On Average Case Complexity of SAT

Johann A. Makowsky

Faculty of Computer Science
Technion–Israel Institute of Technology
Haifa, Israel

janos@cs.technion.ac.il
www.cs.technion.ac.il/∼janos

This talk is based on an old, but barely referenced paper:

On Average Case Complexity of SAT
for Symmetric Distributions
Makowsky J. A. and Sharell A.,
Journal of Logic and Computation, 5(1), 71-92 (1995)

I want to put these results into an actual perspective.

The slides were essentially prepared by **Yoni Mircae**

Outline

- Efficient on the average
- Flat distributions
- Distributions for SAT
- Symmetric distributions
- Fixed density distributions
- Resolution of clauses
- More recent work

Efficient on the average

Back to Outline

How to Define "Efficient on the Average"?

- A possible definition would be: an algorithm A is efficient-on-average if it runs in expected polynomial time.
- **Problem**: suppose A runs in time n^2 on all inputs of length n except on one input that takes 2^n. Then, the expected running time of A is:

$$E[T_A] = \frac{2^n - 1}{2^n} \cdot n^2 + \frac{1}{2^n} \cdot 2^n = \mathcal{O}(n^2)$$

- However, if B is a simulation of A that takes T_A^2, the expected running time of B is:

$$E[T_B] = \frac{2^n - 1}{2^n} \cdot n^4 + \frac{1}{2^n} \cdot 2^{2n} = \mathcal{O}(2^n)$$

Average Case Complexity : Basic Definitions

- A function $\mu : S \to [0,1]$ is a *probability density function* (pdf) on a countable or finite set S if

$$\Sigma_{x \in S} \mu(x) = 1.$$

- A *size function* for a set S if a function $|\cdot| : S \to \mathbb{N}^+$ such that the set $S_n = \{x \in S : |x| = n\}$ is finite.

- An *input set* S is a pair $< S, |\cdot| >$.

- Let S be an input set and μ a pdf on S. The pair $< S, \mu >$ is called a *global randomization* of S.

- Let $< S, \mu >$ be a global randomization and μ_n defined by $\mu_n(x) = Pr_\mu\{x | x \in S_n\}$. The sequence $< S_n, \mu_n >$ is called a *local randomization* of S. Note that each μ_n is a pdf on S_n.

Distributional Problem

- Let $< S, \mu >$ be a global randomization, $\leq$ be a linear ordering on S which is polynomial time computable and $D \subseteq S$.
  - Let μ^* be defined by
    $$\mu^*(x) = \Sigma_{y \leq x} \mu(x)$$
    μ is effectively computable if μ^* is polynomial time computable.
  - A pair $< D, \mu >$ with μ effectively computable is called a *distributional problem*.

  We think of D as the set of positive instances of some problem.

Weight Function

- Given a global randomization $< S, \mu >$ we define its *weight function* w by $w(n) = Pr_{\mu}\{S_n\}$. Note that w is a pdf on $\mathbb{N}^+$.

- If for some constant $c > 0$ and for every $n \in \mathbb{N}^+$, $w(n) \geq n^{-c}$ then we say that w is a *regular weight function* and that the global randomization is regular.

- If for every $\varepsilon > 0$:

  $$\Sigma w(n) n^{\varepsilon} = \infty$$

  then we say that w is a *strongly regular weight function* and that the global randomization is strongly regular.

  Example: $w(n) = n^{-1}(logn)^{-2}$

Note that a local randomization with a weight function defines a unique global randomization.

Local Probabilistic Bounds

- Let $< S_n, \mu_n >$ be a local randomization on S and let $f : \mathbb{R}^+ \to \mathbb{R}^+$. Let $T : S \to \mathbb{R}^+$, $E_{\mu_n}(T) = \Sigma_{x \in S_n} T(x)\mu_n(x)$ is the expectation of T on inputs of size n with respect to μ_n. We say that:
  - f is an upper bound on the expectation of T if
$$E_{\mu_n}(T(x)) \leq f(n).$$
  - f is a (local) upper bound in probability on T if
$$lim_{n \to \infty} Pr_{\mu_n}\{T(x) \leq f(n)\} = 1.$$
  - f is a (local) lower bound in probability on T if
$$lim_{n \to \infty} Pr_{\mu_n}\{T(x) > f(n)\} = 1.$$

Results in probabilistic analysis of algorithms are usually expressed with these types of local bounds on T.

