"Mick gets more than pie"

Dimitris Achlioptas

Microsoft Research

- Formulas:
  - Boolean variables $x_1, x_2, \ldots, x_n$
  - Conjunctions of disjunctions (CNF)

    e.g. $(x_6) \wedge (\overline{x}_8 \vee x_9 \vee \overline{x}_{12}) \wedge (x_{14} \vee \overline{x}_{12} \vee \overline{x}_6 \vee x_8) \wedge (x_3 \vee \overline{x}_6)$

- SATisfiability:

  Given formula $\mathcal{F}$, decide if $\mathcal{F}$ is satisfiable.

- K-SAT:

  All clauses have exactly k literals.

  e.g. $(x_2 \vee \overline{x}_4) \wedge (x_6 \vee x_9) \wedge (x_4 \vee x_7) \cdots \wedge (\overline{x}_9 \vee x_{12})$

- 2-SAT : Polynomial time
    - Pick any unset variable and set it (1 choice)
    - Repeat until no 1-clauses are left:

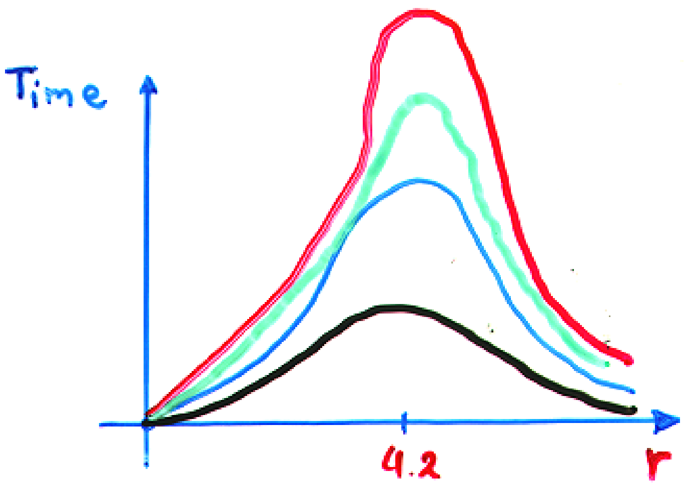
      Pick any 1-clause and satisfy it.

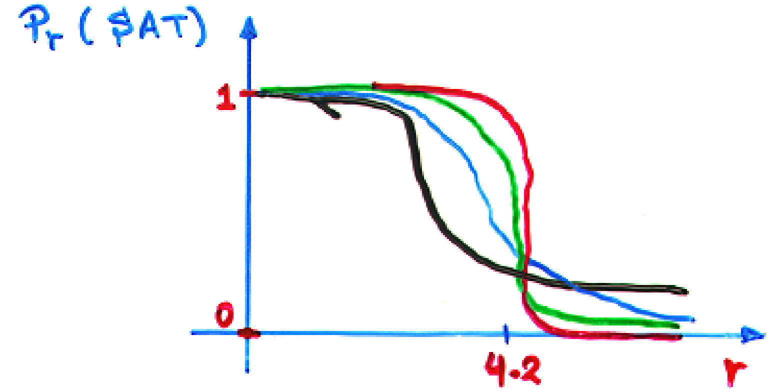
"Either we get a contradiction or a subformula."

- $k \geq 3$: k-SAT is NP-complete.

- For a number of formula distributions, satisfiability can be solved in polynomial time on average.
- For many of these distributions, picking truth assignments at random yields a satisfying assignment in $O(1)$ steps.

Experiments with $F_3(n, r \cdot n)$





[Mitchell, Selman, Levesque '91]

Satisfiability Threshold Conjecture

For each $k \geq 2$, there exists r_k such that

$$\mathcal{F}_k(n, r \cdot n) \begin{cases} \text{is a.s. satisfiable for } r < r_k \\ \text{is a.s. unsatisfiable for } r > r_k \end{cases}$$

"Mick gets some (the odds are on his side)" [Chvátal, Reed 92]

Random k-SAT

- Let A_k be the set of all $2^k \binom{n}{k}$ k-clauses on $x_1, x_2, \dots x_n$.

- $\mathcal{F}_k(n, m)$: Pick m clauses from A_k uniformly at random with replacement.

- Almost surely (a.s.): $\lim_{n \to \infty} P_r[\mathcal{E}_n] = 1$.

Known results

$k=2$

- $r_2 = 1$ [CR 92, Goerdt 92, Fernandez de la Vega 92]
- The transition occurs for $m = n + \lambda \cdot n^{2/3}$, as λ goes from $-\infty$ to $+\infty$.

