ON RANDOM 3-SAT

SVANTE JANSON

ABSTRACT. We prove that a random Boolean 3-SAT formula on n variables with $4.596n$ clauses is almost certainly unsatisfiable, in the sense that the probability of it being satisfiable tends to 0 as $n \to \infty$.

1. Introduction and results

The random 3-SAT problem is the following: Let $x_1, \dots, x_n$ be n Boolean variables and consider the $8\binom{n}{3}$ clauses consisting of the disjunction of three literals x_i or $\neg x_i$ using three different variables. Construct a random formula $\varphi_{n,m}$ as the conjunction of m of these clauses, selected at random (with replacement, for definiteness; this does not matter asymptotically). Now let $n, m \to \infty$. Experiments suggest strongly that there is a threshold $\gamma \approx 4.2$ such that if $m/n \le \gamma - \varepsilon$ for any $\varepsilon > 0$, then $\mathbb{P}(\varphi_{n,m}$ is satisfiable$) \to 1$, and if $m/n \ge \gamma + \varepsilon$, then $\mathbb{P}(\varphi_{n,m}$ is satisfiable$) \to 0$.

This is not yet proven, but Friedgut [2] has shown that there exists a sequence γ_n such that if $m/n \le \gamma_n - \varepsilon$, then $\mathbb{P}(\varphi_{n,m}$ is satisfiable$) \to 1$, and if $m/n \ge \gamma_n + \varepsilon$, then $\mathbb{P}(\varphi_{n,m}$ is satisfiable$) \to 0$. It is still not known, however, if these γ_n converge (or, equivalently, can be taken as a constant γ). We therefore define $\gamma_* = \underline{\lim}_{n\to\infty}\gamma_n = \sup\{c : \mathbb{P}(\varphi_{n,\lfloor cn\rfloor}$ is satisfiable$) \to 1\}$ and $\gamma^* = \overline{\lim}_{n\to\infty}\gamma_n = \inf\{c : \mathbb{P}(\varphi_{n,\lfloor cn\rfloor}$ is satisfiable$) \to 0\}$. Hence $\gamma_* \le \gamma^*$ (and presumably equality holds).

Many authors have given numerical bounds for γ_* or γ^* using various techniques. The best bounds that have appeared so far to our knowledge are

$$3.003 < \gamma_* \le \gamma^* < 4.601+,$$

where the lower bound is due to Frieze and Suen [6] and the upper to Kirousis, Kranakis, Krizanc and Stamatiou [5]. The purpose of the present paper is to present a small improvement in the upper bound.

Theorem 1. $\gamma^* < 4.596$.

The proof is based on the proof in [5], improving one estimate only.

The improvement concerns the sum

$$S_n = \sum_{\varepsilon_1,\dots,\varepsilon_n\in\{0,1\}} \exp\Big(a\sum_{i=1}^{n}\varepsilon_i + \frac{b}{n}\sum\sum_{1\le i<j\le n}\varepsilon_i(1-\varepsilon_j)\Big), \tag{1}$$

summing over the 2^n sequences $\varepsilon_1, \dots, \varepsilon_n$ of 0's and 1's. We will prove the following asymptotic formula for S_n. (All logarithms in this paper are natural.)

Date: March 2, 1999; revised March 8, 1999.

Theorem 2. *Let a and b be two fixed real numbers, and define S_n by* (1). *Then, as $n \to \infty$,*

$$\tfrac{1}{n} \log S_n \to \psi(a,b),$$

where $\psi(a,0) = \log(1+e^a)$ and for $b \neq 0$

$$\begin{aligned}\psi(a,b) &= \frac{(h-a)^2}{2b} + \frac{1}{b}\int_h^{h+b} \log(1+e^x)\,dx \\ &= \frac{(h-a)^2}{2b} + \frac{1}{b}\bigl(\operatorname{dilog}(1+e^h) - \operatorname{dilog}(1+e^{h+b})\bigr), \end{aligned} \tag{2}$$

with $h = h(a,b)$ denoting the unique root of

$$h - \log(1+e^h) + \log(1+e^{h+b}) = a. \tag{3}$$

Remark 1. The dilogarithm is defined by $\operatorname{dilog}(x) = -\int_1^x \frac{\log t}{t-1}\,dt$ [1], and the second equality in (2) follows by a simple change of variables.

Remark 2. It is easily shown that ψ is continuous, also at $b = 0$.