Probabilistic Bounds

- For Average Case Complexity Theory we present here a definition of at most f on the average
  that was developed in *:

  Let $< S, \mu >$ be a global randomization on S and $T : S \to \mathbb{R}^+$. For a *strictly increasing* function $f : \mathbb{R}^+ \to \mathbb{R}^+$ we say that T is at most f on the average w.r.t the global randomization $< S, \mu >$ if

$$E_\mu(\frac{f^{-1}(T(x))}{|x|}) < \infty$$

  and denote this by $T \in AVB(< S, \mu >, f)$ or simply $T \in AVB(f)$ if the randomization is evident from context.

*Shai Ben-David, Benny Chor, Oded Goldreich, and Michael Luby. On the theory of average case complete complexity. Journal of Computer and System Sciences, 44(2):193-219, April 1992.

Probabilistic Bounds (cont.)

We can now define a (average) complexity class:

Let $< D, \mu >$ be a distributional problem. We say that $< D, \mu >$ is polynomial on the average and write $< D, \mu > \in AverP$ if there is a deterministic algorithm A for D with run-time T_A and there is a polynomial p such that $T_A \in AVB(p)$.

Theorem 1 *(Transfer Theorem for Upper Bounds): Let $< S, \mu >$ be a global randomization, $< S_n, \mu_n >$ the implied local randomization and $T : S \to \mathbb{R}^+$. For any function $f : \mathbb{R}^+ \to \mathbb{R}^+$:*

- *If f is a convex function then: $E_{\mu_n}(T(x)) \leq f(n) \Rightarrow T \in AVB(f)$*
- *If f is a concave function then: $T \in AVB(f) \Rightarrow E_{\mu_n}(T(x)) \leq f(n)$*

Proof:
Follows from Jensen's inequality: $\phi(\frac{\Sigma a_i x_i}{\Sigma a_i}) \leq \frac{a_i \phi(x_i)}{\Sigma a_i}$ for a real convex function ϕ and positive weights a_i. The inequality is reversed if ϕ is concave.

■

Probabilistic Bounds (cont.)

Theorem 2 *(Transfer Theorem for Lower Bounds *): Let* $< S, \mu >$ *be a global randomization with weight function* w *and let* $f, g : \mathbb{R}^+ \to \mathbb{R}^+$ *be two strictly increasing functions. If* f *is a lower bound in probability on* T *w.r.t* $< S_n, \mu_n >$ *and* g *is sufficiently small for*

$$\Sigma_{n=1}^{\infty} \frac{w(n)}{n} g^{-1}(f(n)) = \infty$$

to hold then

$$T \notin AVB(g).$$

*Abraham Sharell. On the average case complexity of SAT for flat distributions. Master's thesis, Technion-Israel Institute of Technology, 1992.

Proof:
Since g is strictly increasing it suffices to show that $E_\mu(\frac{g^{-1}(T(x))}{|x|}) = \infty$.

Reminder: Markov's inequality: If X is a nonnegative random variable and $a > 0$, then $P(x \geq a) \leq \frac{E(X)}{a}$.

Applying Markov's inequality to the (strictly positive) random variable $g^{-1}(T(x))$ we derive for every $n \in \mathbb{N}^+$

$$E_{\mu_n}(g^{-1}(T(x))) > g^{-1}(f(n))Pr_{\mu_n}\{g^{-1}(T(x)) > g^{-1}(f(n))\} = g^{-1}(f(n))Pr_{\mu_n}\{T(x) > f(n)\}$$

Observing that

$$E_\mu(\frac{g^{-1}(T(x))}{|x|}) = \Sigma_{n=1}^{\infty}\frac{w(n)}{n}E_{\mu_n}(g^{-1}(T(x))) > \Sigma_{n=1}^{\infty}\frac{w(n)}{n}g^{-1}(f(n))Pr_{\mu_n}\{T(x) > f(n)\}$$

together with the assumptions in the hypothesis gives the desired result.

■

Corollary 3
Let $< S, \mu >$ be a regular global randomization with weight function w.
Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be a strictly increasing function,
and assume that $f(n)$ is a lower bound in probability on T w.r.t $< S_n, \mu_n >$.

(i) If w is regular then there exists $0 < \varepsilon < 1$ so that $T \notin AVB(f(n^\varepsilon))$.

(ii) If w is strongly regular then for every $0 < \varepsilon < 1$ we have $T \notin AVB(f(n^\varepsilon))$.