  [Bollobás, Borgs, Chayes, Kim, Wilson 99]

$k \geq 3$

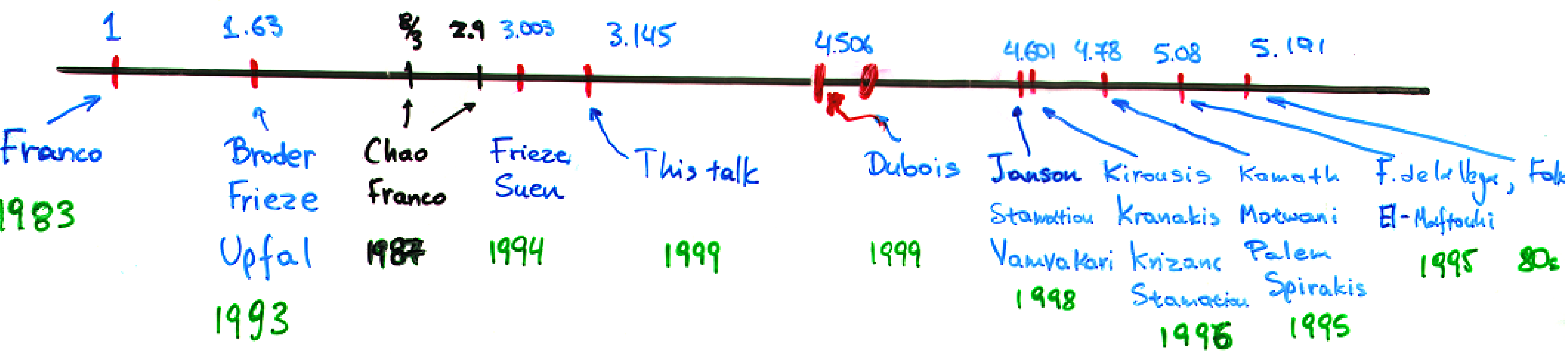
- We don't even know if r_k exists.
- $\frac{2^k}{k} < r_k < 2^k$ [Chao, Franco 87, CR 92, Frieze, Suen 94] – [Folklore]
- For each $k \geq 2$, there exists a function $r_k(n)$ such that for every $\epsilon > 0$,

  $\mathcal{F}_k(n, (r_k(n) - \epsilon) \cdot n)$ is a.s. satisfiable

  $\mathcal{F}_k(n, (r_k(n) + \epsilon) \cdot n)$ is a.s. unsatisfiable

  [Friedgut 97]

Random 3-SAT



- Pure literal heuristic: pick a safe variable and set it.

  $\gamma_3 = 1.637\ldots$, $\gamma_k = \Theta(k)$ [Mitzenmacher 97, Pittel, Spencer, Wormald 96]

- Probabilistic counting arguments for upper bounds.

- Friedgut's Theorem allows "constant probability" results to yield bounds for r_k.

Algorithms

If there exist 1-clauses
 then pick one u.a.r. and satisfy it
 else select a literal ℓ and satisfy it

- Value assignments are permanent
- Continues until all variables are set
- Fails if a 0-clause is ever generated

---

select

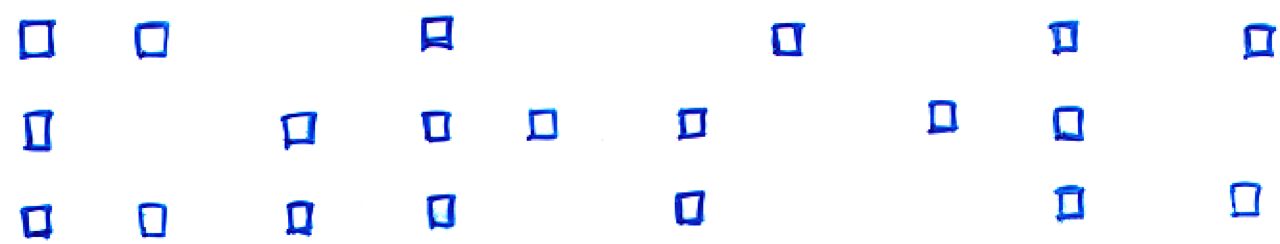
U.C. Pick an unset variable x u.a.r.; pick $\ell \in \{x, \bar{x}\}$ u.a.r. — 8/3

