The threshold for random k-SAT is $2^k \ln 2 - O(k)$

Dimitris Achlioptas

Microsoft

Yuval Peres

Berkeley

Satisfiability

Given a Boolean formula (in 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 $C_k(n)$ be the set of all $2^k \binom{n}{k}$ possible k-clauses on $x_1, x_2, \ldots, x_n$.
- Form a random k-CNF formula $\mathcal{F}_k(n, m)$ as follows:

  Select uniformly, independently and with replacement m clauses from $C_k(n)$.

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

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

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

First moment method

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

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

First moment method

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

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

- 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
```

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

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

$$r < 2^k / k$$

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 $\lim_{k\to\infty} c_k = 1.817....$ **[Frieze Suen 96]**

No asymptotic progress over

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

in more than 15 years.

Many natural DPLL algorithms require exponential time if **[A., Beame, Molloy 01]**

$$r > c_{\mathcal{A}}\, \frac{2^k}{k}$$

Random NAE k-SAT

Given a k-CNF, is there a truth assignment such that every clause has:

at least one satisfied literal **AND** at least one unsatisfied literal?

Random NAE k-SAT

Given a k-CNF, is there a truth assignment such that every clause has:

at least one satisfied literal **AND** at least one unsatisfied literal?

[A., Moore 02]: For all $k \geq 3$, if

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

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

Random NAE k-SAT

Given a k-CNF, is there a truth assignment such that every clause has:

at least one satisfied literal **AND** at least one unsatisfied literal?

[A., Moore 02]: For all $k \geq 3$, if

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

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

So,

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

This talk

Main Theorem: For all $k \geq 3$, if

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

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

Remark: In fact, for all such r there exist "balanced" satisfying assignments.

Explicit bounds for the threshold for random k-SAT for all $k \geq 3$

---

k	3	4	5	7	10	20	21
Upper bound	4.51	10.23	21.33	87.88	708.94	726,817	1,453,635
Lower bound	2.68	7.91	18.79	84.82	704.94	726,809	1,453,626

Preliminaries

[Laplace method]: For any twice-differentiable function f:

$$\sum_{z=0}^{n} \binom{n}{z} f\left(z/n\right)^n = \left[\max_{0\leq\alpha\leq 1} \frac{f(\alpha)}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\right]^n \times \Theta(1)$$

Preliminaries

[Laplace method]: For any twice-differentiable function f:

$$\sum_{z=0}^{n} \binom{n}{z} f\left(z/n\right)^n = \left[\max_{0\le\alpha\le 1} \frac{f(\alpha)}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\right]^n \times \Theta(1)$$

[Friedgut 97]: If $\liminf_{n\to\infty} \Pr[\mathcal{F}_k(n, r^* n)$ is [NAE-]satisfiable$] > 0$ then for all $r < r^*$

$$\Pr[\mathcal{F}_k(n, rn) \text{ is [NAE-]satisfiable}] = 1 - o(1) \ .$$

Preliminaries

[Laplace method]: For any twice-differentiable function f:

$$\sum_{z=0}^{n} \binom{n}{z} f\left(z/n\right)^n = \left[\max_{0\le\alpha\le1} \frac{f(\alpha)}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\right]^n \times \Theta(1)$$

[Friedgut 97]: If $\liminf_{n\to\infty} \Pr[\mathcal{F}_k(n, r^*n)$ is [NAE-]satisfiable$] > 0$ then for all $r < r^*$

$$\Pr[\mathcal{F}_k(n, rn) \text{ is [NAE-]satisfiable}] = 1 - o(1) \ .$$

Useful fact: $\mathcal{F}_k(n, m)$ is asymptotically equivalent ("contiguous") to

- Step 1: Generate randomly $k \times m$ i.i.d. literals (uniformly).
- Step 2: Partition the literals randomly into k-clauses.

Second moment method

For any non-negative random variable X,

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

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 $\mathcal{F}_k(n, m)$. Then

$$\mathbf{E}[X^2] \quad = \quad \mathbf{E}[(I_1 + \cdots + I_{2^n})^2]$$

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 $\mathcal{F}_k(n, m)$. Then

$$\begin{aligned} \mathbf{E}[X^2] &= \mathbf{E}[(I_1 + \cdots + I_{2^n})^2] \\ &= \sum_{\sigma,\tau} \mathbf{E}[I_\sigma I_\tau] \end{aligned}$$

