On the Threshold for Random [Max] k-SAT

Dimitris Achlioptas
Microsoft

Based on joint works with

Cris Moore (UNM/Santa Fe)

Assaf Naor (Microsoft)

Yuval Peres (Berkeley)

Satisfiability

Given a Boolean formula (CNF), decide if a satisfying truth assignment exists.

$$(\overline{x}_{12} \vee x_5) \wedge (x_{34} \vee \overline{x}_{21} \vee x_5 \vee \overline{x}_{27}) \wedge \cdots \wedge (x_{12}) \wedge (x_{21} \vee x_9 \vee \overline{x}_{13})$$

Cook's Theorem: Satisfiability is NP-complete.

---

k-SAT: Each clause has exactly k literals.

Since the mid-70s a number of models have been proposed for Random SATisfiability.

Most models generate formulas that are too easy.

Random k-SAT

- Let $\mathcal{L}(n)$ be the set of $2n$ literals $x_1, \bar{x}_1, x_2, \bar{x}_2, \ldots, x_n, \bar{x}_n$.
- Form a random k-SAT formula $\mathcal{F}_k(n, m)$ as follows:

Generate $k \times m$ i.i.d. uniformly random literals from $\mathcal{L}(n)$

Does $\mathcal{F}_k(n, m)$ have a satisfying assignment?

Conjecture: For each $k \geq 3$, there exists a constant r_k such that

$$\lim_{n \to \infty} \Pr[\mathcal{F}_k(n, rn) \text{ is satisfiable}] = \begin{cases} 1 & \text{if } r < r_k \\ 0 & \text{if } r > r_k \end{cases}$$

In other words

The energy of a truth assignment $\sigma \in \{-1, +1\}^n$ in a k-SAT formula with clauses $c_1, \ldots, c_m$

$$E(\sigma) = \sum_{c_i} \prod_{j=1}^{k} \left(1 - \frac{1 + \sigma_{ij}\ell_{ij}}{2}\right)$$

So, random k-SAT is a mean-field, diluted, spin glass with k-wise interactions

Satisfying truth assignment are states with energy 0

First moment method

For any non-negative, integer-valued random variable X,

$$\Pr[X > 0] \;=\; \sum_{x>0} \Pr[X = x] \;\leq\; \sum_{x>0} \Pr[X = x]\, x \;=\; \mathbf{E}[X]\ .$$

Let X be the number of satisfying truth assignments of $\mathcal{F}_k(n, m = rn)$.

For every t.a. σ, by clause-independence, $\Pr[\sigma \text{ is satisfying}] = \left(1 - \frac{1}{2^k}\right)^m$. So,

$$\begin{aligned} \mathbf{E}[X] &= \mathbf{E}\left[I_1 + \cdots + I_{2^n}\right] \\ &= \left(2\left(1 - \frac{1}{2^k}\right)^r\right)^n\ . \end{aligned}$$

But $2\left(1 - \frac{1}{2^k}\right)^r < 1$ for all $r \geq 2^k \ln 2$, implying $\mathbf{E}[X] = o(1)$ for such r. Thus,

$$r_k < 2^k \ln 2\ .$$

Unit-Clause Propagation

```
Repeat
   - Pick an unset variable at random and assign it 0/1 at random
   - While there are unit clauses
        pick any one and satisfy it
```

- Value assignments are permanent (no backtracking)
- Failure occurs iff a 0-clause is ever generated

[Chao Franco 86]: For all $k \geq 3$, if

$$r < \frac{2^k}{k}$$

Unit-Clause propagation finds a satisfying t.a. with probability $\phi = \phi(k, r) > 0$.

More previous work

- $r_k \geq \frac{3}{8}\, 2^k/k$ **[Chvátal Reed 92]**

- $r_k \geq c_k\, 2^k/k$, where $\lim_{k\to\infty} c_k = 1.817\ldots$. **[Frieze Suen 96]**

- $r_k \leq 2^k \ln 2 - d_k$, where $\lim_{k\to\infty} d_k = (1+\ln 2)/2$ **[Kirousis et al. 98]**

No asymptotic progress over

$$\frac{2^k}{k} < r_k < 2^k$$

in more than 15 years.

