like MaxClique and combinatorial auction problems, is faster by encoding them to MinSAT than reducing them to MaxSAT.

Compared with MaxSAT, there is a lack of knowledge about the bounds of random MinSAT. A $k$-CNF of $q$-satisfiable asks for a truth assignment satisfying a fraction of $1 - 2^{-k} - q(1 - 2^{-k})$ of all clauses. In this paper, we give upper and lower bounds of $r$ for $k$-CNF to be $q$-satisfiable. The upper bound is obtained by the first moment method, while the proof of lower bound is weighted second moment used by [2].

We also present experimental results to demonstrate the tightness of these lower and upper bounds. For MinSAT, the experimental results explain the excellent performance of the state-of-the-art solver MinSatz. Moreover, we investigate the relationship between MaxSAT and MinSAT, and propose a conjecture about

the expected sum of the maximum number and the minimum number of satisfied clauses for a random SAT instance.

This paper is organized as follows. In the next section, we review some basic definitions about SAT and MinSAT. Then the lower and upper bounds of $k$-CNF to be $q$-satisfiable are proved respectively. After that, experimental results are presented, as well as a discussion on the relationship between MaxSAT and MinSAT suggested by the experimental results. Finally, we conclude our work and point out future research directions.

## 2   Preliminaries

A Boolean formula in conjunctive normal form $F$ is a set of clauses $\{C_1, C_2, ..., C_m\}$, where $m$ is the number of clauses in $F$. A clause is a disjunction of literals, $x_{i_1} \vee x_{i_2} \vee \cdots \vee x_{i_k}$, where $k$ is the length of the clause. A literal is either a Boolean variable $x$ or its negation $\bar{x}$. The SAT problem is to determine the existence of an assignment satisfying all the clauses. If there is no assignment to satisfy all the clauses, a natural but more practical question is, how far or how close can one get to satisfiability? This is the optimization version of SAT, which is to find an assignment satisfying the most or the least number of clauses.

**Definition 1.** *(MinSAT) Given a SAT instance $F$, MinSAT asks for an assignment to all the variables such that the minimum number of clauses are satisfied. This number is called the value of the MinSAT instance.*

**Definition 2.** *(q-satisfiable) Given a k-CNF formula $F$, if there exists a truth assignment satisfying a fraction of $1 - 2^{-k} - q(1 - 2^{-k})$ $(0 < q < 1)$ of all clauses, it is q-satisfiable.*

For simplicity, we use $x$ in lieu of $\lfloor x \rfloor$, the largest integer no more than $x$. The following standard asymptotic notations will be used in this paper.

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0 \ \Rightarrow \ f(x) = o(g(x))$$

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1 \ \Rightarrow \ f(x) \simeq g(x)$$

$$\lim_{x \to \infty} \sup \frac{f(x)}{g(x)} \leq M(M > 0) \ \Rightarrow \ f(x) = O(g(x))$$

Especially, while $M = 1$, $f(x) \lesssim g(x)$ means $f(x)$ is less than or equal to $g(x)$ asymptotically.

## 3   The Upper Bound

Given a $k$-CNF instance $F$ to be $q$-satisfiable. Let $P_r$ be the probabilistic distribution and let $N$ denote the solutions number of the instance.

**Theorem 1.** *Let* $r > \frac{2\ln 2}{q^2(2^k-1)}$, *we have*

$$\lim_{n\to\infty} Pr[F \text{ is } q-satisfiable] = 0$$

*Proof.* The expected value of $N$, denoted as $E(N)$, is given by

$$E(N) = 2^n \sum_{i=0}^{\rho rn} \binom{rn}{i} \left(\frac{1}{2^k}\right)^{rn-i} \left(1 - \frac{1}{2^k}\right)^i$$

where $\rho = 1 - 2^{-k} - q(1 - 2^{-k})$.

The last term is maximized for $q \in (0,1)$, so $E(N)$ is upper bounded by

$$E(N) \le 2^n(\rho rn + 1) \binom{rn}{\rho rn} \left(\frac{1}{2^k}\right)^{rn-\rho rn} \left(1 - \frac{1}{2^k}\right)^{\rho rn}$$

According to Stirling's formula $n! \simeq (n/e)^n O(n)$, and

$$\binom{rn}{\rho rn} \simeq \left(\rho^{-\rho}(1-\rho)^{\rho-1}\right)^{rn},$$

We have

$$E(N) \le 2^n(\rho rn + 1) \left(\rho^{-\rho}(1-\rho)^{\rho-1} \left(\frac{1}{2^k}\right)^{1-\rho} \left(1 - \frac{1}{2^k}\right)^{\rho}\right)^{rn}$$

It is easy to prove that $E(N) < 0$ while $r > \frac{2\ln 2}{q^2(2^k-1)}$. By the Markov inequality $Pr(SAT) \le E(N)$, the upper bound is obtained.