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 $\mathcal{F}_k(n, m)$. Then

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

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 $\mathcal{F}_k(n, m)$. Then

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

Focus on the middle

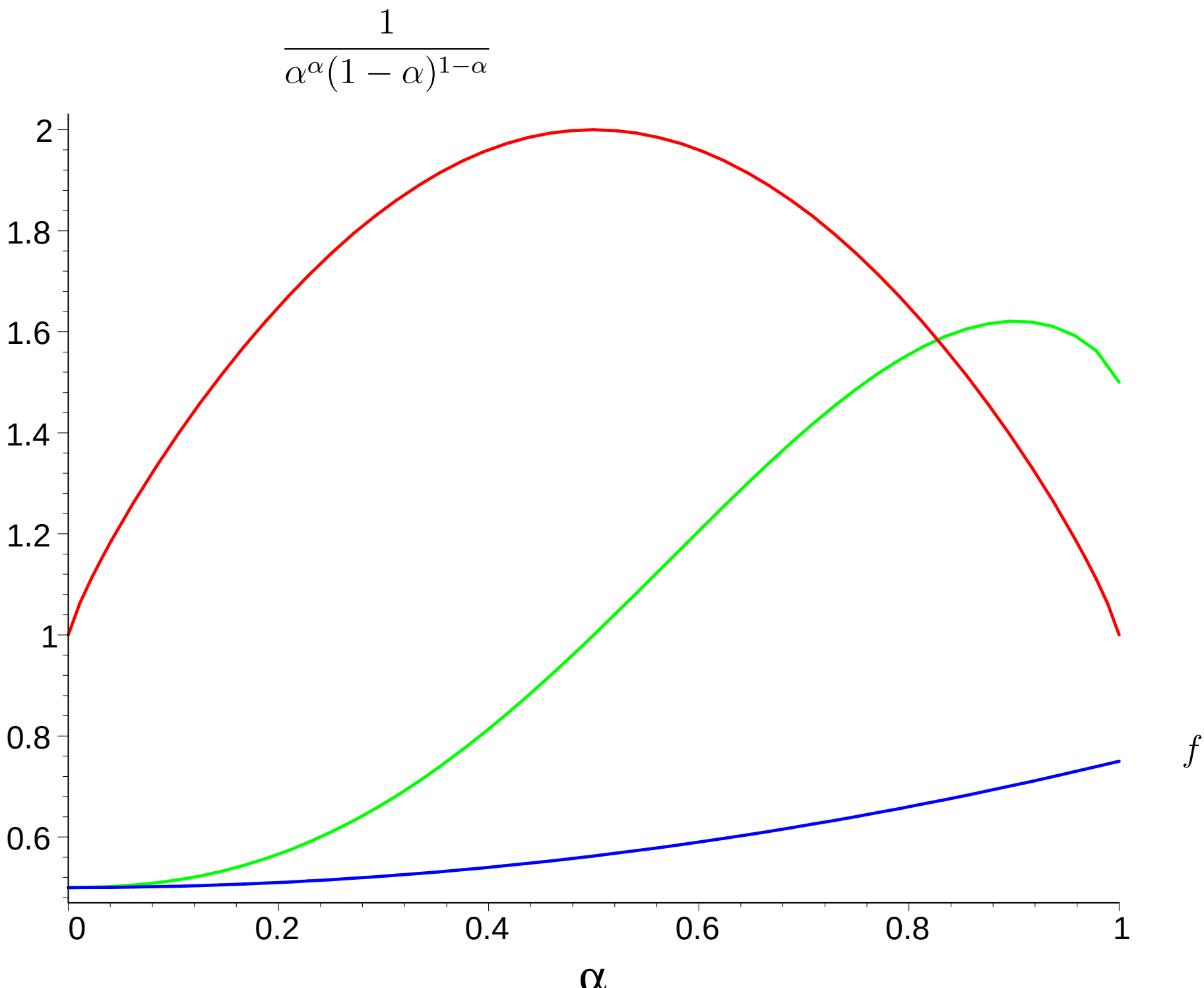
If σ, τ agree on $z = \alpha n$ variables

$$\Pr[\text{Both } \sigma \text{ and } \tau \text{ satisfy a random clause } c] \quad = \quad 1 - 2^{-k+1} + \frac{\alpha^k}{2^k} \quad \equiv \quad f(\alpha) \ .$$

So, all in all, $\mathbf{E}[X^2]$

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

It is easy to see that $\mathbf{E}[X]^2 = g_r(1/2)^n$. (Sharing $n/2$ bits $=$ independence)


$\frac{1}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}$
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
f

Random NAE k-SAT

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

$$\begin{aligned}\mathbf{E}[Y^2] &= \sum_{\sigma,\tau} \Pr[\text{ Both } \sigma \text{ and } \tau \text{ are NAE-satisfying}] \\ &= \sum_{z=0}^{n} \binom{n}{z} f_N(z/n)^m\end{aligned}$$

where

$$f_N(\alpha) = \left(1 - 2^{-k+2} + \frac{\alpha^k + (1-\alpha)^k}{2^{k-1}}\right)$$

Random NAE k-SAT

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

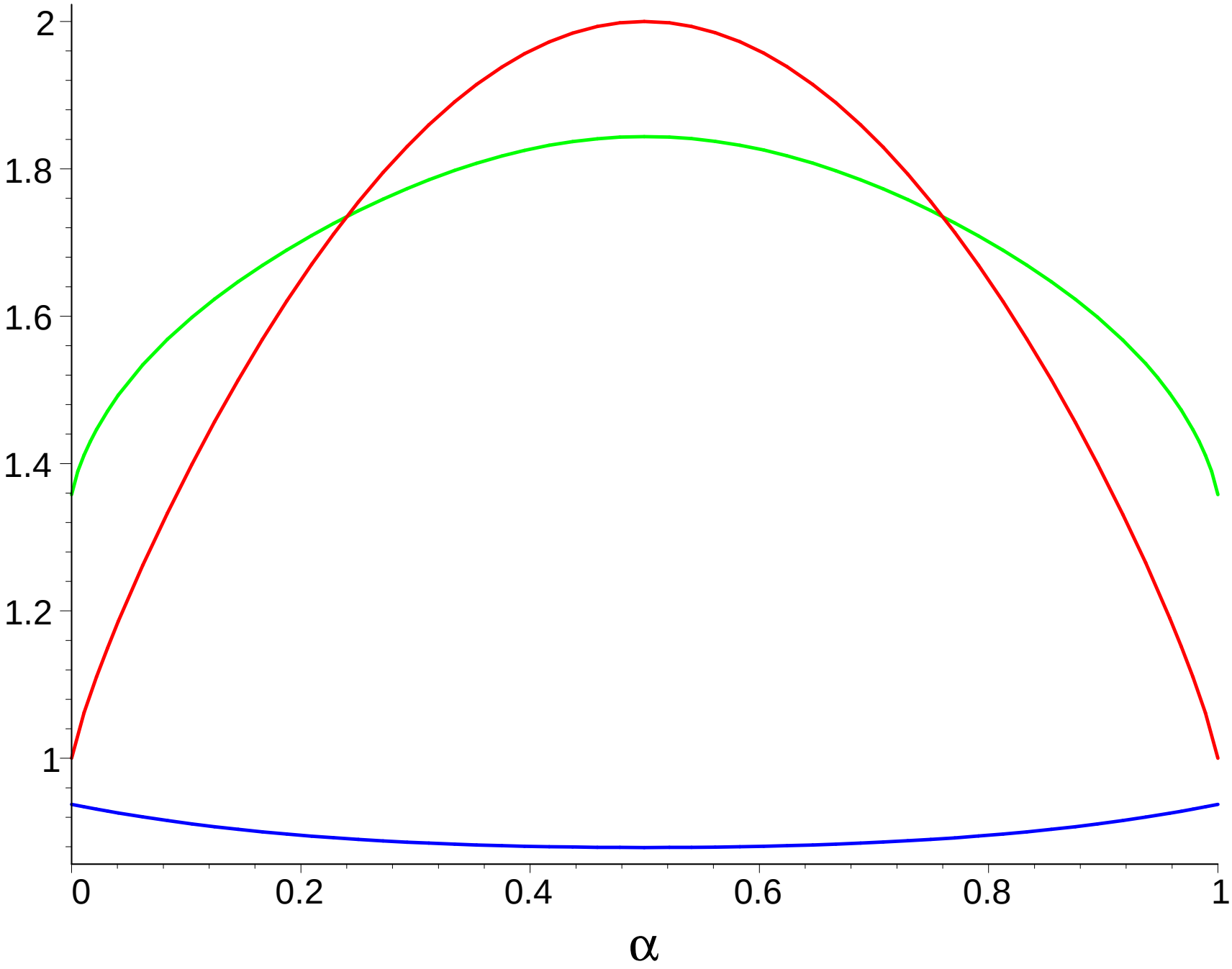
$$\begin{aligned}\mathbf{E}[Y^2] &= \sum_{\sigma,\tau} \Pr[\text{ Both } \sigma \text{ and } \tau \text{ are NAE-satisfying}] \\ &= \sum_{z=0}^{n} \binom{n}{z} f_N(z/n)^m\end{aligned}$$

where

$$f_N(\alpha) = \left(1 - 2^{-k+2} + \frac{\alpha^k + (1-\alpha)^k}{2^{k-1}}\right)$$