This talk

[A., Moore ’02]:

$$2^{k-1}\ln 2 - 2 < r_k < 2^k \ln 2$$

[A., Peres ’03]:

$$\frac{r_k}{2^k \ln 2} \;\rightarrow\; 1$$

[A., Naor, Peres ’03]: For all $p \in [0,1]$, let $r_k(p)$ be the threshold for having a truth assignment that satisfies $(1-2^{-k}+p2^{-k})m$ clauses.

$$\frac{r_k(p)}{2^k \ln 2} \;\rightarrow\; \frac{1}{p+(1-p)\log(1-p)}$$

Second moment method

For any non-negative random variable X,

$$\Pr[X > 0] \geq \frac{\mathbf{E}[X]^2}{\mathbf{E}[X^2]} \ .$$

Let X be the number of satisfying truth assignments of $\mathcal{F}_k(n, m = rn)$.

$$\begin{aligned} \mathbf{E}[X^2] &= \mathbf{E}[(I_1 + \cdots + I_{2^n})^2] \\ &= \sum_{\sigma,\tau} \mathbf{E}[I_\sigma I_\tau] \\ &= \sum_{\sigma,\tau} \Pr[\text{Both } \sigma \text{ and } \tau \text{ are satisfying}] \ . \end{aligned}$$

Overlap is what matters. If σ, τ agree on $z = \alpha n$ variables and c is a random clause,

$$\begin{aligned} \Pr[\text{Both } \sigma \text{ and } \tau \text{ satisfy } c] &= 1 - 2^{-k+1} + \frac{\alpha^k}{2^k} \\ &\equiv f(\alpha) \ . \end{aligned}$$

Focus on the middle terms

$$\sum_{\sigma,\tau} \Pr[\text{Both } \sigma, \tau \text{ are satisfying}] = 2^n \sum_{z=0}^{n} \binom{n}{z} f(z/n)^{rn}$$

