Analysis of Boolean Functions **(CMU 18-859S, Spring 2007)**

Lecture 29: Open Problems

May 3, 2007

Lecturer: Ryan O'Donnell *Scribe: Ryan O'Donnell*

1 Miscellaneous problems

Small total influence implies a large coefficient: Prove or disprove: For every $f : \{-1,1\}^n \to \{-1,1\}$ there exists some S such that $|\hat{f}(S)| \geq 2^{-O(\mathbb{I}(f))}$. One might also try to add the condition that the S satisfies $|S| \leq O(\mathbb{I}(f))$. A lower bound of $2^{-O(\mathbb{I}(f)^2)}$ follows from Friedgut's Theorem, and with this one can also get that $|S| \leq O(\mathbb{I}(f))$.

The result definitely holds for *monotone* functions: From the proof of Friegut/KKL one can show that if $\mathrm{Inf}_i(f) \leq \tau$ for all i, then $\mathbb{I}(f) \geq \Omega(\mathbf{Var}[f]\log(1/\tau))$ (this is usually credited to Talagrand [Tal94]). Thus either $\hat{f}(\emptyset) \geq \Omega(1)$, or there exists i such that $\mathrm{Inf}_i(f) \geq 2^{-O(\mathbb{I}(f))}$. But for monotone functions, $\mathrm{Inf}_i(f) = \hat{f}(\{i\})$.

Bounding level k weight by level 1 weight: Let $f : \{-1,1\}^n \to \{-1,1\}$, and let $\mathbb{III}(f)$ denote $\sum_{i=1}^n \mathrm{Inf}_i(f)^2$. Recall also that we write $W_k(f)$ for $\sum_{|S|=k}\hat{f}(S)^2$; note that $\mathbb{III}(f) = W_1(f)$ if f is monotone. Talagrand [Tal96] showed that for any $f : \{-1,1\}^n \to \{-1,1\}$, $W_2(f) \leq O(\mathbb{III}(f)\log(1/\mathbb{III}(f)))$. Benjamini, Kalai, and Schramm [BKS99] generalized this to show that for each $k \geq 2$, $W_k(f) \leq C_k \cdot \mathbb{III}(f)\log^{k-1}(1/\mathbb{III}(f))$, for some constant C_k. Unpublished work of Kindler shows that in fact one can make the C_k's *smaller* as k increases, with a bound $C_k \leq O(1/k)$. A conjecture is that one can get $C_k \leq O(1/k!)$; if so, this would be tight by considering the Tribes function.

2 Decision trees

Decision trees and influences for real-valued functions [OSSS05]: Recall we proved that for $f : \{-1,1\}^n \to \{-1,1\}$, $\mathbf{Var}[f] \leq \sum_{i=1}^n \delta_i(f)\mathrm{Inf}_i(f)$. The question is to what extent this is true for functions $f : \{-1,1\}^n \to \mathbb{R}$; in particular, is it true that $\mathbf{Var}[f] \leq C \cdot \sum_{i=1}^n \delta_i(f)\mathrm{Inf}_i(f)$ for some universal constant C? By an explicit example it is known that C can't be 1, but the best example only gives a lower bound like $C \geq 1.1$.

3 DNFs

Total influence of DNF: As came up on Problem 1 of Homework #3: If f is computable by a DNF of width w, must it hold that $\mathbb{I}(f) \leq w$? This would be sharp, by Parity, and proving a $2w$

upper-bound is easy.

Fourier concentration for DNF: Let f be computable by a poly-sized DNF. Is f ϵ-concentrated on a set of Fourier coefficients of cardinality at most $n^{C(\epsilon)}$ (i.e., polynomial for constant ϵ)? This question is not actually very interesting for learning theory, since the immediate learning consequence is already superseded by Jackson's algorithm. Also, it's not clear whether or not Tribes already rules out this conjecture.

