Improved Pseudorandom Generators for Depth 2 Circuits

Anindya De[*] Omid Etesami[†] Luca Trevisan[‡] Madhur Tulsiani[§]

December 19, 2009

Abstract

We prove the existence of a $poly(n,m)$-time computable pseudorandom generator which "$1/poly(n,m)$-fools" DNFs with n variables and m terms, and has seed length $O(\log^2 nm \cdot \log\log nm)$. Previously, the best pseudorandom generator for depth-2 circuits had seed length $O(\log^3 nm)$, and was due to Bazzi (FOCS 2007).

It follows from our proof that a $1/m^{\tilde{O}(\log mn)}$-biased distribution $1/poly(nm)$-fools DNFs with m terms and n variables. For inverse polynomial distinguishing probability this is nearly tight because we show that for every m, δ there is a $1/m^{\Omega(\log 1/\delta)}$-biased distribution X and a DNF ϕ with m terms such that ϕ is not δ-fooled by X.

For the case of *read-once* DNFs, we show that seed length $O(\log mn \cdot \log 1/\delta)$ suffices, which is an improvement for large δ.

It also follows from our proof that a $1/m^{O(\log 1/\delta)}$-biased distribution δ-fools all read-once DNF with m terms. We show that this result too is nearly tight, by constructing a $1/m^{\tilde{\Omega}(\log 1/\delta)}$-biased distribution that does not δ-fool a certain m-term read-once DNF.

Keywords: DNF, pseudorandom generators, small bias spaces

[*] Computer Science Division, University of California, Berkeley, CA, USA. anindya@cs.berkeley.edu. Supported by the "Berkeley Fellowship for Graduate Study" and US-Israel BSF grant 2006060.

[†] Computer Science Division, University of California, Berkeley, CA, USA. etesami@cs.berkeley.edu. This material is based upon work supported by the National Science Foundation under grant No. CCF-0729137 and by the BSF under grant 2006060.

[‡] Computer Science Division, University of California, Berkeley, CA, USA. luca@cs.berkeley.edu. This material is based upon work supported by the National Science Foundation under grant No. CCF-0729137 and by the BSF under grant 2006060.

[§] Institute for Advanced Study, Princeton, NJ, USA. madhurt@math.ias.edu. This material is based upon work supported by the National Science Foundation under grant No. CCF-0832797 and IAS Sub-contract no. 00001583. Work done partly when the author was a graduate student at UC Berkeley.

1 Introduction

One of the main open questions in *unconditional* pseudorandomness and derandomization is to construct logarithmic-seed pseudorandom generators that "fool" bounded-depth circuits.[1] Ajtai and Wigderson [AW89] first considered the problem of pseudorandomness against bounded-depth circuits, and constructed a pseudorandom generator against AC^0 with a seed of length $O(n^\epsilon)$ for any $\epsilon > 0$. This was substantially improved by Nisan [Nis91], who used the hardness of parity against AC^0 [Hås86] to construct a pseudorandom generator against depth d circuits with a seed of length $O(\log^{2d+6} n)$. This remains the best known result for AC^0.

Even for depth-2 circuits, the construction of optimal pseudorandom generators remains a challenging open question. A depth-2 circuit is either a CNF or a DNF formula, and a pseudorandom generator that fools DNFs must also fool CNFs with the same distinguishing probability, so from now on we will focus without loss of generality on DNFs, and denote by n the number of variables and m the number of terms. We remark that even constructing an optimal pseudorandom generator for *read-once* DNFs would be interesting as Healy, Vadhan and Viola [HVV04] have shown a connection between such pseudorandom generators and hardness amplification in NP.

Nisan's result quoted above gives a pseudorandom generator for DNFs with seed length $O(\log^{10} nm)$. Luby, Velickovic and Wigderson [LVW93] reduced the seed length to $O(\log^4 nm)$ via various optimizations. For the simpler task of approximating the number of satisfying assignments to a DNF with m terms, Luby and Velickovic [LV96] provide a deterministic algorithm of running time $(m \log n)^{\exp(O(\sqrt{\log\log m}))}$.

The current best pseudorandom generator for DNFs is due to Bazzi [Baz07]. In 1990, Linial and Nisan [LN90] conjectured that depth-d circuits are fooled by every distribution that is $(\log mn)^{O_d(1)}$-wise independent. Bazzi proved the depth-2 case of the Linial-Nisan conjecture, and showed that every $O(\log^2(m/\delta))$-wise independent distribution δ-fools DNFs. This result gives two approaches to constructing a pseudorandom generator for DNFs of seed $O(\log n \cdot \log^2(m/\delta))$, which is $O(\log^3 nm)$ when $\delta = 1/poly(n, m)$. One is to use one of the known constructions of k-wise independent generators of seed length $O(k \log n)$. The other is to use a result of Alon, Goldreich and Mansour [AGM03] showing that every ϵ-biased distribution, in the sense of Naor and Naor [NN93], over n bits is ϵn^k-close to a k-wise independent distribution. This means that, because of Bazzi's theorem, every $exp(-O(\log n \cdot \log^2(m/\delta)))$-biased distribution fools DNFs; Naor and Naor [NN93] prove that an ϵ-biased distribution over n bits can be sampled using a seed of $O(\log(n/\epsilon))$ random bits, and so a $exp(-O(\log n \cdot \log^2(m/\delta)))$-biased distribution can be sampled using $O(\log n \cdot \log^2(m/\delta))$ random bits.

Razborov [Raz09] considerably simplified Bazzi's proof (retaining the same quantitative bounds). In a recent breakthrough, building on Razborov's argument, Braverman [Bra09] has recently proved the full Linian-Nisan conjecture.

For width-w DNF formulas (formulas with each term involving at most w variables), better bounds are known for small w. Luby and Velickovic [LV96] prove the existence of a generator with seed length $O(\log n + w2^w \log 1/\delta)$ which δ-fools all width-w DNFs. It follows from their proof that every $exp(-O(w2^w \log 1/\delta))$-biased distribution δ-fools width-w DNFs. One may always assume without loss of generality that $w = O(\log(m/\delta))$, and so if the Luby-Velickovic result could be improved to a seed length of $O(w + \log(n/\delta))$, the result would be a generator of optimal seed

[1] We say a random variable X, ranging over $\{0,1\}^n$, "δ-fools" a function $f : \{0,1\}^n \to \mathbb{R}$ if $\mathbb{E}f(X) - \mathbb{E}f(U_n)| \le \delta$, where U_n is uniformly distributed over $\{0,1\}^n$. If $\mathcal{C}$ is a class of functions, then we say that X δ-fools $\mathcal{C}$ if X δ-fools every function $f \in \mathcal{C}$.

	DNF Family	Seed length
[Nis91]	general DNFs	$O(\log^{10}(mn/\delta))$
[LVW93]	general DNFs	$O(\log^{4}(mn/\delta))$
[Baz07]	general DNFs	$O(\log n \cdot \log^2(m/\delta))$
This work	general DNFs	$O(\log n + \log^2(m/\delta) \cdot \log\log(m/\delta))$
[LV96]	width-w DNFs	$O(\log n + w2^w \cdot \log(1/\delta))$
This work	width-w DNFs	$O(\log n + w \log w \cdot \log(m/\delta))$
[Baz03]	read-once DNFs	$O(\log n \cdot \log m \cdot \log(1/\delta))$
This work	read-once DNFs	$O(\log n + \log m \cdot \log(1/\delta))$

Figure 1: Pseudorandom generators to δ-fool DNFs with m terms and n variables

length $O(\log(mn/\delta))$.

For read-once DNFs (formulas with each variable appearing in at most one term), Bazzi proves that every $O(\log m \cdot \log 1/\delta)$-wise independent distribution δ-fools every read-once DNF, and hence every $exp(-O(\log n \cdot \log m \cdot \log 1/\delta))$-biased distribution δ-fools read-once DNFs. This gives a generator of seed length $O(\log n \cdot \log m \cdot \log 1/\delta)$, which is $O(\log^2 nm)$ for constant δ.

Our Results

We prove that every width-w DNF is δ-fooled by every $\exp(-O(\log n + w \log w(\log m + \log 1/\delta)))$-biased distribution. This gives a pseudorandom generator of seed length $O(\log^2 mn \cdot \log\log mn)$ for general DNFs and $\delta = 1/\mathrm{poly}(n, m)$.

Regarding read-once DNFs, we show that they are δ-fooled by every $exp(-O(\log m \cdot \log 1/\delta))$-biased distribution, leading to a generator with seed length $O(\log n + \log m \cdot \log 1/\delta)$, which is $O(\log nm)$ for constant δ. Unfortunately this is still not sufficient in order to improve the hardness amplification in [HVV04], which requires a pseudorandom generator with $\delta = 1/\mathrm{poly}(n, m)$.

