Mansour's Conjecture is True for Random DNF Formulas

Adam Klivans
UT-Austin

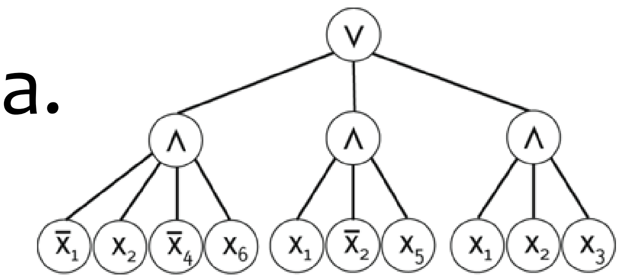
Homin K. Lee
UT-Austin

Andrew Wan
Columbia

[M94]

A Conjecture

Let f be a t-term DNF formula.



∃ a real $t^{O(\log 1/\varepsilon)}$-term poly p s.t.

$^{3}/_{16}x_1x_6 - {}^{1}/_{8}x_3x_7x_8$

$$\mathop{\mathrm{E}}_{x\in\{0,1\}^n}[(p(x)-f(x))^2] \le \varepsilon$$

[KM91]

Sparse Approximators

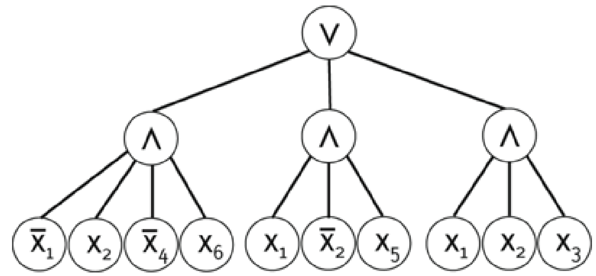
Thm: If $\forall f \in C$ has an s-sparse ε-approx p, then there is a uniform distribution MQ PAC learner for C that runs in time poly(n,s,ε^{-1}).

MC $\Rightarrow$ PAC-learning DNF

[J94]

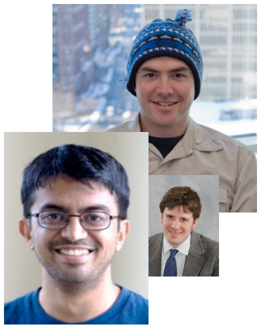
The Harmonic Sieve

A MQ PAC-learner for poly(n)-term DNF formulas over the uniform distribution.



- Didn’t prove MC.
- Used weak-approximator + boosting.

…and a decade passed.



[GKK08]

Sparse Approximators

Thm: If $\forall f \in C$ has an s-sparse ε-approx p, then there is a uniform distribution MQ *agnostic* learner for C that runs in time poly(n,s,ε^{-1}).

MC $\Rightarrow$ agnostic-learning DNF

Agnostic Learning

- f arbitrary Boolean function
- $\text{opt} = \min_{c \in C} \Pr_x[c(x) \neq f(x)]$

An <u>*agnostic learner*</u> is given MQ to f

w.h.p. outputs h s.t.

$$\Pr_x[h(x) \neq f(x)] \leq \text{opt} + \varepsilon$$

Previous Results

$$\mathop{E}_{x\in\{0,1\}^n}[(p(x)-f(x))^2] \leq \varepsilon$$

f a t-term DNF formula; $\exists\varepsilon$-approx p with:

- degree $O(\log(t/\varepsilon)^2)$ [LMN89]
- $t^{O(\log\log t \log(1/\varepsilon))}$ terms [M92]
- degree $O(\log(t/\varepsilon))$ [H01]

Our Results

$$\mathop{E}_{x\in\{0,1\}^n}[(p(x)-f(x))^2] \leq \varepsilon$$

$\exists$ ε-approx p with $t^{O(\log(1/\varepsilon))}$ terms for

- f a t-term *random* DNF formula
- f a t-term *read-k* DNF formula

(and [GK08] gives agnostic learners)

Outline

1. Intro
2. How we didn't prove it.
3. How we did prove it.
   a) Read-once DNF formulas
   b) Random DNF formulas
   c) Read-k DNF formulas
4. Pseudorandomness

How we didn't prove Mansour's Conjecture 1

Every f has a unique real polynomial representation with coeffs $\hat{f}(S)$ (the Fourier representation).

Analyze the large coeffs using Håstad's random restriction machinery [LMN89,M92,H01].