4 LTFs

Noise sensitivity of intersections of halfspaces [KOS04]: Peres's Theorem is that if $f : \{-1,1\}^n \to \{-1,1\}$ is an LTF (halfspace), then $\mathbb{NS}_\epsilon(f) \leq O(\sqrt{\epsilon})$. By the union bound, this implies that if f is the intersection (AND) of k LTFs, then $\mathbb{NS}_\epsilon(f) \leq O(k\sqrt{\epsilon})$. It is conjectured that the following better upper bound holds: $\mathbb{NS}_\epsilon(f) \leq O(\sqrt{\log k}\sqrt{\epsilon})$. This would be tight, by considering k symmetric LTFs with bias $1 - 1/k$ on disjoint sets of variables.. The bound is known to hold if the k LTFs are on disjoint sets of variables.

Most noise sensitive LTF: Let n be odd and fix $0 < \epsilon < 1/2$. Show that the LTF on n bits with highest noise sensitivity at ϵ is Majority. (Peres's Theorem implies this is true up to a constant factor.)

Approximate Chow Parameters: The following problem is attributed to P. Goldberg [Gol06] (see also [Ser06]). Let $f : \{-1,1\}^n \to \{-1,1\}$ be an LTF. It is known [Cho61] and not too hard to show that f's "Chow Parameters" $\hat{f}(\emptyset), \hat{f}(\{1\}), \ldots, \hat{f}(\{n\})$ uniquely determine f among the class of all boolean-valued functions. Now suppose $g : \{-1,1\}^n \to \{-1,1\}$ is another LTF satisfying

$$\sum_{|S|\leq 1} (\hat{f}(S) - \hat{g}(S))^2 \leq \epsilon.$$

Must g be $o_{\epsilon\to 0}(1)$-close to f?

5 Learning

Learning monotone DNF: Can poly-size monotone DNF be PAC-learned under the uniform distribution in polynomial time? (Feel free to assume that the accuracy parameter, ϵ, is a constant.) This is not inherently a Fourier analysis problem, but it's such a big open problem in PAC-learning that it's worth mentioning; also, it's likely that Fourier analysis would play a big role in any solution.

Learning juntas: In addition to the problems for which Avrim Blum will give you prizes, one may ask: Can k-juntas over $\{1,2,3\}^n$ be learned in time $n^{(1-\Omega(1))k}$? How about juntas under the p-biased product distribution, $p \neq 1/2$?

6 Testing

The best 3-bit dictator vs. quasirandom test with perfect completeness: Suppose we want a 3-query test for functions $f : \{-1,1\}^n \to \{-1,1\}$ that accepts all dictators with probability 1 and accepts all $(o(1), o(1))$-quasirandom functions with probability at most $s + o(1)$. How small can s be? Work of Khot and Saket [KS06] implies that s can be as small as $20/27$. A conjecture is that the smallest possible s is $5/8$, but this is not known to be an upper or lower bound.

7 Noise sensitivity

Most noise sensitive code [Kal00]: Let $f : \{-1,1\}^n \to \{0,1\}$ have $\mathbf{E}[f] = 2^{-(1-R)n}$ for some constant $0 < R < 1$. What is the most that $\mathbb{NS}_\epsilon(f)$ can be? Is it achieved, roughly, by a random function of that density? According to Kalai, this is connected to the question as to whether the Gilbert-Varshamov bound from coding theory is optimal.

Cosmic coin flipping [MO05]: Fix $k \geq 2$ and $0 < \epsilon < 1/2$. Suppose $\boldsymbol{x} \in \{-1,1\}^n$ is chosen at random and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ are each formed by letting $\boldsymbol{y}_i = N_\epsilon(\boldsymbol{x})$, independently across i's. We wish to pick an odd function $f : \{-1,1\}^n \to \{-1,1\}$ in an effort to maximize $\mathbf{Pr}[f(\boldsymbol{y}_1) = f(\boldsymbol{y}_2) = \cdots = f(\boldsymbol{y}_k)]$. (Imagine k players are independently measuring a "cosmic" source of random bits, with each experiencing ϵ noise. They wish to try to use their noisy strings to agree on a random bit.) Prove (or disprove): there exists some $r = r(k, \epsilon)$ such that Maj_r is maximizing.