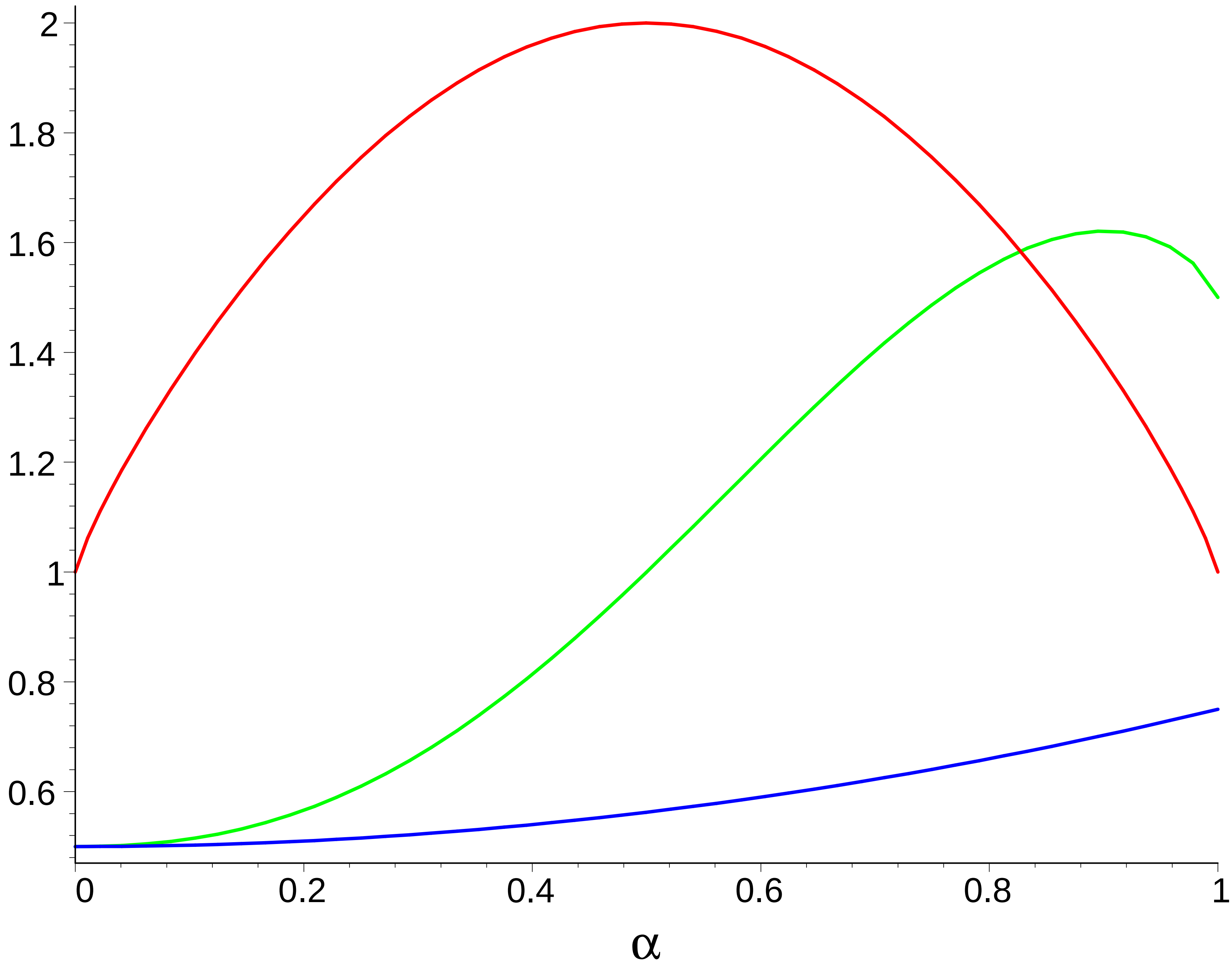
$$\geq 2^n \max_z \binom{n}{z} f(z/n)^{rn}$$

$$\sim \left[\max_{\alpha \in [0,1]} \frac{2 f(\alpha)^r}{\alpha^\alpha (1-\alpha)^{1-\alpha}} \right]^n \qquad (\alpha \equiv z/n)$$

$$\equiv \left(\max_{\alpha \in [0,1]} g_r(\alpha) \right)^n .$$

Observe that $\mathbf{E}[X]^2 = g_r(1/2)^n$. So, g_r better be maximized at $\alpha = 1/2$.

But $f'(1/2) \neq 0$:-(



2
1.8
1.6
1.4
1.2
1
0.8
0.6
0
0.2
0.4
0.6
0.8
1
α

Random NAE k-SAT

Given a k-CNF, is there a truth assignment under which every clause has

at least one satisfied literal and at least one unsatisfied literal?

Let X be the number of NAE-satisfying truth assignments of $\mathcal{F}_k(n, m = rn)$.

$$\begin{aligned} \mathbf{E}[X^2] &= \mathbf{E}[(I_1 + \cdots + I_{2^n})^2] \\ &= \sum_{\sigma,\tau} \mathbf{E}[I_\sigma I_\tau] \\ &= \sum_{\sigma,\tau} \Pr[\text{ Both } \sigma \text{ and } \tau \text{ are NAE-satisfying}] \; . \end{aligned}$$

Again, overlap is what matters. If σ, τ agree on $z = \alpha n$ variables and c is a random clause,