How we didn't prove Mansour's Conjecture 2

Entropy-Influence Conjecture: $E(f)=O(I(f))$

$\sum_S \hat{f}(S)^2 = 1$

$E(f) := \sum_S -\hat{f}(S)^2 \log(\hat{f}(S)^2)$

$I(f) := \sum_S |S| \hat{f}(S)^2$

$EI \Rightarrow MC$

Outline

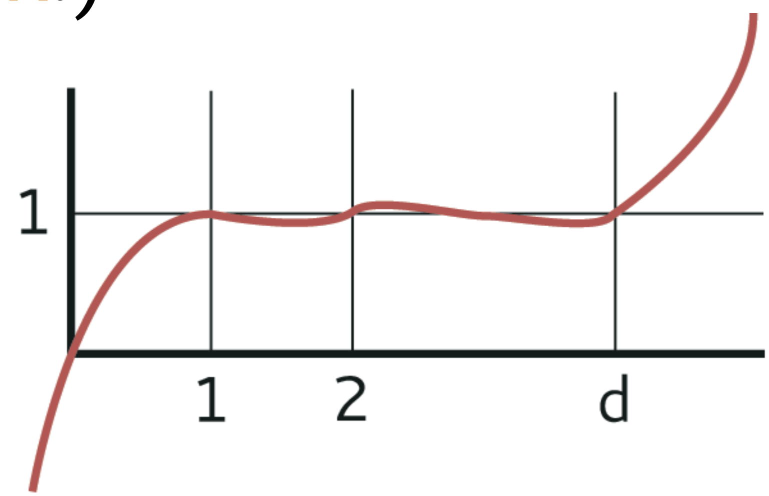
1. Intro
2. How we didn't prove it.
3. How we did prove it.
   a) Read-once DNF formulas
   b) Random DNF formulas
   c) Read-k DNF formulas
4. Pseudorandomness

Polynomial Interpolation

$f = T_1 \vee T_2 \vee \cdots \vee T_t$

Let $y_f(x) = T_1 + T_2 + \cdots + T_t$
(# of terms satisfied by x.)

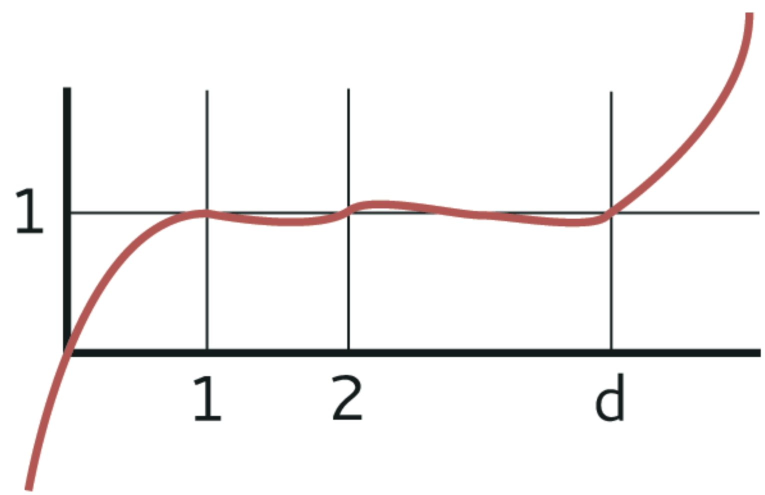
Interpolate the values of f on $\{x : y_f(x) \leq d\}$



The Polynomial

$$P_d(y) = ((-1)^{d+1}/d!)(y-1)(y-2)\cdots(y-d) + 1$$

- $P_d(0)=0$
- $P_d(y)=1,\ y=1\ldots d$
- $|P_d(y)|<\binom{y}{d},\ y>d$



The Polynomial

$$P_d(y) = ((-1)^{d+1}/d!)(y-1)(y-2)\cdots(y-d) + 1$$

- $P_d(y_f(x))$ has $t^{O(d)}$ terms.
- $P_d(y_f(x))=f(x)$ when x satisfies at most d terms.
- Need to show that x satisfies more terms with small probability.

Read-once DNF Formulas

Read-once: each var appears at most once

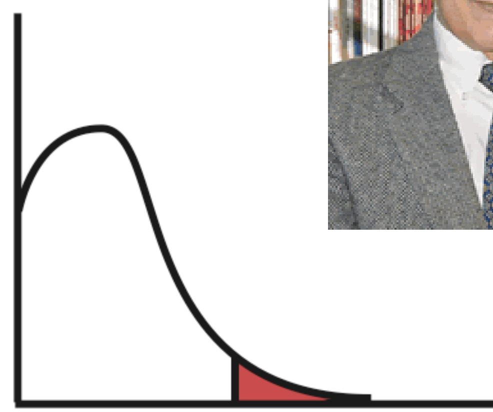
$x_1\bar{x}_5x_8 \vee x_2x_3\bar{x}_{18}x_{31} \vee x_4x_7$

$\Rightarrow$ terms are satisfied independently.

How do we show that sums of independent variables are concentrated in a narrow range?