If X is the number of satisfying truth assignments we instead have

$$f(\alpha) = 1 - 2^{-k+1} + \frac{\alpha^k}{2^k}$$


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

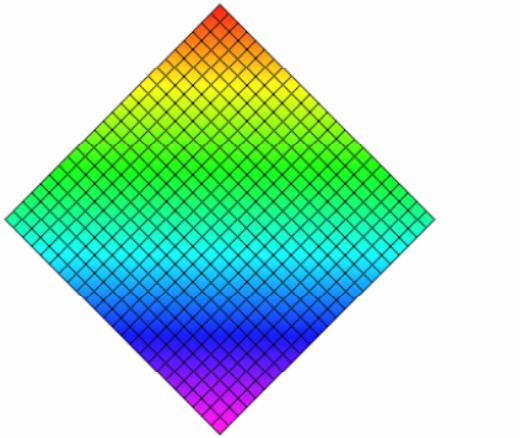
Where does the clustering come from?

Let $L(\sigma)$ be the number of literals satisfied by each $\sigma \in \{0,1\}^n$ at the end of Step 1.

Where does the clustering come from?

Let $L(\sigma)$ be the number of literals satisfied by each $\sigma \in \{0,1\}^n$ at the end of Step 1.

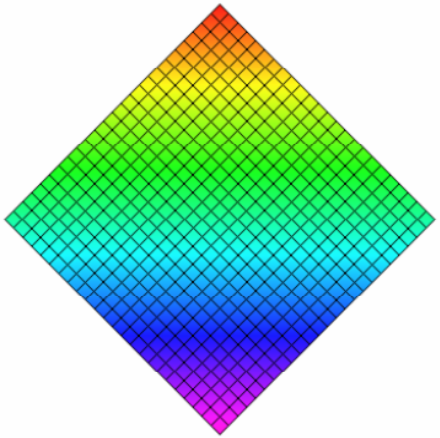
Majority vote assignment



Where does the clustering come from?

Let $L(\sigma)$ be the number of literals satisfied by each $\sigma \in \{0,1\}^n$ at the end of Step 1.

Majority vote assignment

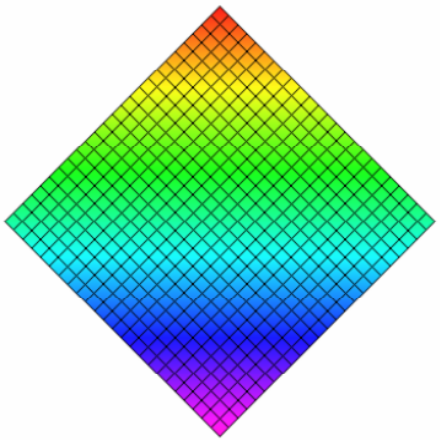


Our starting point (intuition): Correlation towards majority is the main source of clustering.

Where does the clustering come from?

Let $L(\sigma)$ be the number of literals satisfied by each $\sigma \in \{0,1\}^n$ at the end of Step 1.

Majority vote assignment



Our starting point (intuition): Correlation towards majority is the main source of clustering.

Sanity check: NAE k-SAT assignments are not correlated with the majority.

Balanced assignments & Analysis

Let $X = \sum_{\sigma} \text{weight}(\sigma)$ where $\text{weight}(\sigma) = \prod_{c} w(\sigma, c)$ (SAT, NAE)

Balanced assignments & Analysis

Let $X = \sum_{\sigma} \text{weight}(\sigma)$ where $\text{weight}(\sigma) = \prod_{c} w(\sigma, c)$ (SAT, NAE)