Remark 3. The sum S_n equals $\sum_{k=0}^n \binom{n}{k}_q X^k$, where $X = e^a$, $q = e^{b/n}$ and the q-binomial coefficient $\binom{n}{k}_q = \prod_{i=0}^{k-1}(1-q^{n-i})/(1-q^{k-i})$, see [4, 5], but we will not use this representation.

Remark 4. In statistical physics terminology, S_n is the partition function for a system with n sites, each of which has a spin ε_i which may take the values 0 and 1, and with an energy function $H = -a\sum_{i=1}^n \varepsilon_i - \frac{b}{n}\sum\sum_{1\le i<j\le n} \varepsilon_i(1-\varepsilon_j)$. The first term in H corresponds to an external field acting on all spins, and the second to a peculiar interaction which acts only between pairs of sites (of arbitrary distance) where the left site has spin 1 and the right site has spin 0.

This energy function can easily be rewritten into a more conventional form. In fact, $H = \sum_{i=1}^n(-a-b+b\frac{i}{n})\varepsilon_i + \frac{b}{n}\sum\sum_{i<j}\varepsilon_i\varepsilon_j$, or, substituting $\varepsilon_i = (1+s_i)/2$ to have more traditional (and symmetrical) spins with values ± 1, $H = -\frac{a}{2}n - \frac{b}{8}(n-1) + \sum_{i=1}^n(-\frac{a}{2} + b\frac{i-(n+1)/2}{2n})s_i + \frac{b}{4n}\sum\sum_{i<j} s_i s_j$, which shows that the system can be interpreted as a mean-field Ising model with an inhomogeneous (linear) external field.

Remark 5. Theorem 1 has independently been proved by Stamatiou and Vamvakari [7], using another estimate of the sum S_n.

Acknowledgement

I thank Anders Martin-Löf for helpful discussions.

2. Proof of Theorem 2

Theorem 2 is proved by a fairly straightforward optimization of the type common in statistical physics. For later notational convenience we rewrite (1) in the equivalent form (reversing the order of $\varepsilon_1, \dots, \varepsilon_n$)

$$S_n = \sum_{\varepsilon_1,\dots,\varepsilon_n \in \{0,1\}} \exp\Bigl(a\sum_{i=1}^n \varepsilon_i + \frac{b}{n}\sum\sum_{1\le i<j\le n}(1-\varepsilon_i)\varepsilon_j\Bigr), \tag{4}$$

We begin by discretizing in order to obtain a finite-dimensional optimization problem also asymptotically. Fix a large integer N and partition $\{1,\dots,n\}$ into N intervals $I_1,\dots,I_N$ of lengths $\lfloor n/N\rfloor$ or $\lceil n/N\rceil$; let $n_k=|I_k|=n/N+O(1)$. We group the terms in (1) according to $m_k=\sum_{i\in I_k}\varepsilon_i$, $k=1,\dots,N$, and obtain, observing that the sum of all $(1-\varepsilon_i)\varepsilon_j$ with i and j in the same interval I_k is less than $N(n/N)^2$,

$$S_n=\sum_{m_1,\dots,m_k}\prod_{k=1}^{N}\binom{n_k}{m_k}\exp\Big(a\sum_{k=1}^{N}m_k+\frac{b}{n}\sum\sum_{1\le k<l\le N}(n_k-m_k)n_l+O(n/N)\Big).$$

(The implicit constants in the O terms here and below may depend on a and b but not on n or N.) By Stirling's formula,

$$\begin{aligned}\log\binom{M}{m}&=M\log M-m\log m-(M-m)\log(M-m)+O(\log M)\\&=M\Big(-\frac{m}{M}\log\frac{m}{M}-\big(1-\frac{m}{M}\big)\log\big(1-\frac{m}{M}\big)\Big)+O(\log M)\end{aligned}$$

and thus, with $y_k=m_k/n_k$ and summing over the grid $\Lambda_{Nn}=\{(y_1,\dots,y_N)\in[0,1]^N : n_ky_k \text{ is an integer for every } k\}$

$$\begin{aligned}S_n=\sum_{\Lambda_{Nn}}\exp\Big(&\sum_{k=1}^{N}\frac{n}{N}\big(-y_k\log y_k-(1-y_k)\log(1-y_k)\big)+O(N\log n)\\&+a\sum_{k=1}^{N}\frac{n}{N}y_k+\frac{bn}{N^2}\sum\sum_{1\le k<l\le N}(1-y_k)y_l+O(N)+O(n/N)\Big).\end{aligned}$$