## 4   The Lower Bound

Generally, the standard second moment method can be used to prove the lower bound for random problems, such as SAT, CSP. Unfortunately, it fails by using to optimization problem like MaxSAT. To cover this problem, Achlioptas provide a weighting scheme of the second moment to improve the lower bound of kSAT [3] and MaxSAT [2]. We following this line to prove the lower bound of $q$-satisfiable problem.

For any truth assignment $\sigma \in \{0,1\}^n$, let

$$H = H(\sigma, F) = Unsat(l, \sigma) - Sat(l, \sigma)$$
$$S = S(\sigma, F) = Sat(c, \sigma)$$

where $Unsat(l, \sigma)(Sat(l, \sigma))$ is the number of unsatisfied(satisfied) literal, $Sat(c, \sigma)$ is the number of satisfied clauses.

Let $N$ defined as the number of solutions to $q$-satisfiable problem. Define

$$N = \Sigma_\sigma \gamma^{H(\sigma,F)} \eta^{S(\sigma,F)-s_0 rn}$$

where $s_0 = 1 - q - (1-q)2^{-k}$, $0 < \gamma, \eta < 1$.

Using the following two functions,

$$
\begin{aligned}
&f(\alpha, \gamma, \eta) \\
&= \eta^{-2s_0} E[\gamma^{H(\sigma,F)+S(\tau,F)}\eta^{H(\sigma,F)+S(\tau,F)}] \\
&= \eta^{2s_0} \left[ \left( \alpha(\tfrac{\gamma^2+\gamma^{-2}}{2}) + 1 - \alpha \right)^k \right. \\
&\quad - 2(1-\eta) \left( \left( \alpha(\tfrac{\gamma^2+\gamma^{-2}}{2}) + 1 - \alpha \right)^k - \left( \tfrac{\alpha\gamma^{-2}+1-\alpha}{2} \right)^k \right) \\
&\quad \left. + (1-\eta)^2 \left( \left( \left( \alpha(\tfrac{\gamma^2+\gamma^{-2}}{2}) + 1 - \alpha \right)^k - 2\left( \tfrac{\alpha\gamma^{-2}+1-\alpha}{2} \right)^k \right) + 2^{-k} \left( \alpha\gamma^{-2} \right)^k \right) \right]
\end{aligned}
$$

Let

$$
g_r(\alpha, \gamma, \eta) = \frac{f(\alpha, \gamma, \eta)^r}{\alpha^\alpha (1-\alpha)^{1-\alpha}}
$$

We have

$$
E^2[N] = \frac{2^n}{\eta^{s_0 rn}} \left[ (\tfrac{\gamma+\gamma^{-1}}{2})^k - (1-\eta)[\tfrac{\gamma+\gamma^{-1}}{2})^k - (2\gamma)^{-k}] \right]^{rn} = \left( 2g_r \left( \tfrac{1}{2}, \gamma, \eta \right) \right)^n
$$

**Theorem 2.** *Given a k-CNF formula F, if $r \leq \frac{\ln 2}{(q+(1-q)\ln(1-q))2^k}(1-O(k2^{-k}))$,*

$$
\lim_{n \to \infty} [F \ is \ q - satisfiable] = 1.
$$

The proof is similar to the lower bound for MaxSAT in [2], so we ignore here.

## 5 Experimental Results

We conduct experiments to $k$-CNF formula to be $q$-satisfiable, the results are presented in Fig. 1. The upper bound (*upper*) is proved by the first moment
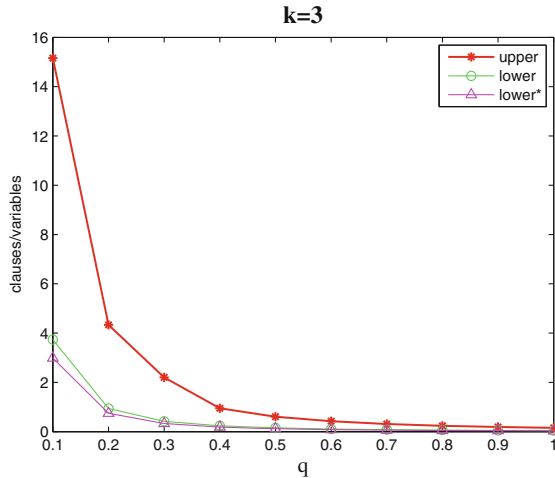