We prove that our quantitative connections between small bias and DNF derandomization are nearly tight. Specifically, we construct an m-term DNF that is not δ-fooled by a certain $1/m^{\Omega(\log 1/\delta)}$-biased distribution, which means that seed length $\Omega(\log n + \log m \cdot \log 1/\delta)$ is necessary if one wants to δ-fool DNFs using a generic small bias distribution. This matches our positive result up to a $\log\log nm$ term when $\delta = 1/poly(n, m)$. It remains open whether seed length $O(\log nm)$ is achievable for constant δ.

We also construct an m-term *read-once* DNF that is not δ-fooled by a certain $1/m^{\tilde{\Omega}(\log 1/\delta)}$-biased distribution (where the $\tilde{\Omega}$ notation hides a $1/\log\log 1/\delta$ term). This means that seed length $\Omega(\log^2 nm/\log\log nm)$ is necessary if one wants to $1/\mathrm{poly}(nm)$-fool read-once DNFs using a generic small bias distribution.

Our Techniques

Our positive results for DNFs and read-once DNFs are based on techniques similar to the ones developed by Bazzi [Baz07] and simplified by Razborov [Raz09].

Bazzi shows that a sufficient (and necessary) condition for a function g to be δ-fooled fooled by a k-wise independent distribution is that the function be "sandwiched" between two bounded real-valued functions f_ℓ, f_u which are degree-k polynomials and such that $f_\ell(x) \leq g(x) \leq f_u(x)$ holds for every x, and $\mathbb{E}_{x\in U_n}[f_u(x) - f_\ell(x)] \le \delta$. We provide a similar sufficient (and necessary) condition for a function g to be δ-fooled by an ϵ-biased distribution in terms of g being sandwiched between functions *whose Fourier transform has small* ℓ_1 *norm.*

Bazzi and Razborov then proceed to show how to construct the sandwiching functions for every DNF by showing that it suffices to find just one low-degree function that approximates the DNF in the ℓ_2 norm, and such a function is provided by a result of Linial, Mansour and Nisan [LMN93] on the Fourier spectrum of DNFs. Our goal, instead, is to find a function of small ℓ_1 Fourier norm which approximates the given DNF well in the ℓ_2 norm. The existence of such a function is guaranteed by a result of Mansour [Man95].

For the case of read-once DNFs we explicitly construct the sandwiching functions with bounded Fourier ℓ_1 norm, using the inclusion-exclusion formula for the DNF. To analyze the error in the truncated inclusion-exclusion formula, we apply an argument which is similar to the one appearing in a paper by Even *et al.* [EGL$^+$92] on the related subject of pseudorandomness for combinatorial rectangles. The technical difference between our argument and the one in [EGL$^+$92] is that while they use the k^{th}-truncations of the inclusion-exclusion series to directly show that k-wise independence fools combinatorial rectangles, we use these to compute functions with low ℓ_1 norm sandwiching the given DNF.

Our negative example for general DNFs is related to a counterexample by Mansour (cited in [LV96]). Mansour shows that there is a k-wise independent distribution that does not δ-fool a certain m-term DNF, where $k = (\log m) \cdot (\log 1/\delta)$, showing that for $\delta = 1/\mathrm{poly}(n, m)$ the analysis of Bazzi is optimal. Mansour's distribution is uniform over the bit strings of odd parity, and so it is not a small-bias distribution. We show that one can use, instead, the uniform distribution over bit strings whose number of ones is not a multiple of 3, which is a small bias distribution.

Our negative example for read-once DNFs is somewhat more technical. We start from a "tribes" function, a read-once DNF with m terms each with $\log m$ literals, and we show how to construct a $1/m^{\tilde{\Omega}(\log 1/\delta)}$-biased distribution that does not δ-fool the tribes function. We show that for every parameter d we can construct a distribution X that is roughly $1/m^d$-biased, and is such that the distinguishing probability of the tribe between X and the uniform distribution is the same as the error of the d-th term of the inclusion-exclusion formula in approximating the tribe. The latter error is roughly $1/d!$, so we get our result by setting $d = (\log 1/\delta)/(\log\log 1/\delta)$.

2 Preliminaries

We start by reviewing some basic Fourier analysis.

Definition 2.1 (Fourier analysis on $\{0,1\}^n$) *The characters of* $\{0,1\}^n$ *is the set of linear functions from* $\{0,1\}^n$ *to* $\{-1,1\}$ *given by*

$$\chi_S(x) = \prod_{i\in S}(-1)^{x_i} \text{ where } S \subseteq [n].$$

It is easy to see that the following identities are true.

- *For any character* χ, $||\chi||_2 = \mathbb{E}_{x\in U_n}[\chi^2(x)] = 1$.
- *For two distinct characters,* χ *and* χ', $\langle \chi, \chi' \rangle = \mathbb{E}_{x\in U_n}[\chi(x)\chi'(x)] = 0$.

Note that there are 2^n *characters and hence they form an orthonormal basis for the functions mapping* $\{0,1\}^n$ *to* $\mathbb{R}$. *Therefore, every function* f *can be expressed as a linear combination of* *these characters which is called the Fourier expansion. The Fourier expansion of* $f : \{0,1\}^n \to \mathbb{R}$ *is*

$$f(x) = \sum_S \hat{f}(S)\chi_S(x).$$

In the above, $\hat{f}(S)$ *is called the Fourier coefficient corresponding to the set* S*. It is easy to check that the following identity (known as Parseval-Plancherel identity) is true*

$$\sum_S \hat{f}^2(S) = \mathop{\mathbb{E}}_{x \in U_n}[f^2(x)]$$

We use the following notation for the Fourier ℓ_1 *norm of* f *and a minor variant of it as below:*

$$\|f\|_1 \;:=\; \sum_S \left|\hat{f}(S)\right| \qquad \text{and} \qquad \|f\|_1^{\neq\emptyset} \;:=\; \sum_{S \neq \emptyset} \left|\hat{f}(S)\right|$$

Definition 2.2 *We say a probability distribution* X *over* $\{0,1\}^n$ ϵ*-fools a real function* $f : \{0,1\}^n \to \mathbb{R}$ *if*

$$|\mathbb{E}[f(X)] - \mathbb{E}[f(U_n)]| \leq \epsilon.$$

We say a probability distribution X *over* $\{0,1\}^n$ *is* ϵ*-biased if it* ϵ*-fools the character functions* χ_S*.*

Proposition 2.3 (Efficient construction of ϵ-biased sets [NN93, AGHP92]) *A subset* $B \subseteq \{0,1\}^n$ *is called an* ϵ*-biased set if the uniform distribution with support* B *is* ϵ*-biased. There exist* ϵ*-biased sets of size* $O(n^2/\epsilon^2)$ *such that a random element from the set can be sampled using a seed of length* $2\log(n/\epsilon) + O(1)$*, in time* $\mathrm{poly}(n, \log(1/\epsilon))$*.*

Definition 2.4 (DNF) *A DNF formula* ϕ *is of the form* $\phi = \bigvee_{i=1}^m C_i$ *where each term* C_i *is an AND of literals (variables or negations). A formula* ϕ *is said to be of* width w *if every term* C_i *involves at most* w *distinct variables. A DNF is said to be* read-once *if every variable appears in at most one of the terms.*

2.1 Sandwich bound

In this section, we state a characterization of functions that can be fooled well by ϵ-biased probability distributions. The characterization derived here is similar to the one derived by Bazzi [Baz07] in context of k-wise independent distributions. The first observation is that if f has a small Fourier ℓ_1 norm, then it is fooled by small ϵ-biased sets:

Lemma 2.5 *Every function* $f : \{0,1\}^n \to \mathbb{R}$ *is* $\epsilon\|f\|_1^{\neq\emptyset}$*-fooled by any* ϵ*-biased probability distribution.*

Proof: Let X be sampled from an ϵ-biased distribution. We have

$$\begin{aligned}
|\mathbb{E}[f(X)] - \mathbb{E}[f(U_n)]| &= \left|\mathbb{E}\left[\sum_S \hat{f}(S)\chi_S(X)\right] - \hat{f}(\emptyset)\right| \\
&= \left|\sum_{S\neq\emptyset} \hat{f}(S)\mathbb{E}[\chi_S(X)]\right| \\
&\leq \epsilon \sum_{S\neq\emptyset} |\hat{f}(S)| \;\leq\; \epsilon\|f\|_1^{\neq\emptyset}.
\end{aligned}$$

■

We can strengthen Lemma 2.5 as follows.