Proof:
For regular w let $c > 1$ be a constant so that for all $n \in \mathbb{N}^+$:

$$w(n) \geq n^{-c}.$$

Set $\varepsilon = \frac{1}{c}$ and $g(n) = f(n^\varepsilon)$. Then $g^{-1}(f(n)) = n^c$ and

$$\Sigma_{n=1}^{\infty} \frac{w(n)}{n} g^{-1}(f(n)) \geq \Sigma_{n=1}^{\infty} \frac{1}{n} = \infty.$$

By the Transfer Theorem for Lower Bounds we conclude that $T \notin AVB(g)$. For strongly regular weight functions let $0 < \varepsilon < 1$ and $g(n) = f(n^\varepsilon)$. Then the general term in the above sum evaluates to $w(n)n^{(\frac{1}{\varepsilon}-1)}$ and by the definition of strongly regular the sum diverges.

■

Probabilistic Bounds (cont.)

Proposition 4
Let $f, g : \mathbb{R}^+ \to \mathbb{R}^+$ *be two strictly increasing functions so that:*

$$lim_{n\to\infty}\frac{g^{-1}(f(n))}{n} = \infty$$

If f *is a lower bound with probability 1 on* T *w.r.t* $< S_n, \mu_n >$ *then there exists a global randomization* $< S, \mu >$ *which is compatible with* $< S_n, \mu_n >$ *such that* $T \notin AVB(< S, \mu >, g)$.

Proof:
Define a new function δ on $a \in \mathbb{R}^+$ by: $\delta(a) = \frac{g^{-1}(f(a))}{a}$. By previous theorem it is sufficient to construct a weight function w so that

$$\Sigma_{n=1}^{\infty} w(n)\delta(n) = \infty$$

We assume without loss of generality that δ is strictly positive and differentiable in $\mathbb{R}^+$. Denote the derivative by δ' and define w by:

$$w(n) = \delta'(n)\delta(n)^{-2}$$

To verify that w is indeed a weight function we have to show that its sum converges. This can be done by bounding the sum with an appropriate integral and using the assumptions as follows:

$$\Sigma_{n=1}^{\infty} w(n) \leq \int_1^{\infty} \delta'(a)\delta(a)^{-2} da = \frac{1}{\delta(1)}$$

On the other hand the sum over $w(n)\delta(n)$ diverges:

$$\Sigma_{n=1}^{\infty} w(n)\delta(n) \geq \int_1^{\infty} \delta'(a)\delta(a)^{-1} da = [ln\delta(a)]_1^{\infty} = \infty.$$

■

Flat distributions

Back to Outline

Flatness

Definition 5
(Gurevich, 1991): Let $< S, \mu >$ be a global randomization and let $< S_n, \mu_n >$ be a local randomization. Then:

- *μ is globally flat if there is a constant $k \in \mathbb{N}^+$ such that for every sufficiently large input $x \in S$, $\mu(x) \leq 2^{-|x|^{1/k}}$.*

- *The sequence μ_n is locally flat if there is a constant $k \in \mathbb{N}^+$ such that for sufficiently large n and input $x \in S_n$, $\mu_n(x) \leq 2^{-n^{1/k}}$.*

Theorem 6 *(Equivalence of global and local flatness): Let $< S, \mu >$ be a global randomization with weight function w and let $< S_n, \mu_n >$ be the implied local randomization:*

- *If the sequence μ_n is locally flat then μ is globally flat.*

- *If w is regular and μ is globally flat, then the sequence μ_n is locally flat.*

Distributions for SAT

Back to Outline

Sets of Clauses

Let V be any (finite or infinite countable) set of boolean variables.

- A literal ℓ over V is either v or $\neg v$ where $v \in V$.
- A clause C over V is a finite set of literals so that for no variable v both v and $\neg v$ occur in C. We denote the set of all clauses over V by $CL(V)$.
- CNF is the set of all finite subsets of $CL(V)$.
- CNFT is the set of all finite ordered tuples over $CL(V)$.
- A truth assignment is a mapping $z : V \to \{0, 1\}$.
  We define $z(\neg v) = 1 - z(v)$.
- An assignment z satisfies a clause C iff $z(\ell) = 1$ for at least one literal $\ell \in C$.
- An assignment satisfies a family Σ of clauses iff it satisfies every clause in Σ.