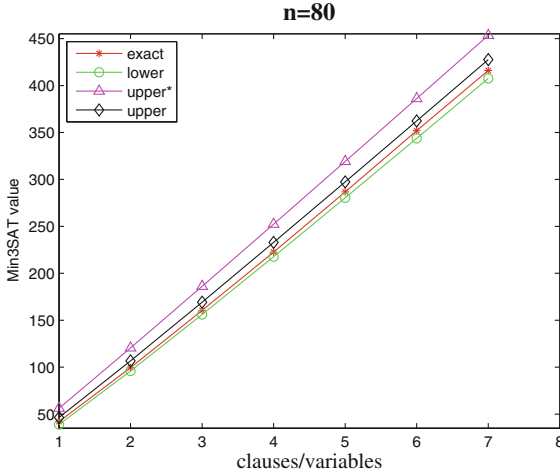**Fig. 1.** The upper and lower bound for $q$-satisfiable

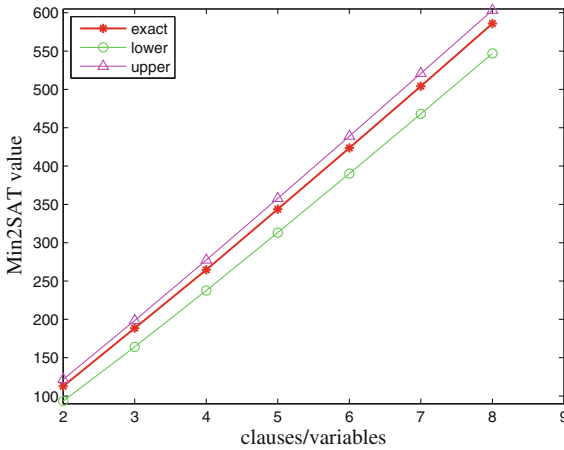**Fig. 2.** The bounds of $n = 80$ for Min3SAT



**Fig. 3.** The bounds of $n = 120$ for Min2SAT

method, while the proof of the lower bound (*lower*) is weighted scheme of the second moment method. We also give a poor lower bound (*lower\**) by algorithm analysis, which was used to MaxSAT in [9].

From Fig. 1, we can see that the space between upper bond and lower bound become smaller as $q$ increase. That is to say, the bounds is tighter while the number of satisfying clauses is less. Based on this, the lower and upper bounds of Min$k$SAT is presented in Fig. 2, which indicate that the bounds for MinSAT ($k = 3$) provided by this paper is very tight. However, the second moment is poor to upper bound for Min2SAT, so algorithm analysis is considered in Fig. 3.

**Table 1.** Comparing with *MinSATz* with branching number

|      | r = 6 | | r = 7 | |
| #n | *MinSatz* + LB | *MinSatz* | *Minsatz* + LB | *MinSatz* |
|-----|------|------|------|------|
| 120 | 6593 | 7331 | 14601 | 14764 |
| 130 | 24881 | 25605 | 319657 | 333949 |
| 140 | 109061 | 112213 | 946257 | 947779 |
| 150 | 49597 | 49744 | 552269 | 552442 |

In the second experiment, we apply our work to the state-of-the-art solver *MinSatz*. In *MinSatz*, UB is the largest number of falsified clauses while extending the current partial assignment to a complete one, and LB is the number of clauses falsified in the best assignment found so far. if LB < UB, MinSatz select a variable and instantiate it, otherwise, the solver backtracks. Besides, *MinSatz* introduces both clique partitioning algorithms and MaxSAT technology to improve the UB so as to prune the search tree quickly. However, we focus on the other side, and give LB an initial number so as to reduce the branching number. We conduct an experiment to test the performances before and after adding our work to *MinSatz* in Table 1. The branching number can be reduced while giving UB a initial value computed by this paper, but the improvements is not so obvious. Further analysis indicate that the genuine *MinSatz* is an excellent solver for MinSAT. However, if the MinSAT solver is not so excellent as *MinSatz*, such as a trivial MinSAT solver with no inference rules (*minsat*), the improvements will be more obvious. The results are presented in Table 2, which indicates that a MinSAT solver with our work outperforms the one without.