Proposition 2.6 (Sandwich bound) *Suppose $f, f_\ell, f_u : \{0,1\}^n \to \mathbb{R}$ are three functions such that for every $x \in \{0,1\}^n$ we have $f_\ell(x) \le f(x) \le f_u(x)$. Furthermore, assume $\mathbb{E}[f(U_n)] - \mathbb{E}[f_\ell(U_n)] \le \delta$ and $\mathbb{E}[f_u(U_n)] - \mathbb{E}[f(U_n)] \le \delta$. Let $l = \max(\|f_\ell(x)\|_1^{\neq\emptyset}, \|f_u(x)\|_1^{\neq\emptyset})$. Then any ϵ-biased probability distribution $(\delta + \epsilon l)$-fools f.*

Proof: Let X be an ϵ-biased random variable. We have

$$\begin{aligned}\mathbb{E}[f(X)] &\le \mathbb{E}[f_u(X)] \\ &\le \mathbb{E}[f_u(U_n)] + \epsilon\|f_u\|_1^{\neq\emptyset} \\ &\le \mathbb{E}[f(U_n)] + \delta + \epsilon\|f_u\|_1^{\neq\emptyset}.\end{aligned}$$

Similarly we have $\mathbb{E}[f(X)] \ge \mathbb{E}[f(U_n)] - \delta - \epsilon\|f_\ell\|_1^{\neq\emptyset}$. Thus the result follows. ■

The following result shows that the condition of Proposition 2.6 is not only a sufficient condition for being fooled by ϵ-biased distributions but also a necessary condition.

Proposition 2.7 (Inverse of the sandwich bound) *Suppose $f : \{0,1\}^n \to \mathbb{R}$ is ϵ'-fooled by any ϵ-biased set. Then there exist functions $f_\ell, f_u : \{0,1\}^n \to \mathbb{R}$ and $\delta, l \in \mathbb{R} \ge 0$ with the following properties:*

- *For every $x \in \{0,1\}^n$ we have $f_\ell(x) \le f(x) \le f_u(x)$.*
- *$\mathbb{E}[f(x)] - \mathbb{E}[f_\ell(x)] \le \delta$ and $\mathbb{E}[f_u(x)] - \mathbb{E}[f(x)] \le \delta$,*
- *$\|f_\ell(x)\|_1^{\neq\emptyset} \le l$, $\|f_u(x)\|_1^{\neq\emptyset} \le l$, and $\delta + \epsilon l \le \epsilon'$.*

Proof: Consider the following linear program in variables p_x:

$$\begin{array}{ll}\min & \sum_x f(x)p_x \\ & \sum_x p_x = 1 \\ \forall S \neq \emptyset & \sum_x p_x\chi_S(x) \ge -\epsilon \\ \forall S \neq \emptyset & \sum_x p_x\chi_S(x) \le \epsilon \\ \forall x & p_x \ge 0\end{array}$$

where $x \in \{0,1\}^n$ and $S \subseteq \{1, \ldots, n\}$. The constraints specify that p_x is the probability distribution of an ϵ-biased random variable. Since f is ϵ'-fooled by ϵ-biased sets, the optimum value of the LP is $\ge \mathbb{E}[f(U_n)] - \epsilon'$.

We now write the dual of the above LP:

$$\begin{array}{ll}\max & z - \epsilon\sum_{S\neq\emptyset}(y_S^+ + y_S^-) \\ \forall x & z + \sum_{S\neq\emptyset}\chi_S(x)(y_S^+ - y_S^-) \le f(x) \\ \forall S \neq \emptyset & y_S^+, y_S^- \ge 0\end{array}$$

which is equivalent to

$$\begin{array}{ll}\max & z - \epsilon\sum_{S\neq\emptyset}|y_S| \\ \forall x & z + \sum_{S\neq\emptyset}\chi_S(x)y_S \le f(x)\end{array}$$

Since the optimum value of the primal is $\geq \mathbb{E}[f(U_n)] - \epsilon'$, there exists a feasible set of values z^* and y^*_S for the above optimization program such that $z^* - \epsilon \sum_{S \neq \emptyset} |y^*_S| \geq \mathbb{E}[f(U_n)] - \epsilon'$. Let $f_\ell(x) = z^* + \sum_{S \neq \emptyset} y^*_S \chi_S(x)$. Clearly $\mathbb{E}[f_\ell(U_n)] = z^*$ and $\sum_{S \neq \emptyset} |y^*_S| = \|f_\ell\|_1^{\neq \emptyset}$; Set $\delta = \mathbb{E}[f(U_n)] - z^*$ and $l = \sum_{S \neq \emptyset} |y^*_S|$. It is easy to check that f_ℓ, δ, l so defined satisfies all the constraints. Similarly, one can consider a different primal where the objective is to maximize $\sum_x f(x) p_x$ and then use its dual to define f_u which satisfies the aforementioned conditions. ■

It is easy to observe the following properties of ℓ_1 norm of functions over the Fourier domain.

Observation 2.8 *If $f, g : \{0,1\}^n \to \mathbb{R}$, then $\|f+g\|_1 \leq \|f\|_1 + \|g\|_1$ and $\|fg\|_1 \leq \|f\|_1 \|g\|_1$*

Observation 2.9 *If $\phi : \{0,1\}^n \to \{0,1\}$ is an AND of some subset of literals (i.e., variables or their negations), then $\|\phi\|_1 = 1$.*

3 Fooling Read-once DNF formulas

In this section, we show that ϵ-biased sets can fool read-once DNFs. In particular, we show the following theorem.

Theorem 3.1 *Let ϕ be a read-once DNF formula with m terms. For $1 \leq k \leq m$, ϵ-biased distributions $O(2^{-\Omega(k)} + \epsilon m^k)$-fool ϕ. In particular, we can δ-fool ϕ by an ϵ-biased distribution, for $\epsilon = m^{-O(\log(1/\delta))}$.*

If we plug in the construction from Proposition 2.3, we get a pseudorandom generator which δ-fools a read-once DNF with n variables and m terms and has seed length $O(\log n + \log m \cdot \log(1/\delta))$. Before going into the proof of Theorem 3.1, we recall the inclusion-exclusion principle.

Let $A_1, \ldots, A_m$ be m arbitrary events in a probability space. The principle of inclusion and exclusion asserts that

$$\Pr[A_1 \cup \cdots \cup A_m] = \sum_{j=1}^{m} (-1)^{j-1} T_j,$$

where

$$T_j = \sum_{S \subseteq [m], |S|=j} \Pr\left[\bigcap_{i \in S} A_i\right].$$

Moreover, the partial sum $\sum_{j=1}^{r} (-1)^{j-1} T_j$ is an upper bound for $\Pr[A_1 \cup \cdots \cup A_m]$ for odd values of r, and a lower bound for $\Pr[A_1 \cup \cdots \cup A_m]$ for even values of r.

We now return to the proof of Theorem 3.1. The proof follows that of Theorem 2 in [EGL+92].

Proof: [of Theorem 3.1] Let $\phi = C_1 \vee \cdots \vee C_m$ be the read-once formula. For $1 \leq i \leq m$, let A_i denote the event that term C_i is satisfied. We divide the analysis into two cases depending on whether $\sum_{i=1}^{m} \Pr[A_i] \leq k/(2\mathrm{e})$ or not.

Case 1: $\sum_{i=1}^{m} \Pr[A_i] \leq k/(2\mathrm{e})$.

Let T_k denote the kth term of the inclusion-exclusion formula. Since the terms are disjoint, we have

$$T_k = \sum_{S \subseteq [m], |S|=k} \prod_{i \in S} \Pr[A_i].$$

We now observe that $T_k \leq 2^{-k}$. Indeed, subject to the restriction $\sum_{i=1}^{m} \Pr[A_i] = \alpha$ and $\Pr[A_i] \geq 0$, a convexity based argument implies that T_k is maximized when all the $\Pr[A_i]$'s are equal implying that $T_k \leq \binom{m}{k}(2\mathrm{e}m/k)^{-k} \leq 2^{-k}$.

Consider the rth approximation to ϕ, obtained by inclusion-exclusion:

$$\phi_r(x) = \sum_{j=1}^{r}(-1)^{j-1} \sum_{S\subseteq[m],|S|=j} \bigwedge_{l\in S} C_l(x),$$

where $\bigwedge$ is the AND function. The functions ϕ_{k-1} and ϕ_k sandwich ϕ and we shall use them in applying Proposition 2.6. To verify the conditions, we note that the function $\bigwedge_{l\in S} C_l(x)$ is an AND of AND terms, therefore $\| \bigwedge_{l\in S} C_l(x)\|_1^{\neq\emptyset} = O(1)$, and hence $\|\phi_r\|_1^{\neq\emptyset} = O(m^r)$. We also have $|\mathbb{E}[f_k(U_n)] - \mathbb{E}[f_{k-1}(U_n)]| = T_k \leq 2^{-k}$. and hence, by Proposition 2.6, ϕ is $O(2^{-k} + \epsilon m^k)$-fooled by ϵ-biased distributions.