Chernoff Bounds

- $T = \sum_{i=1}^{t} T_i$ (i.r.v.'s T_i=1 w.p. μ_i)
- $\mu = \sum_{i=1}^{t} \mu_i = E[T]$
- Can assume
  $\mu \leq \log(1/\varepsilon)$, or $f \approx 1$.



Chernoff : $\Pr[\, T = j \,] \leq (e\mu/j)^j$

MC is true for RO DNFs

$$\mathop{E}_{x \in \{0,1\}^n}[(p(x)-f(x))^2] \leq \varepsilon$$

$$\sum_{j=0}^{t} \Pr[y_f(x)=j]\,(P_d(y_f(x))-f(x))^2$$

$$\leq \sum_{j=d+1}^{t} (ed/j)^j \binom{j}{d}^2 \leq \varepsilon$$

for $d = \log(1/\varepsilon)$.

Outline

1. Intro
2. How we didn't prove it.
3. How we did prove it.
   a) Read-once DNF formulas
   b) Random DNF formulas
   c) Read-k DNF formulas
4. Pseudorandomness

MC is true for random DNFs

Our model: choose each term of a t-term DNF from the set of all terms of length $\log(t)$.

Show that w.h.p. random DNFs behave like RO DNFs using the method of bouded differences.

Outline

1. Intro
2. How we didn't prove it.
3. How we did prove it.
   a) Read-once DNF formulas
   b) Random DNF formulas
   c) Read-k DNF formulas
4. Pseudorandomness

Read-k DNF Formulas

Read-k: each var appears at most k times

$x_1\overline{x}_5x_8 \vee x_1x_2\overline{x}_3x_4 \vee x_5x_7$

Terms are no longer independent!

The Modified Construction

$f = T_1 \vee T_2 \vee \cdots \vee T_t$ (ordered from longest to shortest)

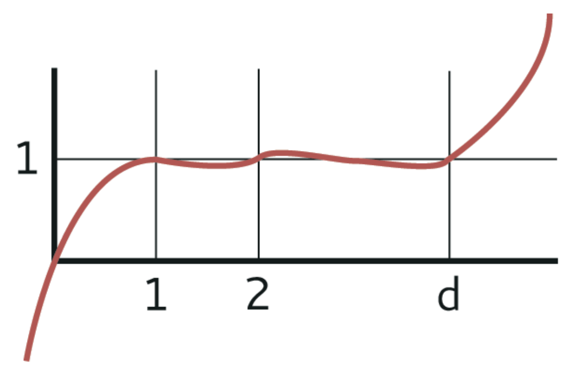
Let $z_f(x) = A_1 + A_2 + \cdots A_t$

$A_i = T_i \wedge (\wedge_{j \sim i, j \leq i} \neg T_j)$

(# of ind. terms sat. by x)

Interpolate the values

of f on $\{x : z_f(x) \leq d\}$



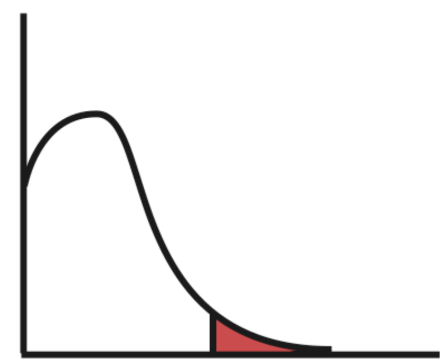
The Polynomial

$$P_d(z) = ((-1)^{d+1}/d!)(z-1)\,(z-2)\cdots(z-d) + 1$$

- $P_d(z_f(x))$ has $t^{O(kd)}$ terms.
- $P_d(z_f(x))=f(x)$ when x satisfies $\leq d$ ind. terms.
- Need to show that x satisfies more indep. terms with small probability.

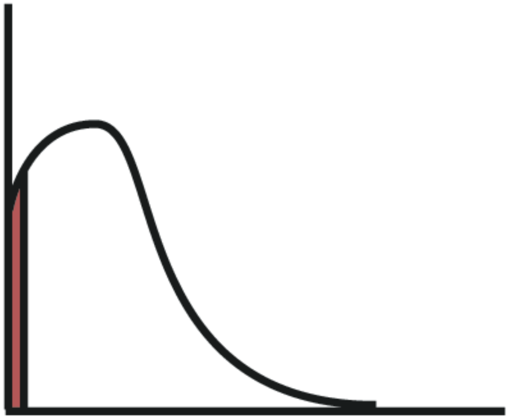
Concentration for Read-k

- T_i are r.v.'s 1 w.p. μ_i
- $\mu = \sum_{i=1}^{t} \mu_i$
- $A = \sum_{i=1}^{t} A_i \quad (A_i = T_i \wedge (\wedge_{j \sim i, j \leq i} \neg T_j))$
- $\Pr[\ A = j\ ] \leq \sum_{|S|=j} \Pi_{i \in S} T_i \leq (e\mu/j)^j$



Janson Bounds

- T_i are r.v.'s 1 w.p. μ_i
- $\mu = \sum_{i=1}^{t} \mu_i$
- $\Delta = \sum_{i \sim j} E[\, T_i T_j \,]$
- $\Pr[\, T=0 \,] \leq \exp(-\mu^2/\Delta)$



By Janson, can assume

$\mu \leq 16^k \log(1/\varepsilon)$, or $f \approx 1$.