Symmetric Distributions of clauses

Back to Outline

Symmetric Distributions

Definition 7
(Negation-symmetry):
Let $\Pi = \{\pi : V \to \{v, \neg v : v \in V\} | \forall v \in V : \pi(v) \in \{v, \neg v\}\}$. *We extend* $\pi \in \Pi$ *to literals, clauses and* $\Sigma =< C_1, ..., C_n > \in CNFT$ *in the natural way.* $\pi(\Sigma)$ *is structurally the same as* Σ *but for some variables* $v \in V$ *the literals* v *and* $\neg v$ *are exchanged.*

- *If for* $\Sigma_1, \Sigma_2 \in CNFT$ *there exists a* $\pi \in \Pi$ *such that* $\Sigma_1 = \pi(\Sigma_2)$ *we say that* Σ_1 *and* Σ_2 *are negation-symmetric.*

- *For* $\Sigma \in CNFT$ *we define the symmetry-class of* Σ *by*

$$[\Sigma]_\Pi = \{\pi(\Sigma) : \pi \in \Pi\}.$$

Symmetric Distributions (cont.)

Definition 8
(Negation-symmetric invariant randomization):
Let $S \subseteq CNFT$ be the union of some symmetry-classes and let S_n be the CNFT-instances in S with n clauses. We say that a (local) randomization $< S_n, \mu_n >$ is negation-symmetry invariant if for all $\Sigma_1, \Sigma_2 \in S$ that are negation-symmetric we have $\mu_n(\Sigma_1) = \mu_n(\Sigma_2)$ where $n = |\Sigma|$.

Proposition 9
For $\Sigma \in CNFT$ let $var(\Sigma)$ denote the number of distinct variables appearing in Σ. Then $|[\Sigma]_\Pi| = 2^{var(\Sigma)}$.

Symmetric Distributions (cont.)

Theorem 10 (Flatness Theorem, JAM and Sharell, 1992): *Let $S \subseteq CNFT$ and let $< S_n, \mu_n >$ be a local randomization of S which is negation-symmetric invariant. If there is a constant $c \in \mathbb{N}^+$ such that for all $\Sigma \in S$ the number of clauses in Σ is bounded by $var(\Sigma)^c$ then $< S_n, \mu_n >$ is locally flat.*

Proof:
Let $n \in \mathbb{N}^+$ and $\Sigma \in S_n$. Since S is the union of some symmetry-classes $[\Sigma]_\Pi \subset S$ and since all instances in a symmetry class have the same number of clauses and so also the same size we have $[\Sigma]_\Pi \subset S_n$. Combining this with $var(\Sigma) \geq n^{1/c}$ we can get a bound on $\mu_n(\Sigma)$ as follows:

$$1 \geq \Sigma_{\Sigma'} \in [\Sigma]_\Pi \mu_n(\Sigma') = |[\Sigma]_\Pi| \mu_n(\Sigma) = 2^{var(\Sigma)} \mu_n(\Sigma) \geq 2^{n^{1/c}} \mu_n(\Sigma).$$

Therefore:

$$\mu_n(\Sigma) \leq 2^{-n^{1/c}}.$$

■

Symmetric Distributions (cont.)

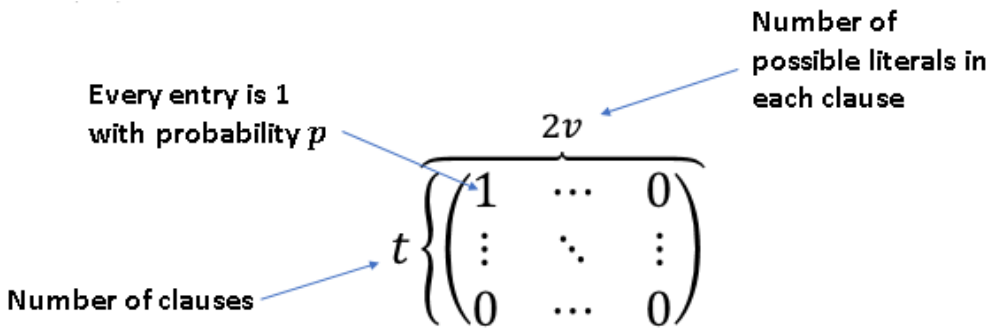
Remark: An interesting special case is $k - CNFT$ since if all clauses have exactly k literals for some constant k then the number of possible clauses is bounded by a polynomial in the number of variables. So any negation-symmetric-invariant randomization on a subset of $k-CNFT$ (where the same clause is not allowed to appear multiple times in an instance) is flat.