We are free to choose any w on $\{0,1\}^k$ such that:

$$w(00\cdots0) = 0 \quad (1)$$

$$f'_w(1/2) = 0 \quad (2)$$

Balanced assignments & Analysis

Let $X = \sum_{\sigma} \text{weight}(\sigma)$ where $\text{weight}(\sigma) = \prod_{c} w(\sigma, c)$ (SAT, NAE)

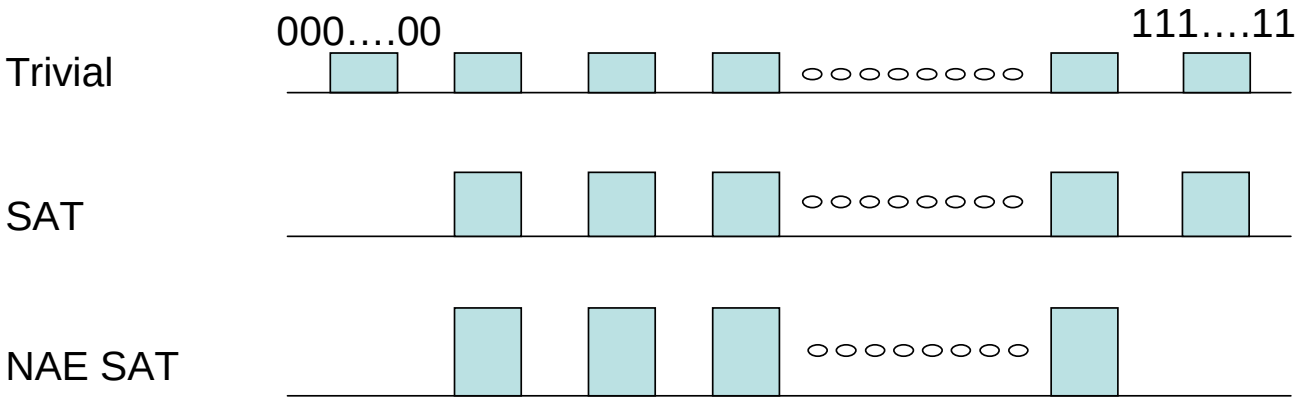
We are free to choose any w on $\{0,1\}^k$ such that:

$$w(00\cdots0) = 0 \quad (1)$$

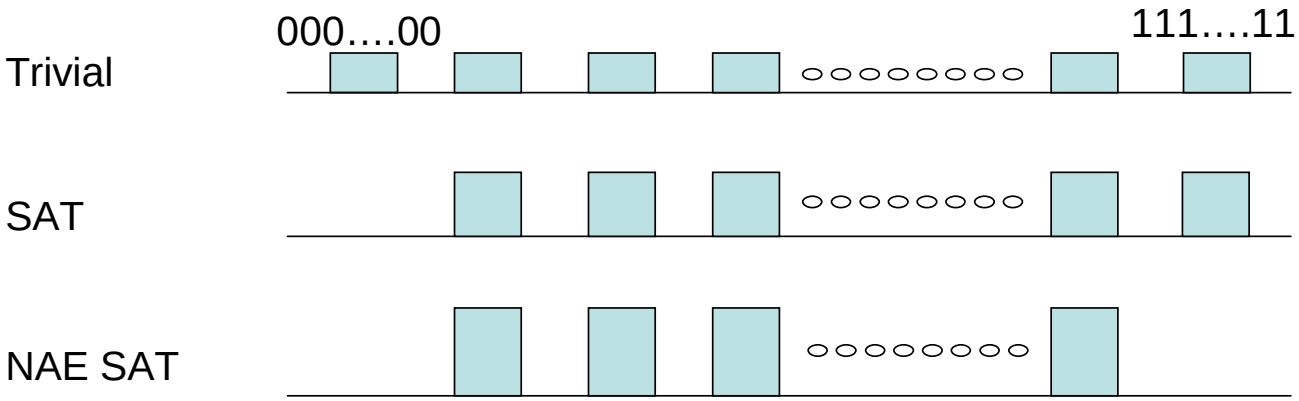
$$f'_w(1/2) = 0 \quad (2)$$

Lemma: Requirement (2) $\iff$ $\mathbf{E}[w(\cdot)] = k/2$.

Balanced assignments & Information Theory

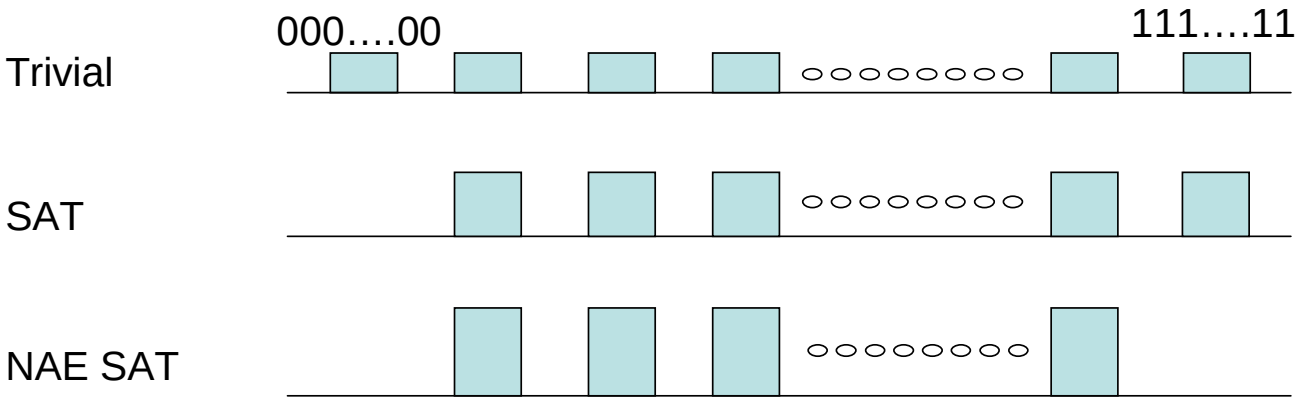


Balanced assignments & Information Theory



What is the least constrained balanced weighting?

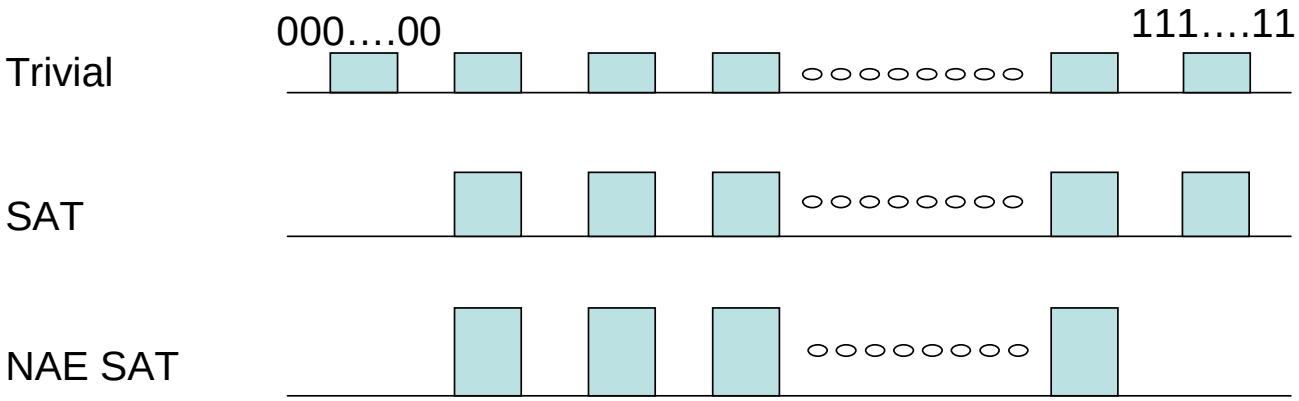
Balanced assignments & Information Theory



What is the least constrained balanced weighting?

Question: Maximize entropy of w subject to: $w(00\cdots0) = 0$ and $\mathbf{E}[w(\cdot)] = k/2$.

Balanced assignments & Information Theory



What is the least constrained balanced weighting?

Question: Maximize entropy of w subject to: $w(00\cdots0) = 0$ and $\mathbf{E}[w(\cdot)] = k/2$.

Answer: Assign to each bucket with $i \geq 1$ ONES weight γ^i, where $\gamma = \gamma(k) < 1$.

The buckets with many satisfied literals “share the pain”.

Concretely

Let $H(\sigma, F)$ be the number of satisfied literals in F under σ. For any $0 < \gamma \leq 1$, let

$$X(F) \quad = \quad \sum_{\sigma} \gamma^{H(\sigma,F)} \, \mathbf{1}_{\sigma \models F}$$

Concretely