Recap

$$\underset{x \in \{0,1\}^n}{E}[(p(x)-f(x))^2] \leq \varepsilon$$

$\exists\,\varepsilon$-approx p with $t^{O(\log(1/\varepsilon))}$ terms for

f a t-term *random* DNF formula w.h.p.

$\exists\,\varepsilon$-approx p with $t^{O(16^k \log(1/\varepsilon))}$ terms for

f a t-term *read-k* DNF formula

Outline

1. Intro
2. How we didn't prove it.
3. How we did prove it.
   a) Read-once DNF formulas
   b) Random DNF formulas
   c) Read-k DNF formulas
4. Pseudorandomness

Pseudorandomness

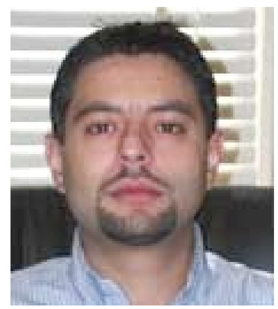
A distribution X ϕ-fools C if $\forall f \in C$

$$|E[f(X)] - E[f(U)]| \leq \phi$$

Seed length is # of random bits used by X.

PRGs against DNFs

Seed length for pseudorandom generators against t-term DNF formulas:



- $O(\log^4(tn/\phi))$ [LVW93]
- $O(\log(n)\log^2(t/\phi))$ [B07]
- $O(\log(n) + \log^2(t/\phi)\log\log(t/\phi))$ [DETT10]

The Sandwich Bound

If $\exists$ $s(\phi)$-sparse g & h s.t.

$\forall x,\ g(x) \leq f(x) \leq h(x)$

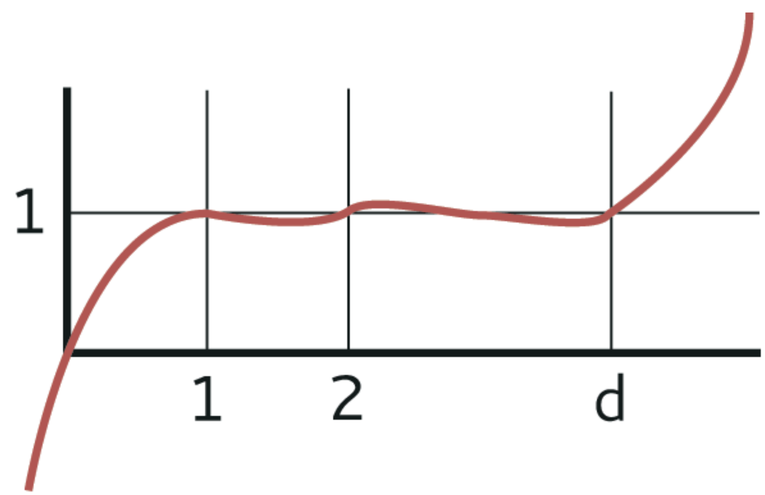
$E[h(x) - f(x)] \leq \phi$, $\quad$ $E[f(x) - g(x)] \leq \phi$

Then $\exists$ dist. that ϕ-fools f with seed length $O(\log n + \log s(\phi))$ [B07,DETT10]

The Polynomial

$$P_d(y) = ((-1)^{d+1}/d!)(y-1)(y-2)\cdots(y-d) + 1$$

- $P_d(0)=0$
- $P_d(y)=1,\ y=1\ldots d$
- $|P_d(y)|<\binom{y}{d}$
- $P_d(y)>1,\ y>d,\ d$ odd
- $P_d(y)<0,\ y>d,\ d$ even



PRGs against DNFs

- t-term random DNFs are fooled by PRGs w/ seed length
  $O(\log(n) + \log(t)\log(1/\phi))$ w.h.p.

- t-term read-k DNFs are fooled by PRGs w/ seed length
  $O(\log(n) + \log(t)16^k\log(1/\phi))$

([DETT10] showed $O(\log(n) + \log(t)\log(1/\phi))$ for RO DNFs)

Open Problems

- Prove Mansour's Conjecture for all t-term DNF formulas.

- Show PRGs against DNFs with seed length $O(\log(t)\log(1/\phi))$.

The End