Case 2: $\sum_{i=1}^{m} \Pr[A_i] > k/(2\mathrm{e})$.

Consider the first m' where $\sum_{i=1}^{m'} \Pr[A_i] \geq k/(2\mathrm{e})$. Define $\phi' = C_1 \vee \cdots \vee C_{m'}$. Observe that the DNF ϕ' is satisfied with probability $1 - 2^{-\Omega(k)}$, for it is not satisfied with probability $\prod_{i=1}^{m'}(1 - \Pr[A_i]) \leq (1 - k/(2\mathrm{e}m'))^{m'} \leq 2^{-\Omega(k)}$. (Again by a convexity argument, $\prod_i(1 - \Pr[A_i])$ is maximized when $\Pr[A_i]$s are equal.)

Let $\phi'_r(x)$ denote the rth approximation to ϕ'. Also, (without loss of generality) let k be even so that $\phi'_k \leq \phi' \leq \phi$. Note that while ϕ'_{k-1} is a an upper bound on ϕ', it is *not* an upper bound on ϕ. We shall use ϕ'_k and identically 1 function respectively as lower and upper bounds for applying Proposition 2.6 to ϕ.

From argument above, we know that $\mathbb{E}[1 - \phi] \leq \mathbb{E}[1 - \phi'] \leq 2^{-\Omega(k)}$. To bound $\mathbb{E}[\phi - \phi'_k]$, we note that

$$\mathbb{E}\left[\phi - \phi'_k\right] \;=\; \mathbb{E}\left[\phi - \phi'\right] + \mathbb{E}\left[\phi' - \phi'_k\right] \;\leq\; \mathbb{E}\left[1 - \phi'\right] + \mathbb{E}\left[\phi'_{k-1} - \phi'_k\right] \;\leq\; 2^{-\Omega(k)}$$

where in the last inequality we used that $\mathbb{E}[\phi'_{k-1} - \phi'_k]$ as in the previous case, since $\sum_{i=1}^{m'} \Pr[A_i] < k/(2\mathrm{e}) + 1$. The bound on the $\|\phi'_k\|_1^{\neq\emptyset}$ is as before. Applying Proposition 2.6, we then get that ϵ-biased sets $O(2^{-\Omega(k)} + \epsilon m'^k)$-fool ϕ. ■

4 Fooling general DNF formulas

In this section, we show that small biased distributions fool general DNFs. While the seed length will not be as good as in the previous section, the result will be more general. Also, this section will involve use of more analytic tools. Our proof shall be along the lines of Razborov's simplified proof of Bazzi's theorem [Raz09]. The following two theorems will be the main theorems of this section.

Theorem 4.1 *Let ϕ be a width w-DNF formula with m terms. Then, ϕ is δ-fooled by an ϵ-biased distribution where $\epsilon = w^{-O(w\log(m/\delta))}$.*

Theorem 4.2 *Let ϕ be a DNF formula with m terms. Then, ϕ is δ-fooled by an ϵ-biased distribution where $\epsilon = (\log(m/\delta))^{O(-\log^2(m/\delta))}$.*

Plugging in the pseudorandom generator construction from Proposition 2.3 in Theorem 4.1, we get a pseudorandom generator which δ-fools width-w DNFs with m terms over n variables and has a seed of length $O(\log n + w \log w \log(m/\delta))$. Doing the same for Theorem 4.2, we get a pseudorandom generator which δ-fools DNFs with m terms over n variables and has a seed of length $O(\log n + \log^2(m/\delta) \log\log(m/\delta))$. Theorem 4.2 follows by a reduction to Theorem 4.1, by deleting the terms with large width, as we describe later. For most of this section, we will be concerned with DNFs of a bounded width. To prove Theorem 4.1, we will be interested in finding sandwiching functions f_l and f_u to apply Proposition 2.6.

Using an argument similar to [Baz07], we reduce this to the problem of finding a function g such that $\|\phi - g\|_2$ and $\|g\|_1$ are small, and $\phi(x) = 0 \implies g(x) = 0$. We then show how to remove the last condition and then find an appropriate g using a Fourier concentration result of Mansour [Man95]. More formally, we prove the following three lemmas.

Lemma 4.3 *Let $\phi : \{0,1\}^n \to \{0,1\}$ be a DNF with m terms and $g : \{0,1\}^n \to \mathbb{R}$ be such that: $\|g\|_1 \leq l$, $||\phi - g||_2 \leq \epsilon_1$ and $g(x) = 0$ whenever $\phi(x) = 0$. Then, we can get $f_\ell, f_u : \{0,1\}^n \to \mathbb{R}$ such that*

- *$\forall\ x$, $f_\ell(x) \leq \phi(x) \leq f_u(x)$*
- *$\mathbb{E}_{x\in U_n}[f_u(x) - \phi(x)] \leq m\epsilon_1^2$ and $\mathbb{E}_{x\in U_n}[\phi(x) - f_\ell(x)] \leq m\epsilon_1^2$.*
- *$\|f_\ell\|_1$, $\|f_u\|_1 \leq (m+1)(l+1)^2 + 1$*

Lemma 4.4 *Let $\phi : \{0,1\}^n \to \{0,1\}$ be a width-w DNF with m terms. Suppose for every width-w DNF ϕ_1, there is a function $g_1 : \{0,1\}^n \to \mathbb{R}$ such that: $\|g_1\|_1 \leq l_1$ and $||\phi_1 - g_1||_2 \leq \epsilon_2$. Then, we can get $g : \{0,1\}^n \to \mathbb{R}$ such that $\|g\|_1 \leq m(l_1 + 1)$, $||\phi - g||_2 \leq m\epsilon_2$ and $g(x) = 0$ whenever $\phi(x) = 0$.*

Lemma 4.5 *Let $\phi : \{0,1\}^n \to \{0,1\}$ be a width w DNF and $\epsilon_2 > 0$. Then there is a function $g_1 : \{0,1\}^n \to \mathbb{R}$ such that $||\phi - g_1||_2 \leq \epsilon_2$ and $\|g_1\|_1 = w^{O(w \log(1/\epsilon_2))}$*

Before, we prove these lemmas, we show how it implies Theorem 4.1.

Proof: [of Theorem 4.1] Set $\epsilon_2 = \sqrt{\delta/2m^3}$ and $\epsilon_1 = \sqrt{\delta/2m}$. By applying Lemma 4.5, for every width-w DNF ϕ_1, we can get a function $g_1 : \{0,1\}^n \to \mathbb{R}$ such that

- $||\phi_1 - g_1||_2 \leq \epsilon_2 = \sqrt{\delta/2m^3}$
- $||g_1||_1 = w^{O(w\log(1/\epsilon_2))} = w^{O(w\log(m/\delta))}$

Now, we apply Lemma 4.4 with $l_1 = w^{O(w\log(m/\delta))}$ and $\epsilon_2 = \sqrt{\delta/2m^3}$. Then, for the given DNF ϕ, we get a function g such that $||g||_1 = w^{O(w\log(m/\delta))}$ and $||g - \phi||_2 \leq m\epsilon_2 = \epsilon_1 = \sqrt{\delta/2m}$. Finally, we apply Lemma 4.3 with g and ϵ_1 as defined and $l = w^{O(w\log(m/\delta))}$ to get f_ℓ and f_u such that ϕ is sandwiched by f_ℓ and f_u, $||f_\ell||_1, ||f_u||_1 \leq w^{O(w\log(m/\delta))}$ and

$$\mathop{\mathbb{E}}_{x\in U_n}[f_u(x) - \phi(x)] \leq \frac{\delta}{2} \quad \text{and} \quad \mathop{\mathbb{E}}_{x\in U_n}[\phi(x) - f_\ell(x)] \leq \frac{\delta}{2}$$

By applying Proposition 2.6, we get that an $\epsilon = w^{-O(w\log(m/\delta))}$ (for an appropriately large constant inside $O(\cdot)$) biased set fools ϕ by $\delta/2 + \epsilon l \leq \delta$. ■

We now get back to proofs of Lemma 4.3, Lemma 4.4 and Lemma 4.5. We start with proof of Lemma 4.3.

Proof: [of Lemma 4.3] Let $\phi = \bigvee_{i=1}^{m} A_i$ where A_i are the terms. We define f_ℓ and f_u as follows:

- $f_\ell = 1 - (1-g)^2$
- $f_u = 1 - (1 - \sum_{i=1}^{m} A_i)(1-g)^2$

We note that this is the same construction of functions as in Lemma 3.3 in [Baz07]. In particular, the following two things are already proven there.