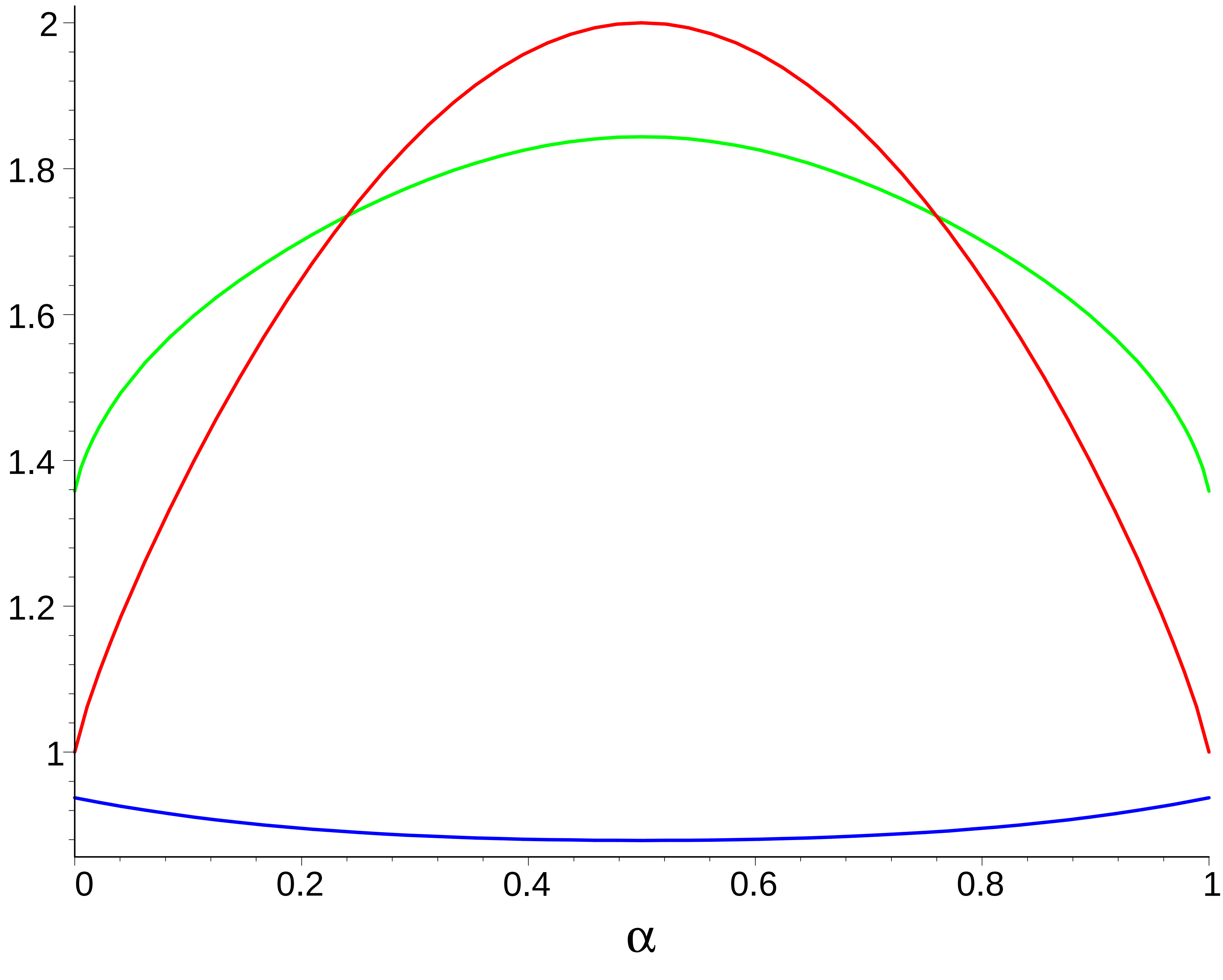
$$\begin{aligned} \Pr[\text{Both } \sigma \text{ and } \tau \text{ NAE-satisfy } c] &= 1 - 2^{-k+2} + \frac{\alpha^k + (1-\alpha)^k}{2^{k-1}} \\ &\equiv f_N(\alpha) \; . \end{aligned}$$