Plurality is Stablest? [MOO05]: Fix $0 < \rho < 1$. Consider the class of functions $f : \{1,2,3\}^n \to \{1,2,3\}$ which are "balanced" ($\mathbf{Pr}[f = c] = 1/3$ for each $c = 1,2,3$) and have small influences ($\mathbf{E}_{\boldsymbol{x}}[\mathbf{Var}_{\boldsymbol{x}_i}[f]] < o(1)$ for all i), where we think of the input domain $\{1,2,3\}^n$ as having the uniform distribution. Let $\boldsymbol{x} \in \{1,2,3\}^n$ be uniformly random and form $\boldsymbol{y}$ by holding each coordinate of $\boldsymbol{x}$ fixed with probability ρ and rerandomizing it with probability $1 - \rho$. Is it true that

$$\mathbb{S}_\rho(f) := \mathop{\mathbf{Pr}}_{\boldsymbol{x},\boldsymbol{y}}[f(\boldsymbol{x}) = f(\boldsymbol{y})] \leq \lim_{n\to\infty} \mathbb{S}_\rho(\text{Plurality}_n) + o(1)?$$

8 Hypercontractivity

Improved Markov for smoothed functions [Tal89]: Michel Talagrand will give you \$1000 if you solve this problem: Fix $0 < \rho < 1$ (think of ρ close to 1). Let $f : \{-1,1\}^n \to [0,1]$ and write $\mu = \mathbf{E}[f]$. Note that $\mathbf{E}[T_\rho f] = \mu$ as well. Clearly, Markov's inequality implies that for large t, $\mathbf{Pr}[(T_\rho f)(\boldsymbol{x}) \geq t\mu] \leq 1/t$. However, since T_ρ "smooths" f out, one might hope for something better. Talagrand conjectures asks for a proof (or disproof) of the better upper bound $O(1/(t\sqrt{\log t})$.

9 Circuit complexity

Small total influence implies small approximating circuits for monotone functions [BKS99]: Linial, Mansour, and Nisan [LMN93] implies that if $f : \{-1,1\}^n \to \{-1,1\}$ has a circuit of depth d and size s, then

$$\mathbb{I}(f) \leq O(\log^d(s)).$$

Boppana [Bop97] improved the exponent to $d-1$, which is sharp (by considering Parity). It's possible that the following "reverse" result holds, approximately, for monotone functions:

Let $f : \{-1,1\}^n \to \{-1,1\}$ be monotone, and let $\epsilon > 0$. Then there is a circuit ϕ which computes f correctly on a $1-\epsilon$ fraction of inputs and has size s and depth d satisfying

$$\mathbb{I}(f) \leq O(\log^d(s)).$$

Note that it is impossible to improve the exponent here to $d-1$, by a recent result [OW07].

10 Threshold phenomena, random graphs, percolation

Total influence lower bounds [Kal00]: Find "general conditions" on functions $f : \{-1,1\}^n \to \{-1,1\}$ that imply $\mathbb{I}(f) \geq n^{\Omega(1)}$. The motivation here is showing that monotone functions have *very* sharp thresholds. Bourgain and Kalai [BK97] have results that can show $\mathbb{I}(f) \geq \mathrm{polylog}(n)$ if f has enough symmetries. The only other method I know is the inequality relating influences and decision tree complexity from Lecture 26.

Influence versus Fourier entropy [FK96]: This is a particular case of the above problem. It would also imply the first problem listed in the Miscellaneous section. Let $f : \{-1,1\}^n \to \{-1,1\}$. Show that

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 \log(1/\hat{f}(S)^2) \leq O(\mathbb{I}(f)).$$

This seems very similar to the Log-Sobolev inequality proven in Homework #4, but it's not clear if they are actually related (in particular, this conjecture clearly needs that f is boolean-valued).