- $\forall\ x,\ f_\ell(x) \leq \phi \leq f_u(x)$
- $\mathbb{E}_{x\in U_n}[f_u(x) - \phi(x)] \leq m||\phi - g||_2^2$ and $\mathbb{E}_{x\in U_n}[\phi(x) - f_\ell(x)] \leq m||\phi - g||_2^2$

Using this, we have the proof of the first two items in the lemma. Only the third item *i.e.,* bound on $\|f_\ell\|_1$ and $\|f_u\|_1$ remains to be proven. To get this, we use Observation 2.8 and Observation 2.9 along with the hypothesis $\|g\|_1 \leq l$. Using this, we get that $\|f_\ell\|_1 \leq 1 + (1+l)^2$ and $\|f_u\|_1 \leq 1 + (m+1)(l+1)^2$ which proves the lemma. ■

We now turn to the proof of Lemma 4.4. The proof follows the proof by Razborov [Raz09] with some changes.

Proof: [of Lemma 4.4] We first observe as in [Raz09] (attributed to Avi Wigderson) that if $\phi = \bigvee_{i=1}^{m} A_i$ where A_i are the individual terms, then ϕ can be rewritten as $\sum_{i=1}^{m} A_i(1 - \bigvee_{j=1}^{i-1} A_j)$. Let us write $\bigvee_{j=1}^{i-1} A_j = \phi_i$ ($\phi_i = 0$ if $i = 1$). Then, we can say that $\phi = \sum_{i=1}^{m} A_i(1-\phi_i)$. Note that each of the ϕ_i is a width w-DNF. Hence, we can apply our hypothesis to get functions $g_1, \ldots, g_m : \{0,1\}^n \to \mathbb{R}$ such that for all i, $\|g_i\|_1 \leq l_1$ and $\|g_i - \phi_i\|_2 \leq \epsilon_2$. Let us now consider the function $g : \{0,1\}^n \to \mathbb{R}$ defined as

$$g = \sum_{i=1}^{m} A_i(1 - g_i)$$

We observe that if $\phi(x) = 0$ for some x, then $\forall\ i,\ A_i(x) = 0$ which implies that $g(x) = 0$. Applying Observation 2.8 and using that A_i's are terms and hence $||A_i||_1 = 1$, we also get that $\|g\|_1 \leq m(l_1+1)$. So, the only thing that remains to be proven is that $\|\phi - g\|_2 \leq m\epsilon_2$. Though this is done in [Raz09], we do it here for the sake of completeness.

$$\begin{aligned}
\|g - \phi\|_2^2 &= \mathop{\mathbb{E}}_{x\in U_n}\left[\left(\sum_{i=1}^{m} A_i(\phi_i - g_i)(x)\right)^2\right] \\
&\leq m \mathop{\mathbb{E}}_{x\in U_n}\left[\sum_{i=1}^{m}(A_i(\phi_i - g_i)(x))^2\right] \qquad \text{(By Jensen's inequality)} \\
&= m\sum_{i=1}^{m} \mathop{\mathbb{E}}_{x\in U_n}\left[(A_i(\phi_i - g_i)(x))^2\right] \\
&\leq m\sum_{i=1}^{m} \mathop{\mathbb{E}}_{x\in U_n}\left[(\phi_i - g_i)(x)^2\right] \qquad \text{(Using } A_i \text{ is bounded by 1)} \\
&= m\sum_{i=1}^{m} ||\phi_i - g_i||_2^2 \quad \leq \quad m^2\epsilon_2^2 \qquad \text{(Using } ||\phi_i - g_i||_2 \leq \epsilon_2\text{)}
\end{aligned}$$

This proves that $||\phi - g||_2 \leq m\epsilon_2$ which finishes the proof. ■

We now come to the proof of Lemma 4.5. The proof is dependent upon the following well-known concentration result by Mansour [Man95] (or see Ryan O'Donnell's lecture notes on Fourier analysis [OD07]).

Theorem 4.6 *[Man95] Let $\phi : \{0,1\}^n \to \{0,1\}$ be a width w-DNF with m terms and $\epsilon_2 > 0$. Let $\sum_{S\subset[n]} \hat{\phi}(S)\chi_S$ be the Fourier expansion of ϕ. Then there is a subset $\Gamma \subset 2^{[n]}$ of size $w^{O(w\log(1/\epsilon_2))}$ such that g defined as $g_1 = \sum_{S\in\Gamma} \hat{\phi}(S)\chi_S$ is such that $||\phi - g_1||_2 \le \epsilon_2$.*

Proof: [of Lemma 4.5] For the given ϕ and ϵ_2, let g_1 be the function given by Theorem 4.6. Clearly, it satisfies $||\phi - g_1||_2 \le \epsilon_2$. To bound $\|g_1\|_1$, note that $\|g_1\|_1 = \sum_{S\in\Gamma} |\hat{\phi}(S)|$ where $|\Gamma| = w^{O(w\log(1/\epsilon_2))}$. Note that $\sum_{S\in\Gamma} |\hat{\phi}(S)|^2 = \alpha$ for some $\alpha \in [0,1]$ (by Parseval-Plancherel identity and the fact that ϕ lies in $[0,1]$). Now, we have

$$\left(\sum_{S\in\Gamma} |\hat{\phi}(S)|\right)^2 \le |\Gamma| \left(\sum_{S\in\Gamma} |\hat{\phi}(S)|^2\right) \le |\Gamma| \qquad \text{(By Jensen's inequality)}$$

Hence, this gives us $\sum_{S\in\Gamma} |\hat{\phi}(S)| \le \sqrt{|\Gamma|} = w^{O(w\log(1/\epsilon_2))}$ which proves the lemma. ■

Theorem 4.2 now follows by reducing the case of arbitrary DNFs to that of bounded width, by deleting the terms with width greater than $\log(m/2\delta)$ and arguing that the change in the distinguishing probability is small.

Proof: [of Theorem 4.2] Let ϕ_w be the DNF obtained by removing all the terms from ϕ which have more than w literals, for a value of w to be specified later. Note that $\forall\ x$, $\phi_w(x) \le \phi(x)$. Also, note that

$$\mathop{\mathbb{E}}_{x\in U_n}[\phi(x) - \phi_w(x)] \le \Pr_{x\in U_n}[\exists \text{ term present in } \phi \text{ but not in } \phi_w \text{ which is satisfied}] \le m2^{-w}$$

The last inequality uses that all the terms present in ϕ but not ϕ_w have more than w literals and hence are satisfied with probability at most 2^{-w} under the uniform distribution. Also, let D be any ϵ-biased distribution. We can again say that

$$\mathop{\mathbb{E}}_{x\in D}[\phi(x) - \phi_w(x)] \le \Pr_{x\in D}[\exists \text{ term present in } \phi \text{ but not in } \phi_w \text{ which is satisfied}] \le m(2^{-w} + \epsilon)$$

The last inequality uses that under a ϵ-biased distribution, a term of width-w is satisfied with probability at most $2^{-w} + \epsilon$. This is because a term has ℓ_1 norm 1 and hence is ϵ fooled by a ϵ-biased distribution. Using the above two inequalities as well as $\phi_w \le \phi$, we can say

$$\mathop{\mathbb{E}}_{x\in D}\phi(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi(x) \ge \mathop{\mathbb{E}}_{x\in D}\phi_w(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi(x) \ge \mathop{\mathbb{E}}_{x\in D}\phi_w(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi_w(x) - m2^{-w}$$

$$\mathop{\mathbb{E}}_{x\in D}\phi(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi(x) \le \mathop{\mathbb{E}}_{x\in D}\phi(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi_w(x) \le \mathop{\mathbb{E}}_{x\in D}\phi_w(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi_w(x) + m(\epsilon + 2^{-w})$$

which together imply that

$$|\mathop{\mathbb{E}}_{x\in D}\phi(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi(x)| \le |\mathop{\mathbb{E}}_{x\in D}\phi_w(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi_w(x)| + m(\epsilon + 2^{-w})$$

Let us put $w = \log(2m/\delta)$. Then, Theorem 4.1 says that $|\mathbb{E}_{x\in D}\phi_w(x) - \mathbb{E}_{x\in U_n}\phi_w(x)|$ is $\delta/4$ fooled by an ϵ biased distribution where $\epsilon = w^{-O(w\log(m/\delta))} = (\log(m/\delta))^{-O(\log^2(m/\delta))}$. Then,

$$|\mathop{\mathbb{E}}_{x\in D}\phi(x) - \mathop{\mathbb{E}}_{x\in U_n}\phi(x)| \le \frac{\delta}{4} + m(\epsilon + 2^{-w}) \le \frac{\delta}{4} + \frac{\delta}{2} + m(\log(m/\delta))^{-O(\log^2(m/\delta))} \le \delta$$

■

5 Limitations of small biased spaces

In this section we provide various lower bounds on fooling DNFs by ϵ-biased distributions. Recall that in Section 3, we showed that a bias less than $m^{-O(\log(1/\delta))}$ is sufficient to δ-fool a read-once DNF with m terms. We first give a simple example which shows that this bound is optimal when δ is a small constant.