Focus on the middle terms (again)

$$\begin{aligned}\sum_{\sigma,\tau} \Pr[\text{Both } \sigma,\tau \text{ are NAE-satisfying}] &= 2^n \sum_{z=0}^{n} \binom{n}{z} f_N(z/n)^{rn} \\ &= 2^n \times \sum_{\alpha} \left[\binom{n}{\alpha n} f_N(\alpha)^{rn} \right] \qquad (\alpha \equiv z/n) \\ &\leq C \times \left[\max_{\alpha \in [0,1]} \frac{2 f_N(\alpha)^r}{\alpha^\alpha (1-\alpha)^{1-\alpha}} \right]^n \\ &\equiv C \times \left(\max_{\alpha \in [0,1]} \psi_r(\alpha) \right)^n .\end{aligned}$$

Again, $\mathbf{E}[X]^2 = \psi_r(1/2)^n$. So, for all r such that ψ_r is maximized at $\alpha = 1/2$,

$$\Pr[X > 0] \geq \frac{\mathbf{E}[X]^2}{\mathbf{E}[X^2]} \geq 1/C .$$


2
1.8
1.6
1.4
1.2
1
0
0.2
0.4
0.6
0.8
1
α

The random NAE k-SAT threshold

Theorem: There exists a sequence $\epsilon_k \to 0$ such that for all $k \geq 3$, if

$$r \leq 2^{k-1} \ln 2 - \frac{\ln 2}{2} - \frac{1}{2} - \epsilon_k \ ,$$

then w.h.p. $\mathcal{F}_k(n, rn)$ is NAE-satisfiable.

[Refined f.m.]: There exists a sequence $\delta_k \to 0$ such that for all $k \geq 3$, if

$$r \geq 2^{k-1} \ln 2 - \frac{\ln 2}{2} - \frac{1}{4} - \delta_k \ ,$$

then w.h.p. $\mathcal{F}_k(n, rn)$ is not NAE-satisfiable.

---

k	3	4	5	6	7	8	9	10	11	12
Upper bound	2.214	49/12	10.505	21.590	43.768	88.128	176.850	354.295	709.186	1418.969
Lower bound	3/2	4.969	9.973	21.190	43.432	87.827	176.570	354.027	708.925	1418.712

A challenge for 1-step RSB

1-step RSB matches the rigorous upper bound

$$2^{k-1}\ln 2 - \frac{\ln 2}{2} - \frac{1}{4} - o(1)$$

Conjecture: The NAE k-SAT threshold occurs at the rigorous lower bound

$$2^{k-1}\ln 2 - \frac{\ln 2}{2} - \frac{1}{2} - o(1)$$

Why?

Intuition: NAE-assignments look like a “mist" on $\{-1,+1\}^n$. SAT-assignments don’t.

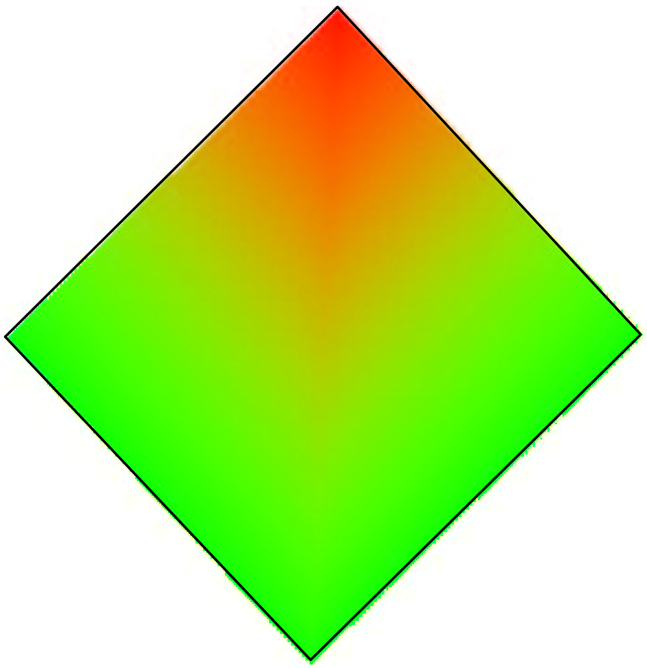
Where does the clustering come from?