Fixed Density Distributions

Back to Outline

Fixed Density Distributions (FD)

- Let v be the number of variables, t the number of clauses and $p \in [0,1]$. We consider matrices M of 0's and 1's with $2v$ columns and t rows.
- For $i \in [v]$, $M(i,j) = 1$ iff the variable x_i occurs in clause j positively, and $M(v+i,j) = 1$ iff it occurs negatively.
- We denote by $l(M)$ the number of 1's in M.



Fixed Density Distributions (FD) – cont.

- Now, let $p : \mathbb{N}^+ \to [0, 1]$ and $t : \mathbb{N}^+ \to \mathbb{N}^+$ be functions such that $p(n) \leq \frac{1}{2}$ and t is non-decreasing.
- p is the probability that an entry is 1 as a function of the number of variables v.
- t is the number of clauses as a function of the number of variables v.
- Let $S(t)$ be the set of all Boolean $2v \times t(v)$ matrices M and $S(t)_n$ the set of matrices with $n = 2v \times t(v)$ ($S(t)_n$ is empty if there is no $v \in \mathbb{N}^+$ s.t. $n = 2vt(v)$).
- A Fixed-density (FD)-randomization [*] [†] is given by $< S(t)_n, \mu_n(t, p) >$ with

$$\mu_n(t, p)(M) = p(v)^{l(M)} \cdot (1 - p(v))^{n-l(M)}.$$

[*] Cynthia Brown, Allen Goldberg, and Paul Purdom. Average time analysis of simplified davis-putnam procedures. Information Processing Letters, 15(2), September 1982.

[†] Cynthia Brown and Paul Purdom. The pure literal rule and polynomial average time. SIAM Journal of computing, 14(4), November 1985.

Fixed Density Distributions (FD) – cont.

FD-distributions are clearly negation-symmetric. However, the number of clauses is not bounded by a polynomial in the number of variables, so this is not sufficient to make them flat.

Theorem 11 (JAM and Sharell, 1992): *A FD-distribution* $< S(t)_n, \mu_n(t,p) >$ *is locally flat iff there is a* $k \in \mathbb{N}^+$ *such that for all sufficiently large* $n \in \mathbb{N}^+$

$$-log_2(1-p_n) \geq n^{(\frac{1}{k}-1)}$$

where $p_n = p(v)$ *and* v *is the unique solution to* $2vt(v) = n$.

Proof:
$-log_2(1-p_n) \geq n^{(\frac{1}{k}-1)}$ iff $(1-p_n)^n \leq 2^{n^{-1/k}}$ iff for every input $M \in S(t)_n$

$$p_n^{l(M)}(1-p_n)^{n-l(M)} \leq 2^{n^{-1/k}}$$

(using that $p(n) \leq \frac{1}{2}$) iff (by definition of $\mu_n(t,p)$) for every input $M \in S(t)_n$

$$\mu_n(t,p)(M) \leq 2^{n^{-1/k}}$$

■

Fixed Size Distributions (FS)

- Let C be the set of clauses over the variables $x_i, i \in \mathbb{N}^+$ such that for no variable x both x and $\neg x$ occur in the same clause. For $v, k \in \mathbb{N}^+$ let $C(v,k) \subseteq C$ be the set of clauses with exactly k literals over the variables $x_1, ..., x_v$. Note that $C(v,k)$ has exactly $\binom{v}{k}2^k$ many elements.
- Let S be the set of size function defined by $|\Sigma|$=number of clauses in Σ.
- For $\alpha > 0$ set $S_n(\alpha, k) = C(\lfloor \alpha n \rfloor, k)^n$, that is all n-tuples of clauses over αn variables of length k. For $\Sigma \in S_n(\alpha, k)$ define $\mu_n(\Sigma_1) = |C(\alpha n, k)|^{-n}$.

Theorem 12 *For every choice of weight functions, the global version of these FS-distributions is flat.*

Proof:
An immediate consequence of the theorem about equivalence of global and local flatness.

■

Resolution of clauses

Back to Outline

Resolution

- Resolution is a particular widely used algorithm to solve SAT.
- If A, B are clauses and x is a variable such that $x \in A$ and $\neg x \in B$, then the clause $(A - \{x\}) \cup (B - \{\neg x\})$ is called a *resolvent* of A and B.
- Every truth assignment satisfying both A and B satisfies all their resolvents.