Thresholds for subgraph containment: Let H be any fixed graph on up to n vertices, and let f be the monotone graph property (in the $G(n,p)$ model) of containing a copy of H. For which graph H is the threshold sharpest? If one could show $\mathbb{I}^{(p_c)}(f) \leq O(\sqrt{v})$ for any subgraph containment property f (with p_c the appropriate critical probability), then one could use the results of Lecture 26 to recover the result of [Grö92], showing $R(f) \geq \Omega(v^{3/2})$ for subgraph containment properties.

Distance variance of first passage percolation: Consider the graph on $\mathbb{Z}^2$ where each vertex is connected to its 4 neighbors at distance 1. Choose each edge to have "length" either 1 or 2, independently and with probability 1/2 each. Now let $\boldsymbol{f}$ denote shortest-path distance from $(0,0)$ to (v,v), where $v \in \mathbb{N}$ is thought of as large. Using the result of Talagrand [Tal94] (mentioned in the first problem of the Miscellaneous section), [IB03] have shown that $\mathbf{Var}[\boldsymbol{f}] \leq O(v/\log v)$. The goal is to prove that $\mathbf{Var}[\boldsymbol{f}] = \Theta(v^{2/3})$, which the statistical physicists "know" to be the right answer.

Percolation on the grid: The scenario here is similar to the previous one. Consider an $(m+1) \times m$ subgrid. Let each of the $n = 2m^2 - 1$ edges be present or absent with probability $1/2$, and let $f : \{-1,1\}^n \to \{-1,1\}$ be the indicator of a "crossing"; i.e., a path from the left side to the right side. (A cute exercise: show $\mathbf{E}[f] = 0$.) One problem is to prove the following conjecture made by physicists: $\mathbb{I}(f) = \Theta(n^{3/8})$. Another is to prove the following conjecture from [BKS99]: For every $\epsilon > 0$, for sufficiently large m, the following holds:

$$\mathop{\mathbf{Pr}}_{\text{horizontal edges}}\left[\left|\mathop{\mathbf{Pr}}_{\text{vertical edges}}[\text{crossing} \mid \text{horizontal edges}] - 1/2\right| \geq \epsilon\right] \leq \epsilon.$$

if one chooses just the

11 Arithmetic Combinatorics

Triangle removal in $\mathbb{F}_2^n$ [Gre04b]: Suppose $f : \mathbb{F}_2^n \to \{0,1\}$ is ϵ-far from being triangle-free (meaning that there are no x, y, z such that $x + y + z = 0$ and $f(x) = f(y) = f(z) = 1$). Prove or disprove: the no-triangles test (pick $\boldsymbol{x}, \boldsymbol{y}$ at random and check that $f(\boldsymbol{x}), f(\boldsymbol{y}), f(\boldsymbol{x}+\boldsymbol{y})$ are not all 1) rejects with probability at least $\mathrm{poly}(\epsilon)$.

Cosets in sumsets (see [Gre04a]): Let $A \subseteq \mathbb{F}_2^n$ have density at least $1/4$. Green has shown that the set $A + A := \{a + b : a, b \in A\}$ must contain coset of codimension at least $\Omega(n)$. On the other hand, Ruzsa has shown that there exists an A of density at least $1/4$ (specifically, the set of vectors with at least $n/2 + \sqrt{n}/2$ 1's) such that $A + A$ doesn't contain any coset of codimension at most $\sqrt{n}$. Narrow this gap.

Polynomial Freiman-Ruzsa Conjecture: This important open problem in arithmetic combinatorics is attributed to Marton by Ruzsa (see, e.g., [Gre04a]. It has many equivalent formulations, including the following: Let $f : \mathbb{F}_2^n \to \mathbb{F}_2^m$ (not just $\to \mathbb{F}_2$) satisfy $\mathbf{Pr}_{\boldsymbol{x},\boldsymbol{y}}[f(\boldsymbol{x}) + f(\boldsymbol{y}) = f(\boldsymbol{x}+\boldsymbol{y})] \geq \epsilon$. Then there is some affine linear function $g : \mathbb{F}_2^n \to \mathbb{F}_2^m$ such that $\mathbf{Pr}[f(\boldsymbol{x}) = g(\boldsymbol{x})] \geq \mathrm{poly}(\epsilon)$.