Useful fact: $\mathcal{F}_k(n,m)$ is “equivalent" to k-SAT formulas generated by

- Step 1: Creating X_i copies of each literal, where $\{X_i\}_{i=1}^{2n}$ are i.i.d. Poisson r.v.
- Step 2: Partitioning the literals randomly into k-clauses.

Modest assignments

- For a given $\sigma \in \{-1,+1\}^n$, let $S(\sigma)$ be the number of literal copies satisfied by s.
- At the end of Step 1, S is a smooth function on $\{-1,+1\}^n$.
- An exponential number of t.a. “can feel" the majority assignment...



Satisfiability and Populism

For a random truth assignment σ in a random formula with m k-clauses

$$\mathbf{E}[S(\sigma)] = \frac{km}{2}$$

But if we condition on σ being a satisfying truth assignment in $\mathcal{F}_k(n, m)$,

$$\mathbf{E}[S(\sigma)] = \frac{km}{2} \times \frac{2^k}{2^k - 1} \ .$$

Observe: But NAE-satisfiability does not increase the conditional expectation of $L(s)$.

Idea: Look for satisfying assignments with $S(\sigma) = \dfrac{km}{2} \pm O(\sqrt{km})$.

Modest assignments via weighting

- Given any k-SAT formula F, let $\mathcal{G} \subseteq \{-1,+1\}^n$ be the set of satisfying t.a. of F.
- Given $\sigma \in \{-1,+1\}^n$ let $H = H(\sigma, F)$ be the number of satisfied literal copies F under σ minus the number of unsatisfied literal copies.
- For any $0 < \gamma \leq 1$, let

$$X = X(F) = \sum_{\sigma} \gamma^{H(\sigma,F)} \mathbf{1}_{\sigma \in \mathcal{G}(F)} \quad .$$

- **Proof:** Apply second moment method to $X(\mathcal{F}_k(n,m))$ for the right value of $\gamma = \gamma(k)$.

For

$$r \leq 2^k \ln 2 - \frac{k}{2} - O(1)$$

the maximum occurs at $\alpha = 1/2$.

Modest assignments for random Max k-SAT

- Define H as before.
- Given $\sigma \in \{-1,+1\}^n$ let $U = U(\sigma, F)$ be the number of unsatisfied clauses by σ in F.
- For any $0 < \gamma \leq 1$ and $0 < \eta \leq 1$, let

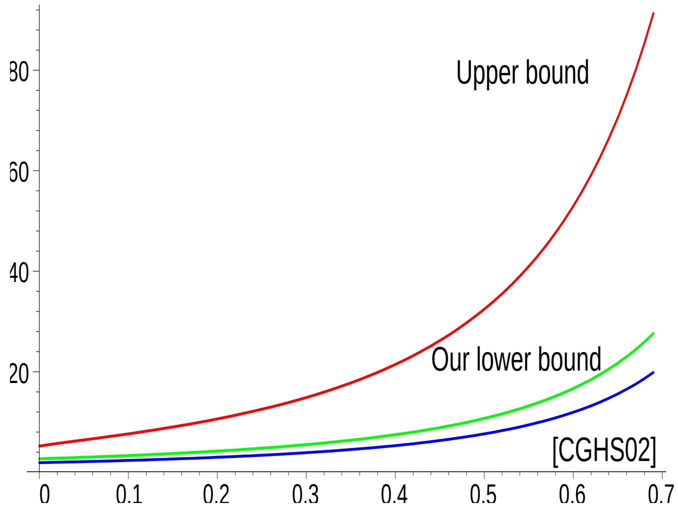
$$X = X(F) = \sum_{\sigma} \gamma^{H(\sigma,F)} \eta^{U(\sigma,F)} .$$

- **Proof:** Apply second moment method to $X(\mathcal{F}_k(n,m))$ for the right combination of γ, η.