Li et al. have found from experiments that the relationship between MaxSAT and MinSAT is counter-intuitive [17,18]: for the same instances, the bigger the MaxSAT value is, the smaller the MinSAT value is, the opposite is also true. They focus these instances at the threshold ($c = 4.25$ for 3SAT). We follow this line, and conjecture that, the sum of the MaxSAT and MinSAT value is a constant value. In our experiments, the number of variables in random instances ranges from 40 to 140, and the density $r$ considered are 6, 7 and 8. We compare the sum of exact MinSAT value and MaxSAT value ('sum' in Table 3) with our conjectured value ('conjecture' in Table 3). From experiments, the accuracy of our conjecture is found to be over 99 % ('accuracy' in Table 3). This indicates

**Table 2.** Application to MinSAT solver

|      | r = 0.8 | | | r = 1 | | | r = 2 | | |
| #n | *minsat* | *minsat* + LB | incr | *minsat* | *minsat* + LB | incr | *minsat* | *minsat* + LB | incr |
|-----|------|------|------|------|------|------|------|------|------|
| 20 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| 30 | 0 | 0 | 0 | 1 | 1 | 0 | 100 | 85 | 15 % |
| 40 | 500 | 420 | 16 % | 600 | 500 | 16.67 % | 1783 | 1458 | 18.22 % |
| 50 | 1284 | 1076 | 16.2 % | 2117 | 1916 | 9.49 % | 4839 | 4279 | 11.57 % |

**Table 3.** The relationship of MinSAT and MaxSAT

| #n | r = 6 | | | r = 7 | | | r = 8 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sum | Conjecture | Accuracy | Sum | Conjecture | Accuracy | sum | Conjecture | Accuracy |
| 40 | 357 | 360 | 99.17 % | 418 | 420 | 99.52 % | 478 | 480 | 99.58 % |
| 60 | 535 | 540 | 99.07 % | 627 | 630 | 99.52 % | 716 | 720 | 99.44 % |
| 80 | 715 | 720 | 99.31 % | 834 | 840 | 99.29 % | 954 | 960 | 99.37 % |
| 100 | 892 | 900 | 99.11 % | 1041 | 1050 | 99.14 % | 1191 | 1200 | 99.25 % |
| 120 | 1071 | 1080 | 99.17 % | 1251 | 1260 | 99.29 % | 1431 | 1440 | 99.37 % |
| 140 | 1248 | 1260 | 99.05 % | 1459 | 1470 | 99.25 % | 1669 | 1680 | 99.35 % |

that our conjecture is close to the exact value. For $k$SAT instances, the sum of MinSAT value and MaxSAT value is approximately $(2-2^{1-k})cn$. If the relationship between MinSAT and MaxSAT values is clear, this conjecture can be used to get a better upper bound for MaxSAT. In other words, give the exact value of MinSAT, the value of MaxSAT for the same instance can be approximated. Besides, this value is significantly better than the upper bound obtained by [9], see Table 4.

**Table 4.** Conjecture about the upper bound of MaxSAT. 'MinSAT value' ('MaxSAT value') is the exact value of MinSAT (MaxSAT), 'ub04' is the upper bound of MaxSAT obtained by [9], $ub^\star$ is our guess value.

| #n | MinSAT value | MaxSAT value | $ub^\star$ | ub04 |
|---|---|---|---|---|
| 40 | 142 | **215** | **218** | **229** |
| 60 | 213 | **322** | **327** | **344** |
| 80 | 283 | **432** | **437** | **459** |
| 100 | 353 | **539** | **547** | **574** |
| 120 | 423 | **648** | **657** | **689** |
| 140 | 493 | **755** | **767** | **804** |

## 6  Conclusions and Future Work

We have presented upper and lower bounds of the minimization versions of the SAT problem. For the upper bound, the first moment argument is used, while the lower bound is derived by weighting second moment. The experimental results confirm the correctness and accuracy of our work. Furthermore, we consider the relationship between MinSAT and MaxSAT, and an interesting conjecture is presented. As for future work, the bounds of other optimization problems such as Min-CUT of random graphs [7] and Min-CSP [13] of random CSPs [25] may be considered.