Resolution (cont.)

- Let Σ be a family of clauses and $C_1, C_2, ..., C_N$ be a sequence of clauses such that
  - each C_k belongs to Σ or is a resolvent for some C_i, C_j such that $i, j < k$.
  - C_N is the empty clause.
- induction on k shows that every truth assignment satisfying Σ must satisfy each C_k. since no truth assignment satisfies the empty clause C_N, it follows that Σ is unsatisfiable.
- The sequence $C_1, C_2, ..., C_N$ is called a *resolution proof* of unsatisfiability of Σ.

Complexity of Resolution

- Resolution complexity of an unsatisfiable family Σ of clauses is the smallest N such that there is a resolution proof $C_1, C_2, ..., C_N$ of unsatisfiability of Σ.
- Goldberg [*] showed that for certain FD-distributions SAT can be decided in expected polynomial time.
- Franco and Paul [†] showed that for certain FS-distributions a restricted form of resolution takes more than $2^{\sqrt[4]{n}}$ steps for almost all instances.
- Chvatal and Szemeredi [‡] have shown that for the same distribution resolution is exponential for almost all instances.

[*] A. Goldberg. Average case complexity of the satisfiability problem. In 4th Workshop on Automated Deduction, pages 1-6, 1979. Austin, TX.

[†] John Franco and Marvin Paul. Probabilistic analysis of the davisputman procedure for solving the satisfiability problem. Discrete Applied Mathematics, 5:77-87, 1983.

[‡] Vasek Chvatal and Endre Szemeredi. Many hard examples for resolution. Journal of the ACM, 35(4), October 1988.

Distributions for which SAT is Polynomial

The results here are based on a modified Davis-Putnam-Procedure to test satisfiability of clauses, which we call DPP^*, which is a special case of resolution.

Theorem 13 (Brown-Purdom) *Let $< S(t)_n, \mu_n(t,p) >$ be a FD-distribution and assume either*

i. $t(v) = \mathcal{O}(lnv)$ *or*

ii. $t(v)p(v) = v^{\mathcal{O}(p(v))}$ *or*

iii. $-ln(1-p(v)) = \frac{-ln(1-\mathcal{O}(\sqrt{\frac{lnv}{v}})}{t(v)})$

Then there is a polynomial $P(v)$ such that DPP^*, and hence resolution, has expected run-time $\mathcal{O}(P(v))$.

Distributions for which SAT is Polynomial (cont.)

Definition 14
(global FD-distribution) Let $< S(t), \mu >$ be a global randomization which ia compatible to a local FD-distribution. Then we call $< S(t), \mu >$ a global FD-distribution.

Theorem 15 *Let $< S, \mu_i >$ be a global FD-distribution compatible to a local FD-distribution subject to one of the conditions in Brown-Purdom theorem. Let $< S, \mu >$ be a global randomization that is a finite linear combination of the $< S, \mu_i >$. That is for some constants $a_1, a_2, .., a_m \in \mathbb{R}^+$:*

$$\mu(x) = \Sigma_{i=1}^{m} a_i \mu_i(x).$$

Then the distributional problem that consists of SAT and the randomization $< S, \mu >$ is in **AverP**.

Proof:
From the Transfer Theorem for Upper Bounds we derive that for each $1 \leq i \leq m$ SAT with $< S, \mu_i >$ is in **AverP**. It is easy to show that this is preserved under finite linear combinations.

■

Distributions for which Resolution is Exponential

Theorem 16 (Chvatal and Szemeredi) *Let* $< S_n(\alpha, k), \mu_n >$ *be a FS-randomization with* $k \geq 3$ *and* $\alpha \leq \frac{2^{-k}}{0.7}$. *Then there is a constant* $c \in \mathbb{R}^+$ *such that*

$$lim_{n\to\infty} Pr_{\mu_n}\{T_{res}(\Sigma) \geq 2^{cn}\} = 1,$$

where $T_{res}(\Sigma)$ *is the resolution complexity of* Σ.

The following result shows that the condition on α cannot be relaxed too much in the above theorem.

Theorem 17 (Franco) *Let* $< S_n(\alpha, k), \mu_n >$ *be a FS-randomization with* $\alpha > 1$. *Then there is an algorithm with runtime* T *so that for some* $c > 0$*:*

$$lim_{n\to\infty} Pr_{\mu_n}\{T(x) > n^c\} = 0.$$

Distributions for which Resolution is Exponential (cont.)