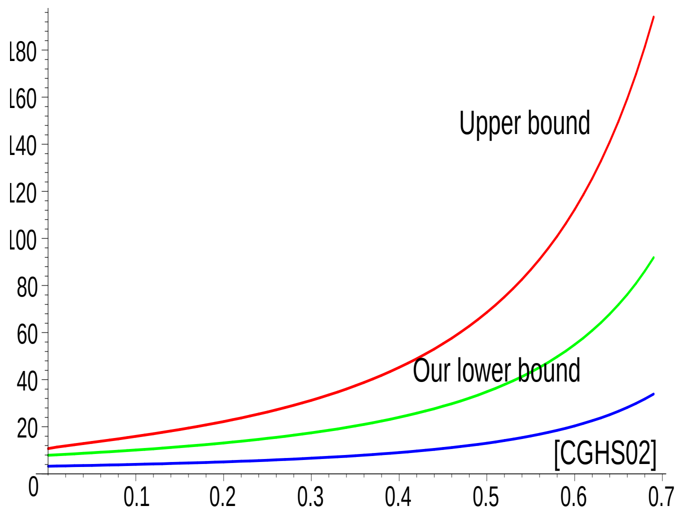
[A., Naor, Peres '03]: For all $p \in [0,1]$, let $r_k(p)$ be the threshold for having a truth assignment that satisfies $(1 - 2^{-k} + p2^{-k})m$ clauses.

$$\frac{r_k(p)}{2^k \ln 2} \to \frac{1}{p + (1-p)\log(1-p)}$$

$k = 3$

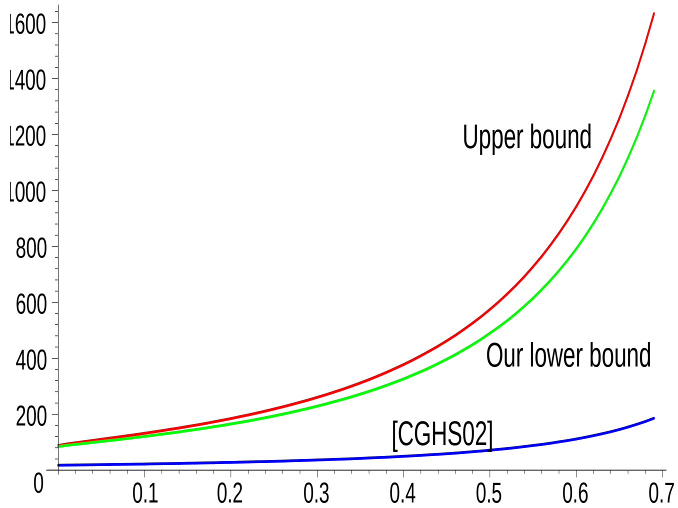


$k = 4$

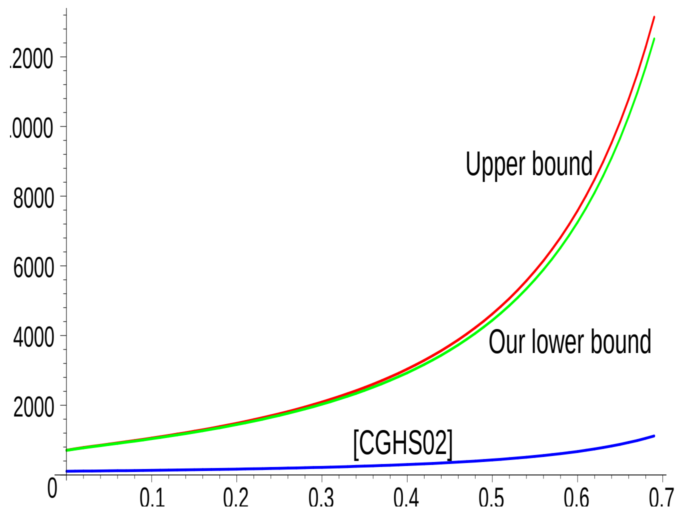


Upper and lower bounds for $r_k(p)$ as a function of $1 - p$.

$k = 7$



$k = 10$



Upper and lower bounds for $r_k(p)$ as a function of $1 - p$.