Define, for $y=(y_1,\dots,y_N)\in[0,1]^N$,

$$\begin{aligned}F_N(y)=\frac{1}{N}\sum_{k=1}^{N}&\big(-y_k\log y_k-(1-y_k)\log(1-y_k)+ay_k\big)\\&+\frac{b}{N^2}\sum\sum_{1\le k<l\le N}(1-y_k)y_l+\frac{b}{N^2}\sum_{k=1}^{N}(y_k-y_k^2/2)\end{aligned}\tag{5}$$

(the last term, which is $O(1/N)$, will be convenient below). Then,

$$S_n=\sum_{y\in\Lambda_{Nn}}\exp\Big(nF_N(y)+O\Big(\frac{n}{N}+N\log n\Big)\Big).$$

The number of terms in this sum is $\prod_1^N(1+n_k)\le n^N$, and thus

$$S_n=\max_{y\in\Lambda_{Nn}}\exp\Big(nF_N(y)+O\Big(\frac{n}{N}+N\log n\Big)\Big)$$

and

$$\tfrac{1}{n}\log S_n=\max_{y\in\Lambda_{Nn}}F_N(y)+O\Big(\frac{1}{N}+N\frac{\log n}{n}\Big).\tag{6}$$

Since F_N is continuous on $[0,1]^N$, $\max_{y\in\Lambda_{Nn}} F_N(y)$ tends to the global maximum of F_N on $[0,1]^N$ as $n\to\infty$, and (6) implies

$$\liminf_{n\to\infty} \tfrac{1}{n}\log S_n,\ \limsup_{n\to\infty} \tfrac{1}{n}\log S_n = \max_y F_N(y) + O\big(\tfrac{1}{N}\big). \tag{7}$$

It follows easily that both $\lim_{n\to\infty}\frac{1}{n}\log S_n$ and $\lim_{N\to\infty}\max_y F_N(y)$ exist and are equal, but it remains to find this limit.

F_N is continuous on the closed cube $[0,1]^N$ and differentiable in its interior with

$$\begin{aligned} N\frac{\partial F_N}{\partial y_k} &= -\log y_k + \log(1-y_k) + a + \frac{b}{N}\Big(-\sum_{l>k} y_l + \sum_{j<k}(1-y_j) + 1 - y_k\Big) \\ &= -\log y_k + \log(1-y_k) + a + \frac{b}{N}k - \frac{b}{N}\sum_{l=1}^N y_l. \end{aligned} \tag{8}$$

Since this derivative tends to $+\infty$ as $y_k\downarrow 0$, and to $-\infty$ as $y_k\uparrow 1$, the maximum of F_N cannot be attained for $y_k=0$ or $y_k=1$; thus the maximum is attained at an interior point y^* where all $\partial F_N/\partial y_k = 0$.

We define

$$h_N = a - \frac{b}{N}\sum_{k=1}^N y_k^*, \tag{9}$$

and obtain by (8)

$$\log\frac{y_k^*}{1-y_k^*} = h_N + b\frac{k}{N}.$$

Hence,

$$y_k^* = \frac{e^{h_N+bk/N}}{1+e^{h_N+bk/N}}. \tag{10}$$

Here h_N should be such that (9) holds. We observe that $\frac{b}{N}\sum y_k^*$ is a Riemann sum of

$$\int_0^1 b\frac{e^{h_N+bt}}{1+e^{h_N+bt}} = \big[\log(1+e^{h_N+bt})\big]_0^1 = \log(1+e^{h_N+b}) - \log(1+e^{h_N}).$$

Since (9) implies $|h_N|\le|a|+|b|$ and the Riemann sums of this integral converge uniformly for h_N in a compact set,

$$h_N = a - \log(1+e^{h_N+b}) + \log(1+e^{h_N}) + o(1). \tag{11}$$

The function $\varphi(h) = h - \log(1+e^h) + \log(1+e^{h+b}) = \log(1+e^{h+b}) - \log(1+e^{-h})$ is strictly increasing on $(-\infty,+\infty)$, and $\lim_{h\to\pm\infty}\varphi(h)=\pm\infty$. Consequently, there exists a unique $h=h(a,b)$ such that $\varphi(h)=a$, i.e. (3) holds, and (11) implies that $h_N\to h(a,b)$ as $N\to\infty$.