Theorem 18 (JAM and Sharell, 1992): *Let* $< S_n(\alpha, k), \mu_n >$ *be a local randomization with* $k \geq 3$ *and* $\alpha \leq \frac{2^{-k}}{0.7}$.

i. *Let* $< S, \mu >$ *be strongly regular and compatible to* $< S_n(\alpha, k), \mu_n >$. *Then for every* $0 < \varepsilon < 1$, $T_{res} \notin AVB(< S, \mu >, 2^{n^\varepsilon}$.

ii. *Let* $g : \mathbb{R}^+ \to \mathbb{R}^+$ *be a function so that* $g(n)/n \to 0$. *Then there exists a global randomization* $< S, \mu >$ *compatible to* $< S_n(\alpha, k), \mu_n >$ *so that:*

$$T_{res} \notin AVB(< S, \mu >, 2^{g(n)})$$

Proof:

i. Use the theorem by Chvatal and Szemeredi and corollary 3.

ii. Use the theorem by Chvatal and Szemeredi and proposition 4.

■

Average case completeness

Let (D, μ) be a flat randomized decision problem where D is solvable in deterministic exponential time.

Theorem: (Gurevich 1983)

If (D, μ) is average time hard for randomized **NP**, then

NEXPtime = EXPtime.

Conclusion: Unless **NEXPtime = EXPtime** flat distributions do not have maximal average complexity.

More recent work

Back to Outline

Related work: clauses/variables ratio

- **Mitchell et al. (1992)** * tested empirically the hardness of randomly generated 3-SAT formulas:

  m clauses are constructed uniformly and independently at random, each clause is obtained by sampling uniformly and independently 3 of n variables and negating each of them with probability $1/2$.

  Using the Davis-Putnam (DP) procedure, they found an *easy-hard-easy pattern, where the hardest formulas in terms of number of DP calls have a denisty (i.e. the clauses/variables ratio, m/n) of* ≈ 4.3, *near the point where* 50% *of the formulas are satisfiable. This suggests guidelines for constructing distributions of formulas for testing the average complexity of SAT solvers.*

*Mitchell, David, Bart Selman, and Hector Levesque. *"Hard and easy distributions of SAT problems." AAAI. Vol. 92. 1992.*

Related work: Average complexity and approximability

- **U. Feige (2002)** * studies the relationship betwee the clause/variable ratio and approximability of SAT and other NP-complete problems.

  He showed that for a particular distrubution the clause/variable ratio affects the approximability not only of SAT problems but also for many other NP-complete (NP-hard) problems.

* U. Feige, Relations between Average Case Complexity and Approximation Complexity

Related work: Phase transitions

- **Coarfa et al. (2000)** * investigated experimentally the average-case complexity of random 3-SAT formulas for fixed density and varying number of variables.

  They found a phase transition in which the complexity shifts from polynomial to exponential, where the value of density at which the phase transition occurs appears to be solver-dependent: the GRASP algorithm shifts from polynomial to exponential complexity near the density of 3.8, CPLEX algorithm shifts near density 3, while the transition of the CUDD algorithm is observed between densities of 0.1 and 0.5.

*Coarfa, C., Demopoulos, D. D., Aguirre, A. S. M., Subramanian, D., and Vardi, M. Y. *"Random 3-SAT: The plot thickens." International Conference on Principles and Practice of Constraint Programming. Springer, Berlin, Heidelberg, 2000.*

Related work: Walksat

- **Coja-Oghlan and Alan (2014)** * proved the following result:

  Let Φ be a uniformly distributed random k-SAT formula with n variables and m clauses, then the $Walksat$ algorithm finds a satisfying assignment of Φ in polynomial time with high probability if $m/n \leq \rho \cdot 2^k/k$ for a certain constant $\rho > 0$.

- In 2017 **Coja-Oghlan et al.** † proved that the $Walksat$ algorithm is ineffective with high probability if $m/n > c2^k ln^2 k/k$ where $c > 0$ is an absolute constant.

*Coja-Oghlan, Amin, and Alan Frieze. *"Analyzing Walksat on random formulas." SIAM Journal on Computing 43.4 1456-1485. 2014*

†Coja-Oghlan, Amin, Amir Haqshenas, and Samuel Hetterich.*"Walksat Stalls Well Below Satisfiability." SIAM Journal on Discrete Mathematics 31.2: 1160-1173. 2017*

Thank you for your attention
