Weighted Polynomial Approximations: Limits for Learning and Pseudorandomness

Mark Bun and Thomas Steinke

Harvard University

August 24, 2015

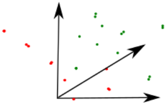
Derandomizing Concentration Inequalities

- Chernoff-Hoeffding Bound: Let $v \in \mathbb{R}^n$ and $U_n \in \{-1, 1\}^n$ be uniform. Then

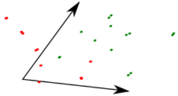
$$\Pr[|\langle U_n, v\rangle| \geq T\|v\|_2] \leq \exp(-\Omega(T^2)).$$

- Algorithmic applications, e.g. dimensionality reduction via Johnson-Lindenstrauss

$$\begin{bmatrix} -1 & 1 & 1 & \dots & -1 \\ 1 & -1 & 1 & \dots & 1 \end{bmatrix} \times$$



$$=$$



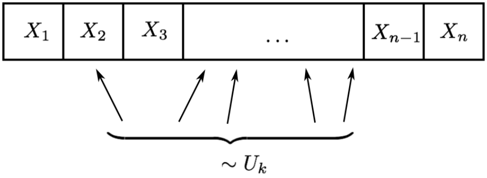
- Motivating question: What pseudorandom X suffice in place of U_n?

This Talk

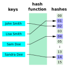
- Main result: Lower bound for derandomizing Chernoff via k-wise independence
- (Non-constructive) proof by way of polynomial approximations
- Similar ideas give lower bounds for agnostically learning halfpsaces

Why k-Wise Independence?

$X \in \{-1,1\}^n$ is <u>k-wise independent</u> if every subset of k variables is uniform:

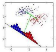


- Simple and pervasive notion of pseudorandomness

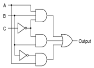


Hashing

Streaming



Dimensionality Reduction



Circuit Complexity

- Naturally gives rise to the Chernoff bound

Chernoff Bound from k-Wise Independence [Schmidt-Siegel-Srinivasan93]

Proof of Chernoff by moment bounds: Let $v \in \mathbb{R}^n$ be a unit vector

$$\begin{aligned}
\Pr[|\langle U_n, v\rangle| \geq T] &= \Pr[(\langle U_n, v\rangle)^k \geq T^k] \\
&\leq \frac{E[\langle U_n, v\rangle^k]}{T^k} && \text{(Markov's Inequality)} \\
&\leq \frac{k^{k/2}}{T^k} && \text{(Khintchine-Kahane)}
\end{aligned}$$

- For a tail bound of δ and $T = \sqrt{\log(1/\delta)}$, it suffices to take $k = O(\log(1/\delta))$
- Can replace U_n with <u>any</u> k-wise independent X, since $E[\langle X, v\rangle^k] = E[\langle U_n, v\rangle^k]$

Pseudorandom Generators for Chernoff

A k-wise independent X requires seed length $O(k \log n)$ [ABI86]
$\implies$ PRG with seed length $O(\log n \cdot \log(1/\delta)) = O(\log^2 n)$
Think of $\delta = 1/\operatorname{poly}(n)$

Question: Is the [SSS93] analysis optimal? (Or can k be reduced?)

Other PRGs for Chernoff:

Construction	Seed length
Probabilistic method	$O(\log n + \log(1/\delta)) = O(\log n)$
Small-bias spaces [NN90]	$O(\log n \cdot \log(1/\delta)) = O(\log^2 n)$
PRG for small space [Nis92,INW94]	$O(\log n \cdot \log(n/\delta)) = O(\log^2 n)$
PRG for Fourier shapes [GKM15]	$\tilde{O}(\log n + \log(1/\delta)) = \tilde{O}(\log n)$

Main Result

Theorem (Main)

Let $\delta \leq 1/\operatorname{poly}(n)$ and $T = \Theta(\sqrt{\log(1/\delta)})$. For $k = \Omega(\log(1/\delta))$, there exists a k-wise independent X for which

$$\Pr[|X_1 + \cdots + X_n| \geq T\sqrt{n}] > \delta.$$

- Matches upper bound of [SSS93]
- Previous lower bound [SSS93] of

$$k \geq \Omega\left(\frac{\log(1/\delta)}{\log n}\right)$$

is constant for $\delta = 1/\operatorname{poly}(n)$

Proof Overview

1. Dual formulation of problem [Bazzi07, DGJSV09]:

   Chernoff bound via k-wise independence

   $\Leftrightarrow$

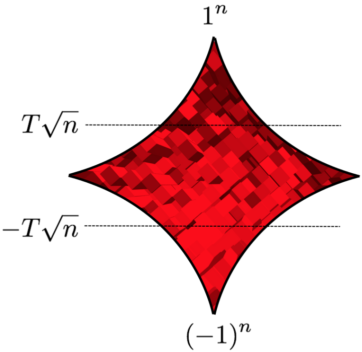
   Threshold function well-approximated by a degree-k polynomial

2. Lower bound k using real approximation theory

LP and dual formulations of derandomizing Chernoff by k-wise independence

Primal Formulation [Bazzi07]

"What is the worst tail bound given by a k-wise independent X?"