Let $g_N(t) = e^{h_N+bt}/(1+e^{h_N+bt})$ and $g(t) = e^{h+bt}/(1+e^{h+bt})$. Thus, by (10), $y_k^* = g_N(k/N)$. Now let $N\to\infty$; then $g_N\to g$ uniformly on $[0,1]$, and it follows from the definition (5) of F_N that

$$\max F_N = F_N(y^*) \to \int_0^1 \Big(-g(t)\log g(t) - \big(1-g(t)\big)\log\big(1-g(t)\big) + ag(t)\Big)\,dt \\ + b \iint_{0\le s<t\le 1} \big(1-g(s)\big)g(t)\,ds\,dt. \tag{12}$$

Denoting the right hand side by $\psi(a,b)$, this together with (7) shows that $\liminf \frac{1}{n}\log S_n = \limsup \frac{1}{n}\log S_n = \psi(a,b)$, as asserted.

It remains to derive the more explicit expressions given for $\psi(a,b)$. The case $b=0$ is trivial, with $S_n = (1+e^a)^n$, so we assume $b \neq 0$. Integrating first over s, we obtain

$$\begin{aligned}\psi(a,b) &= \int_0^1 \Big(-g(t)(h+bt) + \log(1+e^{h+bt}) + ag(t) \\ &\qquad + \big(bt - \log(1+e^{h+bt}) + \log(1+e^h)\big)g(t)\Big)\,dt \\ &= \tfrac{1}{b}\Big[(a-h)\log(1+e^{h+bt}) - \tfrac{1}{2}\big(\log(1+e^{h+bt}) - \log(1+e^h)\big)^2\Big]_0^1 \\ &\qquad + \int_0^1 \log(1+e^{h+bt})\,dt,\end{aligned}$$

which, using $\log(1+e^{h+b}) - \log(1+e^h) = a-h$ from (3), easily is shown to equal the expressions given in (2).

Remark 6. The proof demonstrates the usual conflict between energy and entropy; the largest term in (1) is obtained for a sequence of 0's followed by 1's or conversely (depending on the sign of b), while most terms have about $n/2$ 0's and 1's rather uniformly distributed; neither of these cases contribute significantly to the sum (unless $a=b=0$). The proof shows that the main contribution comes from terms where the density of 1's among the ε_i with $i/n \approx t$ is about $g(1-t)$; these are thus the typical states in the statistical physics interpretation.

3. Proof of Theorem 1

We sketch the argument of Kirousis, Kranakis, Krizanc and Stamatiou [5], omitting most details.

First, an old observation is that a given truth assignment A to $x_1,\dots,x_n$ satisfies $7\binom{n}{3}$ of the $8\binom{n}{3}$ considered clauses with three literals, and hence the probability that $\varphi_{n,m}$ is satisfied by A equals $(7/8)^m$. Since there are 2^n different truth assignments, the expected number of truth assignments that satisfy $\varphi_{n,m}$ is $2^n(7/8)^m$, and if $m/n \ge \log 2/\log(8/7) + \varepsilon$ (with $\varepsilon > 0$), this tends to 0 and thus $\mathbb{P}(\varphi_{n,m}$ is satisfiable$) \to 0$. Consequently $\gamma^* \le \log 2/\log(8/7) \approx 5.191$.

This bound is not sharp, and the reason is that a satisfiable formula usually has many satisfying truth assignments, and thus it is possible that the expected number tends to infinity even when the probability of having any satisfying truth assignment tends to 0. The method used in [5] to obtain better estimates is to count only a subset of all satisfying truth assignments; they choose the set $\mathcal{A}^{2\sharp}$ defined as follows.

A change of a truth assignment to another is called a *single flip* if the value of exactly one variable x_i is changed and this change is from FALSE to TRUE; a change is called a *double flip* if the values of exactly two variables x_i and x_j (with $i<j$) are changed and further x_i is changed from FALSE to TRUE and x_j from TRUE to FALSE. Then $\mathcal{A}^{2\sharp}$ is defined as the set of all truth assignments A such that A satisfies $\varphi_{n,m}$ but no single or double flip of A does. (It is easily seen that $\mathcal{A}^{2\sharp}$ is non-empty if $\varphi_{n,m}$ is satisfiable.)