Singularity probability for random matrices: Let M be a random $n \times n$ matrix, where each entry is an independent random ± 1 bit. Let P_n denote the probability that M has determinant 0. Clearly this will happen if any two rows are the same (up to sign) or any two columns are the same (up to sign). This gives a lower bound:

$$P_n \geq (1 - o(1)) 2n^2 \cdot 2^{-n}.$$

It is conjectured that this bound is correct up to a $1 + o(1)$ factor. A breakthrough result [JK95] gave the upper bound $P_n \leq .999^n$, and the best current result [TV07] gets this down to $(3/4 + o(1))^n$, using Fourier analysis and arithmetic combinatorics. That paper includes some discussion of how one might try to improve the result to $(1/2 + o(1))^n$.

References

[BK97] J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group symmetries. *GAFA*, (7):438–461, 1997.

[BKS99] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[Bop97] R. Boppana. The average sensitivity of bounded-depth circuits. *Inf. Process. Lett.*, 63(5):257–261, 1997.

[Cho61] C. Chow. On the characterization of threshold functions. In *Proc. Symp. Switching Circuit Theory and Logical Design*, pages 34–38, 1961.

[FK96] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. AMS*, (124):2993–3002, 1996.

[Gol06] P. Goldberg. A bound on the precision required to estimate a boolean perceptron from its average satisfying assignment. *SICOMP*, pages 328–343, 2006.

[Gre04a] B. Green. Finite field models in arithmetic combinatorics, 2004. On the arXiv.

[Gre04b] B. Green. A Szemerédi-type regularity lemma in abelian groups with applications, 2004. On the arXiv.

[Grö92] D. Gröger. On the randomized complexity of monotone graph properties. *Acta Cybernetica*, 10:119–127, 1992.

[IB03] O. Schramm I. Benjamini, G. Kalai. First passage percolation has sublinear distance variance. *Ann. Probab.*, 31(4), 2003.

[JK95] E. Szemerédi J. Kahn, J. Komlós. On the probability that a random ± 1 matrix is singular. *J. AMS*, 8:223–240, 1995.

[Kal00] G. Kalai. Combinatorics with a geometric flavor: Some examples. http://citeseer.ist.psu.edu/332522.html, 2000.

[KOS04] A. Klivans, R. O'Donnell, and R. Servedio. Learning intersections and thresholds of halfspaces. *JCSS*, 68(4), 2004.

[KS06] S. Khot and R. Saket. A 3-query non-adaptive PCP with perfect completeness. In *CCC*, 2006.

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

[MO05] E. Mossel and R. O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *RS&A*, 26(4):418–436, 2005.

[MOO05] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *FOCS*, 2005.

[OSSS05] R. O'Donnell, M. Saks, O. Schramm, and R. Servedio. Every decision tree has an influential variable. In *FOCS*, 2005.

[OW07] R. O'Donnell and K. Wimmer. Approximation by DNF: examples and counterexamples. In *ICALP*, 2007.

[Ser06] R. Servedio. Every linear threshold function has a low-weight approximator. In *CCC*, 2006.

[Tal89] M. Talagrand. A conjecture on convolution operators, and operators from l^1 to a Banach space. *Isr. J. Math.*, 68:82–88, 1989.

[Tal94] M. Talagrand. On Russo's approximate zero-one law. 22(3):1576–1587, 1994.

[Tal96] M. Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16:243–258, 1996.

[TV07] T. Tao and V. Vu. On the singularity probability of random Bernoulli matrices. *J. AMS*, 20:603–628, 2007.