Let $H(\sigma, F)$ be the number of satisfied literals in F under σ. For any $0 < \gamma \leq 1$, let

$$\begin{aligned} X(F) &= \sum_{\sigma} \gamma^{H(\sigma,F)} \, \mathbf{1}_{\sigma \models F} \\ &= \sum_{\sigma} \prod_{c} \gamma^{H(\sigma,c)} \, \mathbf{1}_{\sigma \models c} \end{aligned}$$

Concretely

Let $H(\sigma, F)$ be the number of satisfied literals in F under σ. For any $0 < \gamma \leq 1$, let

$$\begin{aligned} X(F) &= \sum_{\sigma} \gamma^{H(\sigma,F)} \mathbf{1}_{\sigma \models F} \\ &= \sum_{\sigma} \prod_{c} \gamma^{H(\sigma,c)} \mathbf{1}_{\sigma \models c} \end{aligned}$$

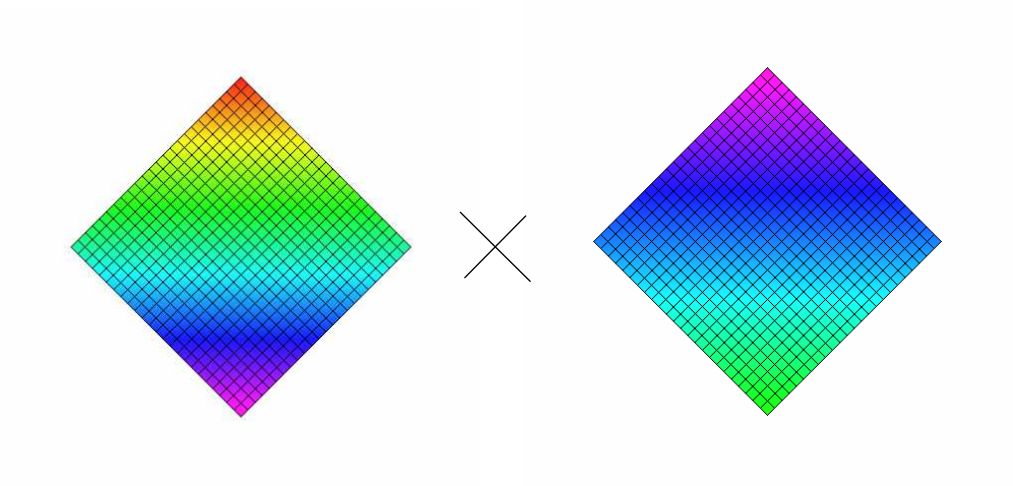
Proof: Apply second moment method to $X(\mathcal{F}_k(n, m))$.

Concretely

Let $H(\sigma, F)$ be the number of satisfied literals in F under σ. For any $0 < \gamma \leq 1$, let

$$\begin{aligned} X(F) &= \sum_{\sigma} \gamma^{H(\sigma,F)} \, \mathbf{1}_{\sigma \models F} \\ &= \sum_{\sigma} \prod_{c} \gamma^{H(\sigma,c)} \, \mathbf{1}_{\sigma \models c} \end{aligned}$$

Proof: Apply second moment method to $X(\mathcal{F}_k(n, m))$.



Conclusions

$$2^k \ln 2 - O(k) < r_k < 2^k \ln 2$$

Satisfying assignments exist way beyond the reach of current algorithms.

Conclusions

$$2^k \ln 2 - O(k) < r_k < 2^k \ln 2$$

Satisfying assignments exist way beyond the reach of current algorithms.

- Is $r_k = 2^k \ln 2 - O(1)$?

Conclusions

$$2^k \ln 2 - O(k) < r_k < 2^k \ln 2$$

Satisfying assignments exist way beyond the reach of current algorithms.

- Is $r_k = 2^k \ln 2 - O(1)$?

Can polynomial time algorithms go beyond $2^k/k$?