For smaller values of δ, we give a somewhat more technical construction, which shows that the bias needs to be less than $m^{-\Omega(\log(1/\delta)/\log\log(1/\delta))}$ to δ-fool a read-once DNF with m terms. Note that this would also imply the optimality for constant δ but we choose to retain the previous example due to its simplicity.

For the case of general DNFs, we give an instance showing that ϵ must be necessarily less than $m^{-\Omega(\log(1/\delta))}$. This does match our bound for the case of read-once DNFs, but is somewhat far from the upper bound we provide in Section 4 (which uses $\epsilon = (\log(m/\delta))^{-O(\log^2(m/\delta))}$).

5.1 Lower bounds for read-once DNFs when $\delta = \Theta(1)$

Our analysis gives that for $\delta = \Theta(1)$ and $m = n^{\Theta(1)}$, an ϵ-biased distribution with $\epsilon = n^{-\Theta(1)}$ suffices to δ-fool a read-once DNF with m terms. The following theorem shows this tradeoff is optimal.

Theorem 5.1 *There is read-once DNF $\phi : \{0,1\}^n \to \{0,1\}$ with $\Theta(n/\log n)$ terms and an ϵ-biased distribution D over $\{0,1\}^n$ where $\epsilon = n^{-\Theta(1)}$ such that*

$$|\Pr_{x\in U_n}[\phi(x)=1] - \Pr_{x\in D}[\phi(x)=1]| = \Omega(1)$$

Proof: Let t be an integer such that $t \equiv 2 (mod\ 4)$ and for $x \in \{0,1\}^t$, define the inner product

$$IP(x) = \left(\sum_{i=1}^{t/2} x_i x_{t/2+i}\right) \quad (mod\ 2)$$

Define distribution D over $\{0,1\}^{t+1}$ as follows. It is a uniform distribution on $x \circ IP(x)$ for $x \in \{0,1\}^t$. The following fact is easy to verify.

Fact 5.2 *For all subsets $S \subset [t]$, $\chi_S : \{0,1\}^t \to \{-1,1\}$,*

$$\left|\mathbb{E}_{x\in U_t}\left[\chi_S(x)(-1)^{IP(x)}\right]\right| = 2^{-t/2}$$

Claim 5.3 *D is $2^{-\Omega(t)}$ biased distribution over $\{0,1\}^{t+1}$.*

Proof: Consider any character $\chi_S : \{0,1\}^{t+1} \to \{0,1\}$. In case, $(t+1) \notin S$, then clearly $\mathbb{E}_{x\in D}[\chi_S(x)] = 0$. If $(t+1) \in S$, then let $S' = S\backslash\{t+1\}$

$$\left|\mathbb{E}_{x\in D}[\chi_S(x)]\right| = \left|\mathbb{E}_{x'\in U_t}[\chi_{S'}(x')(-1)^{IP(x')}]\right| = 2^{-\Omega(t)}$$

This implies that D is $2^{-\Omega(t)}$ biased distribution over $\{0,1\}^{t+1}$. ■

Let $n = (t+1)2^t$ for $t \equiv 2(mod\ 4)$. Split $\{0,1\}^n$ into 2^t chunks. Let the variables in the i^{th} chunk be $y_{i,1}, \ldots, y_{i,t+1}$. Let $D_1, \ldots, D_{2^t}$ be 2^t independent copies of D such that D_i is over $y_{i,1}, \ldots, y_{i,t+1}$. Let D' defined over $\{0,1\}^n$ be the product distribution of $D_1, \ldots, D_{2^t}$. Clearly, D' is a $2^{-\Omega(t)}$ biased distribution. Now, consider the read-once DNF ϕ defined as

$$\phi = \bigvee_{i=1}^{2^t} \left(\bigwedge_{j=1}^{t+1} y_{i,j} \right)$$

Under the uniform distribution, each term is satisfied with probability $1/2^{t+1}$ while note that under D', each term is satisfied with probability $1/2^t$. This is because once the first t variables in a term are 1, the $t+1^{th}$ variable is 1 in D as $t \equiv 2(mod\ 4)$. As the terms are over disjoint sets of variables, hence we can say that

$$\left| \Pr_{y \in D}[\phi(y) = 0] - \Pr_{y \in U}[\phi(y) = 0] \right| = \left| \left(1 - \frac{1}{2^t}\right)^{2^t} - \left(1 - \frac{1}{2^{t+1}}\right)^{2^t} \right| = \Omega(1)$$

This proves the theorem. ■

5.2 Almost tight examples for smaller δ

The obvious scaling of the previous example would give $\epsilon = 2^{-\Omega(\log m + \log\log(1/\delta))}$. Here we give a construction of a specific ϵ-biased distribution which shows that to δ-fool the "tribes" DNF (described below), one must have $\epsilon = m^{-\Omega(\log(1/\delta)/\log\log(1/\delta))}$. We first state the more general form of the theorem claiming the existence of such a DNF and a distribution and as a subsequent corollary, we get the bias in terms of the distinguishing probability.

Theorem 5.4 *For every sufficiently large integer n of the form $n = m \log m$ for m which is power of 2 and for every integer $d \geq 1$, there is an $(m/2)^{-d}$-biased distribution D over $\{0,1\}^n$ and a read-once DNF ϕ with m terms such that ϕ distinguishes D from uniform by at least $1/(2d+3)!$.*

Proof: We first describe the DNF. The DNF is defined by splitting the n variables into m chunks of size $\log m$. Let the variables in the i^{th} chunk be $x_{i,1}, \ldots, x_{i,\log m}$. The DNF is

$$\phi(x) = \bigvee_{i=1}^{m} C_i \quad \text{where } C_i \equiv \bigwedge_{j=1}^{\log m} x_{i,j}$$

The following two claims, describe the required distribution D.

Claim 5.5 *There is a distribution $Y = Y_1 \circ \ldots \circ Y_m$ over $\{0,1\}^m$ with the following properties*

- *for every $1 \leq i \leq m$, $\Pr[Y_i = 1] = 1/m$.*
- *$Y_1, \ldots, Y_m$ are d-wise independent;*
- *For every $y \in Supp(Y)$, $y_1 + \ldots + y_m \leq d$.*

We can now describe the distribution D in terms of the random variables $Y_1, \ldots, Y_m$. Given values $y_1, \ldots, y_m$, we choose $x_{i,1}, \ldots, x_{i,\log m}$ to be all 1, if $y_i = 1$ and uniformly from $\{0,1\}^{\log m} \setminus 1^{\log m}$ if $y_i = 0$. In particular, this ensures that $\bigwedge_{j=1}^{\log m} x_{i,j} = y_i$ and hence C_i is satisfied if and only if $y_i = 1$. We claim that the distribution has a small bias.

Claim 5.6 *The distribution D defined above has bias at most $(m/2)^{-d}$.*

Before proving these two claims, lets see why they suffice to construct the counterexample. First, observe that by Claim 5.6, term C_i being satisfied is equivalent to $y_i = 1$. By inclusion-exclusion principle, the probability that $x \in_r D$ satisfies ϕ is

$$\begin{aligned}
\Pr_{x\in D}[\phi \text{ is satisfied}] &= \sum_{S\in[m],|S|>0} (-1)^{|S|-1} \Pr[\forall i \in S,\ C_i \text{ is satisfied}] \\
&= \sum_{S\in[m],|S|>0} (-1)^{|S|-1} \Pr[\forall i \in S, y_i = 1] \\
&= \sum_{S\in[m],d\geq|S|>0} (-1)^{|S|} \Pr[\forall i \in S,\ y_i = 1] \qquad (\text{Using } \sum_i^m y_i \leq d) \\
&= \sum_{t=1}^{d} (-1)^{t-1} \binom{m}{t} \frac{1}{m^t}
\end{aligned}$$

The last equality uses that y_i's are d-wise independent and $\Pr[y_i = 1] = 1/m$. To estimate the above probability for the uniform distribution, we can obtain upper and lower bounds on it by truncating the inclusion-exclusion respectively at $d+1$ and $d+2$ when d is even (the upper and lower bounds are switched when d is odd). Thus ϕ distinguishes D from uniform with probability at least

$$\begin{aligned}
\binom{m}{d+1}\frac{1}{m^{d+1}} - \binom{m}{d+2}\frac{1}{m^{d+2}} &= \frac{m!}{m^{d+1}(d+1)!(m-d-2)!}\left(\frac{1}{m-d-1} - \frac{1}{m(d+2)}\right) \\
&\geq \frac{m!}{m^{d+1}(d+1)!(m-d-2)!}\frac{1}{2m} \\
&\geq \frac{1}{2(d+1)!}\prod_{i=1}^{d+1}\left(1 - \frac{i}{m}\right) \\
&= \frac{1}{2(2d+2)!}\prod_{i=1}^{d+1}\left((d+1+i)\left(1-\frac{i}{m}\right)\right) \geq \frac{1}{(2d+3)!}
\end{aligned}$$

The last inequality uses that $(d+1+i)(1-i/m) \geq 1$. Hence, we need to prove Claims 5.5 and 5.6. We start with Claim 5.5.