Let $r>0$ and define the following functions of r:

$$\begin{aligned}
u &= e^{-r/7},\\
X &= 1-u^3,\\
Z &= -\frac{6u^6\log(1/u)}{1-u^3}+\frac{18u^9\log^2(1/u)}{(1-u^3)^2}\varphi_2\Bigl(\frac{6u^6\log(1/u)}{1-u^3}\Bigr),
\end{aligned}\tag{13}$$

where $\varphi_2(x)$ is the smallest root of $\varphi_2(x)=e^{x\varphi_2(x)}$. ($\varphi_2$ is defined on $[0,e^{-1}]$ only, but it is easily verified that the argument to φ_2 in (13) lies in this interval for any $r\ge 0$.) A long probabilistic argument [5] shows that if $r>0$ is fixed and $m,n\to\infty$ with $m/n\to r$, then for any truth assignment A with $sf(A)$ possible single flips and $df(A)$ possible double flips,

$$\mathbb{P}(A\in\mathcal{A}^{2\sharp})\le 3m^{1/2}\bigl(\tfrac{7}{8}\bigr)^m\bigl(X+o(1)\bigr)^{sf(A)}\bigl(1+\tfrac{1}{n}Z+o(\tfrac{1}{n})\bigr)^{df(A)}$$

and thus [5, equation (20) (with slightly different notation)]

$$\mathbb{E}\,|\mathcal{A}^{2\sharp}|\le 3m^{1/2}\bigl(\tfrac{7}{8}\bigr)^m\sum_A\bigl(X+o(1)\bigr)^{sf(A)}\bigl(1+\tfrac{1}{n}Z+o(\tfrac{1}{n})\bigr)^{df(A)}.\tag{14}$$

Coding the truth assignments A as sequences $\varepsilon_1,\dots,\varepsilon_n$, where (unconventionally) 0 represents TRUE and 1 FALSE, we see that the sum in (14) is S_n as defined in (1), with $a=\log\bigl(X+o(1)\bigr)=\log X+o(1)$ and $b=n\log\bigl(1+\frac{1}{n}Z+o(\frac{1}{n})\bigr)=Z+o(1)$, and it follows from Theorem 2 (and Remark 2) that

$$\mathbb{E}\,|\mathcal{A}^{2\sharp}|\le\exp\Bigl(n\bigl(r\log(7/8)+\psi(\log X,Z)+o(1)\bigr)\Bigr).$$

In particular, if r is such that

$$r\log(7/8)+\psi(\log X,Z)<0,\tag{15}$$

then $\mathbb{P}(\varphi_{n,m}\text{ is satisfiable})\le\mathbb{E}\,|\mathcal{A}^{2\sharp}|\to 0$ as $n\to\infty$, and thus $\gamma^*\le r$. (By continuity of the left hand side of (15) in r, in fact $\gamma^*<r$.)

A numerical calculation (performed with a small Pascal progam) shows that (15) indeed holds for $r=4.596$.

Remark 7. The proof in [5] uses instead of Theorem 2 the estimate, proved in [4], that if $2a\le b/n\le 0$, then $S_n\le\prod_{k=0}^{n-1}(1+e^{a+ib/2n})$, and thus

$$\overline{\lim}\,\tfrac{1}{n}\log S_n\le\frac{2}{b}\int_a^{a+b/2}\log(1+e^x)\,dx=\frac{2}{b}\bigl(\operatorname{dilog}(1+e^a)-\operatorname{dilog}(1+e^{a+b/2})\bigr).$$

Remark 8. The function φ_2 in (13) comes from an application of a version of Suen's inequality [3]. It is almost certainly not the best possible, but no better results are known at present.

References

[1] M. Abramowitz & I.E. Stegun, Eds., *Handbook of Mathematical Functions.* U.S. Department of Commerce, National Bureau of Standards, Washington, DC, 1972.
[2] E. Friedgut, Necessary and sufficient conditions for sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, to appear.
[3] S. Janson, New versions of Suen's correlation inequality. *Random Struct. Alg.* **13** (1998), 467–483.
[4] L. Kirousis, E. Kranakis & D. Krizanc, An upper bound for a basic hypergeometric series. Tech. Report TR-96-07, Computer Technology Institute, Carleton University, Canada, 1996.
[5] L.M. Kirousis, E. Kranakis, D. Krizanc & Y.C. Stamatiou, Approximating the unsatisfiability threshold of random formulas. *Random Struct. Alg.* **12** (1998), 253–269.
[6] A. Frieze & S. Suen, Analysis of two simple heuristics on a random instance of k-SAT. *J. Algorithms* **20** (1996), 312–355.
[7] Y.C. Stamatiou and M. Vamvakari, A bound for the Rogers–Szegö polynomials with an application to the threshold conjecture of random SAT problems. Preprint, 1999.

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden
E-mail address: svante.janson@math.uu.se