# References

1. Avidor, A., Zwick, U.: Approximating MIN 2-SAT and MIN 3-SAT. Theor. Comput. Syst. **38**(3), 329–345 (2005)
2. Achlioptas D., Naor A., Peres, Y.: On the maximum satisfiability of random formulas. In: FOCS 2003, pp. 362–370 (2003)
3. Achlioptas, D., Peres, Y.: The threshold for random k-SAT is $2^k \log 2 - O(k)$. J. Am. Math. Soc. **17**(4), 947–973 (2004)
4. Anstegui, C., Bonet, M.L., Levy, J.: SAT-based MaxSAT algorithms. Artif. Intell. **196**, 77–105 (2013)
5. Anstegui, C., Malitsky, Y., Sellmann, M: MaxSAT by improved instance-specific algorithm configuration. In: AAAI 2014, pp. 2594–2600 (2014)
6. Bollobas, B., Borgs, C., Chayes, J.T., Kim, J.H., Wilson, D.B.: The scaling window of the 2-SAT transition. Random Struct. Algorithms **18**(3), 201–256 (2001)
7. Bollobas, B.: Random Graphs. Cambridge Studies in Advanced Mathematics, vol. 73, 2nd edn. Cambridge University Press, Cambridge (2001)
8. Cai, S.W., Luo, C., Thornton, J., Su, K.L.: Tailoring local search for partial MaxSAT. In: AAAI 2014, pp. 2623–2629 (2014)
9. Coppersmith, D., Gamarnik, D., Hajiaghayi, M.T., Sorkin, G.B.: Random MAX SAT random MAX CUT, and their phase transitions. Random Struct. Algorithms **24**, 502–545 (2004)
10. Eén, N., Sörensson, N.: Translating pseudo-boolean constraints into SAT. J. Satisfiability Boolean Model. Comput. **2**(2006), 21–26 (2006)
11. Gramm, J., Hirsch, E.A., Niedermeier, R., Rossmanith, P.: Worst-case upper bounds for MAX-2-SAT with an application to MAX-CUT. Discrete Appl. Math. **130**, 139–155 (2003)
12. Hirsch, E.A.: A new algorithm for MAX-2-SAT. In: Reichel, H., Tison, S. (eds.) STACS 2000. LNCS, vol. 1770, pp. 65–73. Springer, Heidelberg (2000)
13. Huang, P., Yin, M.H.: An upper (lower) bound for Max(Min) CSP. Sci. China Inf. Sci. **57**(7), 1–9 (2014)
14. Kohli, R., Krishnamurti, R., Mirchandani, P.: The minimum satisfiability problem. SIAM J. Discrete Math. **7**, 275–283 (1994)
15. Li, C.M., Many, F., Nouredine, N.O., Planes, J.: Resolution-based lower bounds in MaxSAT. Constraints **15**(4), 456–484 (2010)
16. Li, C.M., Manyà, F., Quan, Z., Zhu, Z.: Exact MinSAT solving. In: Strichman, O., Szeider, S. (eds.) SAT 2010. LNCS, vol. 6175, pp. 363–368. Springer, Heidelberg (2010)
17. Li, C.M., Zhu, Z., Many, F., Simon, L.: Minimum satisfiability and its applications. In: IJCAI 2011, pp. 605–610 (2011)
18. Li, C.M., Zhu, Z., Many, F., Simon, L.: Optimizing with minimum satisfiability. Artif. Intell. **190**, 32–44 (2012)
19. Lin, H., Su, K.L., Li, C.M.: Within problem learning for efficient lower bound computation in Max-SAT solving. In: AAAI 2008, pp. 351–356 (2008)
20. Marathe, M.V., Ravi, S.S.: On approximation algorithms for the minimum satisfiability problem. Inf. Process. Lett. **58**, 23–29 (1996)
21. Robinson, N., Gretton, C., Pham, D.N., Sattar, A.: A compact and efficient SAT encoding for planning. In: ICAPS 2008, pp. 296–303 (2008)
22. Spencer, J.H.: Ten Lectures on the Probabilistic Method. SIAM, Philadelphia (1994)

23. Whitley, D., Howe, A.E., Hains, D.: Greedy or Not? Best Improving versus first improving stochastic local search for MAXSAT. In: AAAI 2013 (2013)
24. Xu, X.L., Gao, Z.S., Xu, K.: A tighter upper bound for random MAX 2-SAT. Inf. Process. Lett. **111**(3), 115–119 (2011)
25. Xu, K., Li, W.: Exact phase transitions in random constraint satisfaction problems. J. Artif. Intell. Res. (JAIR) **12**, 93–103 (2000)
26. Zhou, G.Y., G Z.S., Liu J.: On the lower bounds of random Max 3 and 4-SAT. Manuscript