Proof: [of Claim 5.5] Let $p_0, \ldots, p_d \geq 0$ such that $\sum p_i = 1$ (We will non-constructively describe p_i's later). The distribution Y is chosen as following. Pick i, $0 \leq i \leq d$ with probability p_i. Choose a uniformly random subset $S \subset [m]$ of size d and set $y_i = 1$ if $i \in S$ and $y_i = 0$ if $i \notin S$. By construction, trivially the third property is satisfied. We need to set $p_0, \ldots, p_d$ such that the first and the second properties are satisfied. Note that to ensure that Y_i's are d-wise independent, it suffices to show that for every $0 \leq i \leq d$ and $1 \leq j_1 < \ldots < j_i \leq m$, we have $\mathbb{E}[y_{j_1} \cdot \ldots \cdot y_{j_i}] = \mathbb{E}[y_{j_1}] \cdot \ldots \cdot \mathbb{E}[y_{j_i}] = 1/m^i$ (because each variable y_k takes only two possible values.) By symmetry of the construction, it suffices to ensure these properties when $\{j_1, \ldots, j_i\} = \{1, \ldots, i\}$ for every $0 \leq i \leq d$. Thus we only need to select $p_0, \ldots, p_d$ such that for every $0 \leq i \leq d$,

$$\mathbb{E}[y_1 \cdot \ldots \cdot y_i] = \sum_{t=i}^{d} \frac{\binom{m-i}{t-i}}{\binom{m}{t}} p_t = 1/m^i.$$

This is a triangular system of $d+1$ linear equations which has a unique solution $p_0, \ldots, p_d$. However, we must make sure that the values of the solution $p_0, \ldots, p_d$ are nonnegative. We use descent on i to show $p_i \geq 0$. We have $p_d = \binom{m}{d}/m^d \geq 0$. For $i < d$, we have:

$$\begin{aligned}
p_i &= \binom{m}{i}\left[\frac{1}{m^i} - \sum_{t=i+1}^{d} \frac{\binom{m-i}{t-i}}{\binom{m}{t}} p_t\right] \\
&\geq \binom{m}{i}\left[\frac{1}{m^i} - \sum_{t=i+1}^{d} \frac{\binom{m-i-1}{t-i-1}}{\binom{m}{t}} m p_t\right] \\
&= m\binom{m}{i}\left[\frac{1}{m^{i+1}} - \sum_{t=i+1}^{d} \frac{\binom{m-i-1}{t-i-1}}{\binom{m}{t}} p_t\right] = 0
\end{aligned}$$

■

We also give a constructive proof of the above claim in the appendix. However, we choose to retain this argument as the technique used to justify existence of the distribution is more general.

Proof: [of Claim 5.6] To compute the bias of the distribution D, consider any character χ_S where $S \subset [m \log m]$ is non-empty. For any $i \in [m]$, let us define $S_i = S \cap \{(i-1)\log m + 1, \ldots, i \log m\}$. Note that

$$\mathop{\mathbb{E}}_{x \in D}[\chi_S(x)] = \mathop{\mathbb{E}}_{x \in D}\left[\prod_{i: S_i \neq \phi} \chi_{S_i}(x)\right]$$

Our proof will only depend on the number of non-empty sets S_i. Without loss of generality, we can assume that the non-empty sets are $S_1, \ldots, S_t$ for some $t > 0$. We denote the set of variables $x_{i,1}, \ldots, x_{i,\log m}$ by x_i. To compute the bias, we then need to calculate

$$\mathbb{E}_{x \in D}\left[\prod_{i=1}^{t} \chi_{S_i}(x_i)\right] = \mathbb{E}_Y\left[\prod_{i=1}^{t} \mathbb{E}_{x_i}\left[\chi_{S_i}(x_i) | y_i\right]\right]$$

as the variables $x_1, \ldots x_m$ are independent given Y. We now note that

$$\mathbb{E}_{x_i}\left[\chi_{S_i}(x_i) | y_i = 1\right] = (-1)^{|S_i|} \qquad \text{and} \qquad \mathbb{E}_{x_i}\left[\chi_{S_i}(x_i) | y_i = 0\right] = -\frac{(-1)^{|S_i|}}{m-1}$$

If $t \leq d$, then $y_1, \ldots, y_t$ are independent and the bias simply becomes 0 as below.

$$\begin{aligned}
\mathbb{E}_Y\left[\prod_{i=1}^{t} \mathbb{E}_{x_i}\left[\chi_{S_i}(x_i) | y_i\right]\right] &= \prod_{i=1}^{t} \mathbb{E}_{x_i, y_i}\left[\chi_{S_i}(x_i)\right] \\
&= \prod_{i=1}^{t}\left(\frac{1}{m} \cdot (-1)^{|S_i|} - \left(1 - \frac{1}{m}\right) \cdot \frac{(-1)^{|S_i|}}{m-1}\right) = 0
\end{aligned}$$

If $t > d$, we can bound the bias as

$$\begin{aligned}\mathbb{E}_Y\left[\prod_{i=1}^{t}\mathbb{E}_{x_i}\left[\chi_{S_i}(x_i)|y_i\right]\right] &\leq \mathbb{E}_Y\left[\prod_{i=1}^{t}\left|\mathbb{E}_{x_i}\left[\chi_{S_i}(x_i)|y_i\right]\right|\right]\\ &\leq \mathbb{E}_Y\left[\prod_{i=1}^{d}\left|\mathbb{E}_{x_i}\left[\chi_{S_i}(x_i)|y_i\right]\right|\right]\\ &= \prod_{i=1}^{d}\left(\frac{1}{m}+\left(1-\frac{1}{m}\right)\cdot\frac{1}{m-1}\right) = \left(\frac{2}{m}\right)^d\end{aligned}$$

which proves the claim. ■

■

By plugging $d = \log(1/\delta)/\log\log(1/\delta)$ in the above theorem, we get the following corollary.

Corollary 5.7 *For m which is a power of 2 and $\delta > 0$, there is a read-once DNF ϕ over $n = m\log m$ variables and a distribution D over $\{0,1\}^n$ which has bias $m^{-O(\log(1/\delta)/\log\log(1/\delta))}$ and ϕ distinguishes D from uniform by δ.*

5.3 Lower bounds for fooling general DNFs

Below we show that to δ-fool general DNFs with m terms, one requires a $m^{-\Omega(\log 1/\delta)}$ biased set. Before, we state the theorem, we state the following technical lemma.

Lemma 5.8 *For $x \in \{0,1\}^n$, let $MOD_3(x) = \sum_{i=1}^{n} x_i \pmod 3$. Consider the distribution D over $\{0,1\}^n$ which is the uniform distribution on the set D_0 defined as*

$$D_0 = \{x | MOD_3(x) \neq 0\}$$

Then D is $2^{-\Omega(n)}$ biased distribution.

Proof: Consider any linear function $\chi : \{0,1\}^n \to \{-1,1\}$. Lemma 2.9 in [VW08] says that

$$|\Pr_{x:MOD_3(x)=0}[\chi(x)=1] - \Pr_{x:MOD_3(x)\neq 0}[\chi(x)=1]| = 2^{-\Omega(n)}$$

Also, $|x : \chi(x)=1| = \left(\Pr_{x:MOD_3(x)\neq 0}[\chi(x)=1]\right)|D_0| + \left(\Pr_{x:MOD_3(x)=0}[\chi(x)=1]\right)(2^n - |D_0|)$

$$\implies |x : \chi(x)=1| \geq \left(\Pr_{x:MOD_3(x)\neq 0}[\chi(x)=1]\right)|D_0| + \left(\Pr_{x:MOD_3(x)\neq 0}[\chi(x)=1] - 2^{-\Omega(n)}\right)(2^n - |D_0|)$$

$$\implies \frac{|x : \chi(x)=1|}{2^n} + \frac{2^{-\Omega(n)}(2^n - |D_0|)}{2^n} \geq \left(\Pr_{x:MOD_3(x)\neq 0}[\chi(x)=1]\right)$$

$$\implies \frac{1}{2} + 2^{-\Omega(n)} \geq \left(\Pr_{x:MOD_3(x)\neq 0}[\chi(x)=1]\right)$$

Similarly, we can prove that

$$\frac{1}{2} - 2^{-\Omega(n)} \leq \left(\Pr_{x:MOD_3(x)\neq 0}[\chi(x)=1]\right)$$

This implies that $|\mathbb{E}_{x\in D}[\chi(x)]| = 2^{-\Omega(n)}$ which implies that D is a $2^{-\Omega(n)}$ biased set. ■

We now prove the existence of small biased sets which are distinguished by DNFs. The bound on the bias in terms of number of terms and distinguishing probability is in the subsequent corollary.