Primal Formulation [Bazzi07]

"What is the worst tail bound given by a k-wise independent X?"
Let $\psi(x) = \Pr[X = x]$

$$\max_{\psi} \sum_{x \in \{-1,1\}^n} \psi(x) \cdot \mathbb{1}(|x_1 + \cdots + x_n| \geq T\sqrt{n})$$

$$\text{s.t.} \sum_{x \in \{-1,1\}^n} \psi(x) \cdot \chi_S(x) = 0 \qquad \text{for all } |S| \leq k$$

$$\sum_{x \in \{-1,1\}^n} \psi(x) = 1$$

$$0 \leq \psi(x) \leq 1 \qquad \text{for all } x \in \{-1,1\}^n$$

Dual Formulation [Bazzi07]

"What is the smallest L_1-norm of any degree-k upper sandwiching polynomial?"

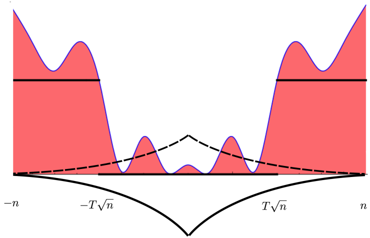
$$\min_p \; 2^{-n} \sum_{x \in \{-1,1\}^n} p(x)$$

$$\text{s.t. } \operatorname{deg}(p) \le k$$

$$p(x) \ge \mathbb{1}(|x_1 + \cdots + x_n| \ge T\sqrt{n}) \quad \text{for all } x \in \{-1,1\}^n$$

Dual Formulation [Bazzi07]

"What is the smallest L_1-norm of any degree-k upper sandwiching polynomial?"



Theorem: There exists a k-wise independent X with a tail bound worse than δ

$\Leftrightarrow$

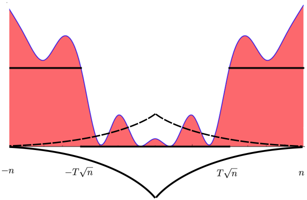
Every upper sandwich with L_1-norm at most δ has degree greater than k

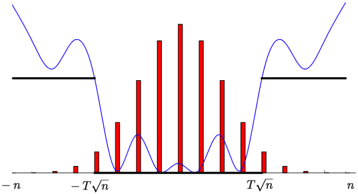
Goal: Lower bound the degree of any upper sandwich with low L_1-norm

Simplifying the Problem

Symmetrization [MinskyPapert69]

Any upper sandwich p can be turned into a univariate p^{sym}:





1. $\deg(p^{\mathsf{sym}}) \leq \deg(p)$
2. p^{sym} is an upper sandwich for univariate threshold function
3. L_1-norm of $p = L_1$-norm of p^{sym} under binomial distribution

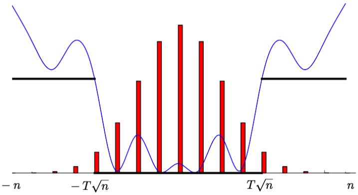
Simplifying the Problem

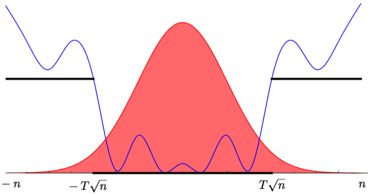
Approximation by a Gaussian

L_1-norm of p^{sym} under binomial distribution

$$\approx$$

L_1-norm of p^{sym} under Gaussian



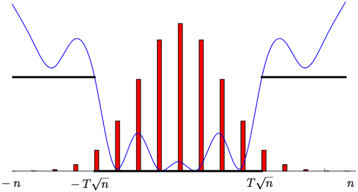


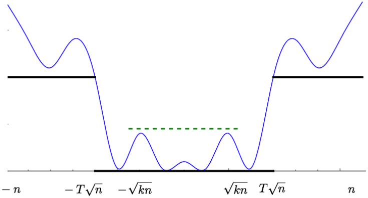
Technical step: p^{sym} bounded at integers $\Rightarrow$ p^{sym} bounded on reals [EhlichZeller64]

Simplifying the Problem

Approximation by a Gaussian

$$\deg(p^{\mathsf{sym}}) = k \text{ and } L_1(p^{\mathsf{sym}}) \le \delta \text{ (under binomial distribution)}$$
$$\Longrightarrow$$
$$p^{\mathsf{sym}}(t) \le \delta n \text{ for all } t \in [-\sqrt{kn}, \sqrt{kn}]$$

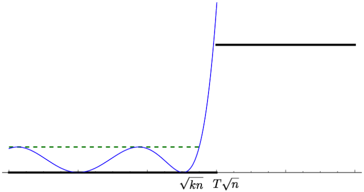




Technical step: p^{sym} bounded at integers $\Rightarrow$ p^{sym} bounded on reals [EhlichZeller64]