U.C.W.M. Pick an unset variable x u.a.r.; pick $\ell \in \{x, \bar{x}\}$ which appears in fewest remaining 3-clauses — 2.9

G.U.C. Pick a shortest remaining clause $c_i = x_{i_1} \vee \cdots \vee x_{i_j}$ u.a.r.; pick $\ell \in \{x_{i_1}, \ldots, x_{i_j}\}$ u.a.r. — 3.003

Uniform Randomness

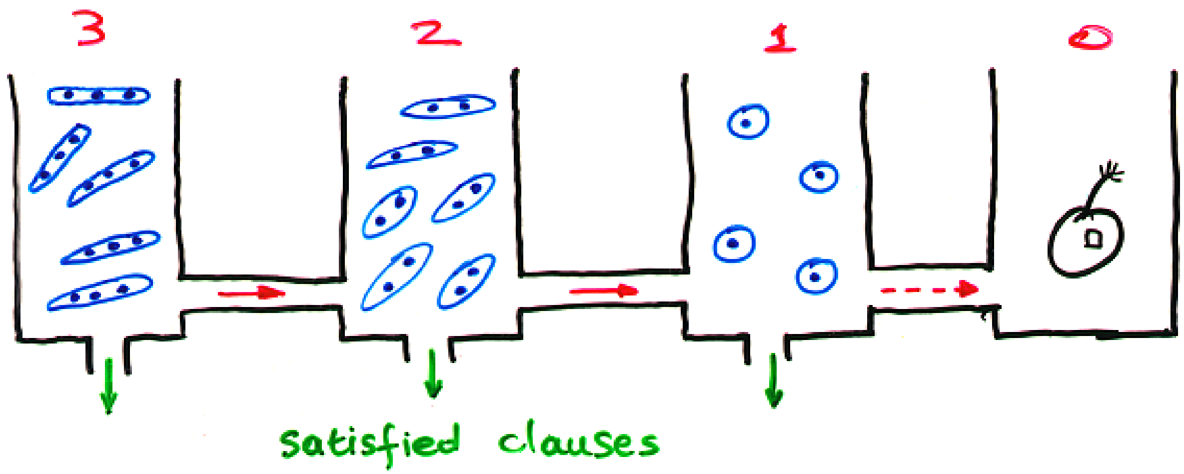
For all $0 \le i \le 3$, the set of i-clauses remaining at the end of each round is uniformly random, conditional on its size.



- Initially all the cards are "face down"; 3 cards per clause
- We can $\left\{ \begin{array}{l} \text{name a variable} \\ \text{or} \\ \text{point at a card} \end{array} \right.$

- All cards with the named/underlying variable turn face up.
- Cards of unsatisfied literals are removed
- Cards of satisfied clauses are removed

Buckets



$C_i(t)$ is the number of i-clauses after t variables are set.

- The probability that an i-clause contains the variable set in round t is $\frac{i}{n-t}$.

- If $t_e = (1-\epsilon)\cdot n$, $\epsilon > 0$, and $\boxed{\sum_{t=0}^{t_e} C_1(t) < B \cdot n}$

  then with probability $e^{-\frac{B}{\epsilon}} > 0$ there are NO 0-clauses at time t_e.

"PROOF":

$$\underbrace{\left(1-\frac{1}{n-t}\right)^{C_1(t)}}_{\text{per step}} < \underbrace{\left(1-\frac{1}{\epsilon\cdot n}\right)^{\sum_{t=0}^{t_e} C_1(t)}}_{\text{total risk}} < e^{-\frac{B}{\epsilon}+o(1)}$$

Tracing $C_i(t)$

- Each 2-clause becomes an 1-clause with probability $\frac{2}{n-t} \times \frac{1}{2}$. Thus, $\boxed{\frac{C_2(t)}{n-t}}$ 1-clauses are generated, on average, in step t. $[\mathrm{Bin}(C_2(t), \frac{1}{n-t})]$
- Therefore, $\sum_{t=0}^{t_e} C_1(t) < B \cdot n$ iff the 2-SAT subformula has density $[C_2(t)/(n-t)] < 1$ for all $0 \le t \le t_e$.
- We need to trace $C_2(t)$.
- $E\left(C_2(t+1) - C_2(t)\right) = f\left(C_3(t), C_2(t), C_1(t), t\right)$.