Theorem 5.9 *For any $t \geq 3, \ell$, there exists a DNF ϕ over ℓt variables with $O(t2^{\ell})$ terms and an $\epsilon = 2^{-\Omega(\ell t)}$-biased distribution D such that ϕ distinguishes D from uniform with probability $2^{-O(t)}$.*

Proof: The distribution D will be the uniform distribution over $D_0 \subset \{0,1\}^{\ell t}$ which is defined as

$$D_0 = \left\{ x \in \{0,1\}^{\ell t} | MOD_3(x) \neq 0 \right\}$$

By Lemma 5.8, the bias of D_0 is $2^{-\Omega(\ell t)}$. To define the DNF ϕ, we partition the variables into t blocks, each block having ℓ variables. The j^{th} variable in the i^{th} block is denoted by x_{ij}. The DNF ϕ is defined as $\phi_1 \vee \ldots \vee \phi_t$ where ϕ_i is a DNF over the i^{th} block of variables which is 1 if and only if the sum of the variables in the i^{th} block is non-zero modulo 3. Note that ϕ is only a function of variables in the i^{th} block. Thus, we can always write ϕ_i using 2^{ℓ} terms. Hence, ϕ can be written using $t2^{\ell}$ terms. We first observe that

$$\Pr_{x \in D}[\phi(x) = 1] = 1$$

This is because if the sum of the variables in all the blocks is non-zero mod 3, then there must be at least one block i in which the sum is non-zero mod 3 which ensures that $\phi_i = 1$ implying $\phi = 1$. Now, note that under the uniform distribution, each $\phi_i = 0$ with probability at least $1/3 - 2^{-\ell} \geq 1/4$. This is because $\phi_i = 1$ iff $\sum_{j=1}^{\ell} x_{ij} \neq 0 (mod\ 3)$. As all ϕ_i's are over disjoint sets of variables, this implies

$$\Pr_{x \in U}[\phi(x) = 1] = 1 - \Pr_{x \in U}[\phi(x) = 0] = 1 - (\wedge_{i=1}^{t} \Pr_{x \in U}[\phi_i(x) = 0]) \leq 1 - \frac{1}{4^t}$$

This implies that ϕ distinguishes D from uniform by $1/4^t = 2^{-O(t)}$. ■

Corollary 5.10 *For arbitrarily large m and arbitrary small δ such that $2^{-m/2} < \delta$, there exists a DNF ϕ over $O(\log m \log(1/\delta))$ variables and a distribution D such that ϕ has m terms, D has bias $m^{-\Omega(\log(1/\delta))}$ and ϕ distinguishes D from uniform with probability δ.*

Proof: From the above theorem, we can say that for every t, ℓ there is a DNF ϕ and a distribution D such that ϕ has $t2^{\ell}$ terms, D is $2^{-\Omega(t\ell)}$ biased and ϕ can distinguish D from uniform by $2^{-O(t)}$. By setting $t = \Theta(\log(1/\delta))$, we can get the distinguishing probability to be equal to δ. Similarly, we set $\ell = \log m - \log\log(1/\delta) - \Theta(1)$, we can get the number of terms to be m. Then the bias of the distribution D guaranteed by the theorem is $2^{-\Omega(t\ell)} = 2^{-\Omega((\log m - \log\log(1/\delta) - \Theta(1))\log(1/\delta)} = m^{-\Omega(\log(1/\delta))}$ as long as $\delta > 2^{-m/2}$. ■

Acknowledgements

We thank Salil Vadhan for suggesting a proof of Lemma 5.8.

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992. 4

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88(3):107–110, 2003. 1

[AW89] Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constand-depth circuits. *Advances in Computing Research - Randomness and Computation*, 5:199–223, 1989. Preliminary version in *Proc. of FOCS'85*. 1

[Baz03] Louay Bazzi. *Minimum Distance of Error Correcting Codes versus Encoding Complexity, Symmetry, and Pseudorandomness*. PhD thesis, MIT, 2003. 2

[Baz07] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 63–73, 2007. 1, 2, 4, 8, 9

[Bra09] Mark Braverman. Poly-logarithmic independence fools AC0 circuits. Technical Report TR09-011, Electronic Colloquium on Computational Complexity, 2009. 1

[EGL$^+$92] Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Approximations of general independent distributions. In *Proceedings of the 24th ACM Symposium on Theory of Computing*, pages 10–16, 1992. 3, 6

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th ACM Symposium on Theory of Computing*, pages 6–20, 1986. 1

[HVV04] Alexander Healy, Salil Vadhan, and Emanuele Viola. Using nondeterminism to amplify hardness. In *Proceedings of the 36th ACM Symposium on Theory of Computing*, pages 192–201, 2004. 1, 2

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

[LN90] Nathan Linial and Noam Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990. 1

[LV96] Michael Luby and Boban Velickovic. On deterministic approximation of DNF. *Algorithmica*, 16(4/5):415– 433, 1996. 1, 2, 3

[LVW93] Michael Luby, Boban Velickovic, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the 2nd ISTCS*, pages 18–24, 1993. 1, 2

[Man95] Yishay Mansour. An $o(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50(3):543–550, 1995. 3, 8, 10

[Nis91] N. Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 12(4):63–70, 1991. 1, 2

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993. 1, 4

[OD07] Ryan ODonnell. Lecture notes on analysis of boolean functions. Available at http://www.cs.cmu.edu/ ∼odonnell/boolean-analysis, 2007. 10

[Raz09] Alexander Razborov. A simple proof of bazzi's theorem. *ACM Trans. Comput. Theory*, 1(1):1–5, 2009. 1, 2, 7, 9

[VW08] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008. 15

A A constructive proof of Claim 5.5

We give below an alternate proof of Claim 5.5 which gives an explicit construction for the d-wise independent distribution mentioned in the claim.

Claim A.1 *There is a distribution $Y = Y_1 \circ \ldots \circ Y_m$ over $\{0,1\}^m$ with the following properties*

- *for every $1 \leq i \leq m$, $\Pr[Y_i = 1] = 1/m$.*
- *$Y_1, \ldots, Y_m$ are d-wise independent;*
- *For every $y \in Supp(Y)$, $y_1 + \ldots + y_m \leq d$.*

Proof: Since m is taken to be a power of 2, there exists a field $\mathbb{F}$ with $|\mathbb{F}| = m$, the elements of which we identify with the numbers $0, \ldots, m-1$. Choose d independent random elements $a_0, \ldots, a_{d-1} \in \mathbb{F}$ and define the (random) degree-d polynomial

$$P(z) \; := \; z^d + a_{d-1} z^d + \ldots + a_0.$$

We define the random variables $Y_1, \ldots, Y_m$ as

$$Y_i \; := \; \begin{cases} 1 & \text{if } P(i-1) = 0 \\ 0 & \text{otherwise} \end{cases}$$

Since P is a degree d-polynomial, for any $y_1, \ldots, y_m \in Supp(Y)$, at most d of $y_1, \ldots, y_m$ are 1 and hence $y_1 + \ldots + y_m \leq d$. Also, since P is equally likely to take any of the m values at the point $i-1$ (as a_0 is uniform in $\mathbb{F}$), $\Pr[Y_i = 1] = \Pr[P(i-1) = 0] = 1/m$.

Note that for any d distinct points $i_1, \ldots, i_d$ and the polynomial P as above, the vector $(P(i_1), \ldots, P(i_d))$ can be computed as

$$(P(i_1), \ldots, P(i_d)) \; = \; (a_0, \ldots, a_{d-1}) \cdot A + (i_1^d, \ldots, i_d^d)$$

where $A \in \mathbb{F}^{d \times d}$ is a matrix with the j^{th} column as $(1, i_j, \ldots, i_j^{d-1})^\mathsf{T}$. Since all the columns of A are linearly independent, and $(a_0, \ldots, a_{d-1})$ is a random element of $\mathbb{F}^d$, $(P(i_1), \ldots, P(i_d))$ is also uniformly distributed in $\mathbb{F}^d$. This gives that the values of P at any d points are independent and hence $Y_1, \ldots, Y_m$ form the required d-wise independent distribution. ■