The Final Step

- Let p be a degree-k upper sandwich for $\mathbb{1}(|x_1 + \cdots + x_n| \geq T\sqrt{n})$ with $L_1(p) \leq \delta \leq 1/n^4$
- $\implies$ There exists univariate p^{sym} with
  1. $\deg(p^{\mathsf{sym}}) \leq k$,
  2. $p^{\mathsf{sym}}(\pm T\sqrt{n}) \geq 1$, and
  3. $0 \leq p^{\mathsf{sym}}(t) \leq \delta n$ for all $t \in [-\sqrt{kn}, \sqrt{kn}]$



- $p^{\mathsf{sym}}(T\sqrt{n}) \leq \delta n \cdot \mathrm{Chebyshev}_k\left(\frac{T\sqrt{n}}{\sqrt{kn}}\right) \leq \delta n \cdot \left(\frac{2T}{\sqrt{k}}\right)^k$
- Conclusion: $k \geq \Omega(\log(1/\delta))$

Recap

Theorem (Main)

Let $\delta \leq 1/\operatorname{poly}(n)$ and $T = \Theta(\sqrt{\log(1/\delta)})$. For $k = \Omega(\log(1/\delta))$, there exists a k-wise independent X for which

$$\Pr[|X_1 + \cdots + X_n| \geq T\sqrt{n}] > \delta.$$

1. Dual formulation of problem [Bazzi07, DGJSV09]:

   Chernoff bound via k-wise independence

   $\Leftrightarrow$

   Threshold function well-approximated by a degree-k polynomial

2. By real approximation theory, $k \geq \log(1/\delta)$

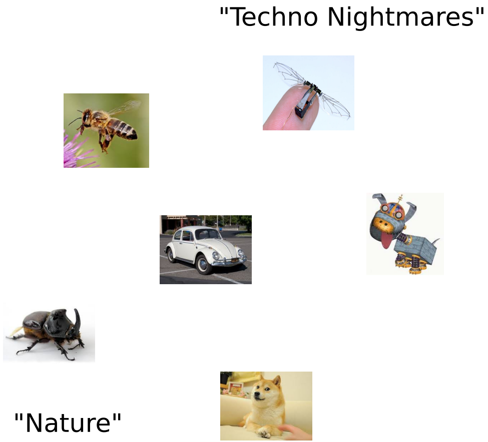
Open Questions

- Explicit bad k-wise independent distribution?
  (cf. dual witnesses for approximate degree lower bounds
  [Špalek08, B.-Thaler13])
- Seed length needed for small-bias spaces? XORs of small-bias spaces?

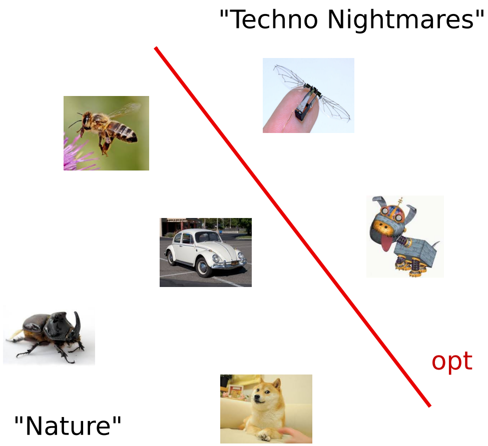
Thank you!

Agnostically Learning Halfspaces

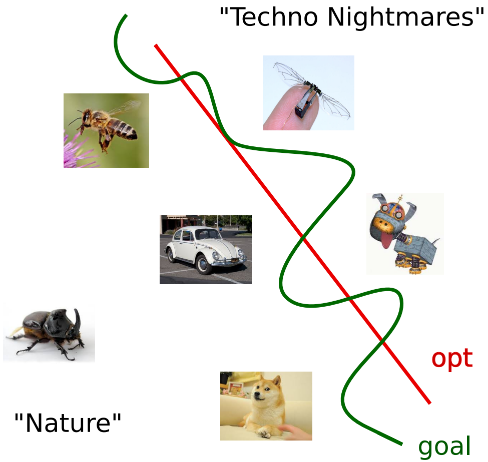
Learning with Noise: The Agnostic Model



Learning with Noise: The Agnostic Model



Learning with Noise: The Agnostic Model



Agnostically Learning Halfspaces

- [KalaiKlivansMansourServedio05] gave an efficient algorithm under distributional assumptions
- E.g., if distribution on examples in $\mathbb{R}^n$ is log-concave, then any halfspace $\approx$ a low-degree polynomial
- Question: Can the log-concavity assumption be relaxed?

- This work: For mildly log-convex distributions, there exist halfspaces that cannot be approximated by polynomials of <u>any degree</u>
- Draws on L_1-approximation theory [NevaiTotik86,87]