A "Mediterranean Server" Lemma

Replace: If there are 1-clauses then pick one u.a.r. and satisfy it
else select a literal ℓ and satisfy it

With: • Flip a coin that comes up **1** with probability $(1+\delta)\cdot\frac{c_2(t)}{n-t}$
$\min\{1, \quad \}$

If the coin came up 1 then
If there are 1-clauses then pick one u.a.r. and satisfy it
else satisfy a random literal
else select a literal ℓ and satisfy it.

Lemma: If $\frac{c_2(t)}{n-t} < \frac{1}{1+\delta}$ for all $0 \le t \le t_e$, then a.s.

$$\sum_{t=0}^{t_e} c_1(t) < B\cdot n \quad , \text{ for some } B = B(\delta).$$

Now: $E\big(c_2(t+1) - c_2(t)\big) = f\big(c_3(t), c_2(t), t\big)$.

Differential Equations

- $E\left(C_i(t+1) - C_i(t)\right) = \Theta(1)$, for all $0 \le t \le t_e$
- $Pr\left(C_i(t+1) - C_i(t)\right)$ is large is very small
- The "law" of $C_i(t+1) - C_i(t)$ is smooth with respect to $C_i(t)$, t.

There exist functions f_2, f_3 such that a.s.

$$C_2(t) = f_2(t/n) \cdot n + o(n), \text{ for all } t.$$

Differential equations

$$\boxed{\begin{array}{l} t = x \cdot n \\ C_i(t) = f_i(t/n) \cdot n \end{array}}$$

3-clauses: $E\left(C_3(t+1) - C_3(t)\right) = -\dfrac{3 \cdot C_3(t)}{n-t}$

$$f_3'(x) = -\frac{3 \cdot f_3(x) \cdot n}{(1-x) \cdot n}, \quad f_3(0) = r$$

$$f_3(x) = r \cdot (1-x)^3$$

2-clauses:

U.C. $E\left(C_2(t+1) - C_2(t)\right) = \dfrac{3 \cdot C_3(t)}{2 \cdot (n-t)} - \dfrac{2 \cdot C_2(t)}{n-t}$

$$f_2'(x) = \frac{3 \cdot r \cdot (1-x)^3}{2 \cdot (1-x)} - \frac{2 \cdot C_2(x)}{1-x}, \quad f_2(0) = 0$$

$$f_2(x) = \frac{3}{2} \cdot r \cdot x \cdot (1-x)^2$$

$$\frac{f_2(x)}{1-x} < 1 \quad \text{for all } x \in [0,1] \Leftrightarrow r < 8/3$$

U.C.w.m.

Let $q_r(x) = E(\min\{Z_1, Z_2\})$ where Z_1, Z_2 are i.i.d. Poisson random variables with mean $\frac{3}{2} \cdot r \cdot (1-x)^2$.

$$f_2'(x) = \frac{c_2(x)}{1-x} \cdot \frac{3}{2} \cdot r \cdot (1-x)^2 + \left(1 - \frac{c_2(x)}{1-x}\right) \cdot q_r(x) - \frac{2 \cdot c_2(x)}{1-x}$$

$$\frac{f_2(x)}{1-x} < 1 \iff r < 2.99$$

G.U.C.

There "always" are 2-clauses around.

$$f_2'(x) = \frac{3}{2} \cdot r \cdot (1-x)^2 - \frac{2 c_2(x)}{1-x} - 1 \cdot \left(1 - \frac{c_2(x)}{1-x}\right)$$

$$f_2(x) = \frac{1}{4} \cdot (1-x) \cdot (3rx(2-x) + 4 \cdot \ln(1-x))$$

$$\frac{f_2(x)}{1-x} < 1 \iff r < 3.003\ldots$$

How to get pie

- When there are no 1-clauses we'll set 2 variables
- There "always" are 2-clauses around
- - Pick a 2-clause u.a.r.; let x_1, x_2 be two underlying variables
  - Sort the 4 possible assignments to x_1, x_2 with respect to the number of 3-clauses they satisfy

e.g.

x_1	x_2	
0	1	4
0	0	7
1	1	9 ← second best
1	0	11 ← best

- If the "best" assignment satisfies the 2-clause then use it; else use the second best.

- With probability $\frac{3}{4}$ the best assignment "works"

  Even the "second best" is "better than average".

$$r_3 > 3.145$$

Pie and coffee

- Minimizing the $2 \to 1$ flow X
- Setting more than 2 variables at a time X

  How much further can the "cards model" go?

- Do pure literal as you go along.
- **Bold idea:** since we understand 2-SAT, maybe we can give up uniform randomness for buckets 1,2.

  2-SAT formula will have density > 1 but remain satisfiable.