Electronic Colloquium on Computational Complexity, Report No. 34 (2013)

Small-bias is not enough to hit read-once CNF *

Louay Bazzi † Nagi Nahas ‡

March 2, 2013

Abstract

Small-bias probability spaces have wide applications in pseudorandomness which naturally leads to the study of their limitations. Constructing a polynomial complexity hitting set for read-once CNF formulas is a basic open problem in pseudorandomness. We show in this paper that this goal is not achievable using small-bias spaces. Namely, we show that for each read-once CNF formula F with probability of acceptance p and with m clauses each of size c, there exists a δ-biased distribution μ on $\{0,1\}^n$ such that $\delta = 2^{-\Omega(\log m \log(1/p))}$ and no element in the support of μ satisfies F, where $n = mc$ (assuming that $2^{-m^{0.3}} \leq p \leq p_0$, where $p_0 > 0$ is an absolute constant). In particular if $p = n^{-\Theta(1)}$, the needed bias is $2^{-\Omega(\log^2 n)}$, which requires a hitting set of size $2^{\Omega(\log^2 n)}$. Our lower bound on the needed bias is asymptotically tight. The dual version of our result asserts that if $f_{low} : \{0,1\}^n \to \mathbb{R}$ is such that and $E[f_{low}] > 0$ and $f_{low}(x) \leq 0$ for each $x \in \{0,1\}^n$ such that $F(x) = 0$, then the $L1$-norm of the Fourier transform of f_{low} is at least $E[f_{low}]2^{\Omega(\log m \log(1/p))}$. Our result extends a result due to De, Etesami, Trevisan, and Tulsiani (APPROX-RANDOM 2010) who proved that the small-bias property is not enough to obtain a polynomial complexity PRG for a family of read-once formulas of $\Theta(1)$ probability of acceptance.

*Research supported by FEA URB grant Program Number 288309, American University of Beirut.

†Department of Electrical and Computer Engineering, American University of Beirut, Beirut, Lebanon. E-mail: lb13@aub.edu.lb.

‡Department of Electrical and Computer Engineering, American University of Beirut, Beirut, Lebanon. E-mail: nhn01@aub.edu.lb.

ISSN 1433-8092

1 Introduction

A *pseudorandom number generator (PRG)* $G : \{0,1\}^r \to \{0,1\}^n$ maps uniformly random *seeds* to longer pseudorandom strings. An α-PRG G for a class $\mathcal{C}$ of boolean functions from $\{0,1\}^n$ to $\{0,1\}$ is a PRG such that the probability distribution supported by $G(\{0,1\}^r)$ α-fools [1] all function in $\mathcal{C}$. Hitting sets are weaker than PRGs but they capture the PRG problem when $Pr[f = 1]$ is small. An *ϵ-hitting set* [ACR98, SZ11] for a class $\mathcal{C}$ of boolean functions is subset $H \subset \{0,1\}^n$ such that for every function $f \in \mathcal{C}$ with $Pr[f = 1] > \epsilon$, there exists $y \in H$ such that $f(y) = 1$. Hence if G is an ϵ-PRG for $\mathcal{C}$, then $H = G(\{0,1\}^r)$ is an ϵ-hitting set for $\mathcal{C}$. More generally, if μ is a probability distribution on $\{0,1\}^n$, we say that μ is an *ϵ-hitting distribution* for $\mathcal{C}$ if for every function $f \in \mathcal{C}$ such that $Pr[f = 1] > \epsilon$, there exists $y \in Support(\mu)$ such that $f(y) = 1$.

An ultimate goal of pseudorandomness is to construct an $n^{-O(1)}$-PRG (computable in uniform polynomial time) of logarithmic seed length for all polynomial size circuits. Without hardness assumptions [NW88, IW97], the problem is still open even for simple families of circuits such as polynomial-size depth-d circuits constructed from AND, OR, and NOT gates, for $d \geq 2$. The $d = 2$ case corresponds to DNF formulas (OR of AND gates) and CNF formulas (AND of OR gates). The simpler problem of constructing polynomial-size $n^{-O(1)}$-hitting sets (computable in uniform polynomial time) for depth-2 circuits is also an open problem, even if we assume that the depth-2 circuit is read-once. It is not hard to show that any ϵ/m-biased probability distribution [2] on $\{0,1\}^n$ is an ϵ-hitting distribution for DNF formulas consisting of m clauses (this follows from a union bound and the fact that any ϵ-biased distribution is an ϵ-hitting distribution for AND gates). The CNF case is different due to the lack of symmetry between 0 and 1 in the hitting set definition. Constructing a hitting set for read-once CNF formulas captures a key difficulty of the problem of constructing a PRG for depth-2 circuits.

1.1 Fooling depth-2 circuits

Nisan constructed [Nis91] a PRG for polynomial-size constant-depth circuits of polylogarithmic seed-length. In the special case of depth-2 circuits with m clauses and n variables, the seed length is $O(log^{10}(mn/\alpha))$. By optimizing Nisan's generator, Luby, Velickovic, and Wigderson [LVW93] achieved $O(log^4(mn/\alpha))$ seed length. By resolving a conjecture due to Linial and Nisan [LN90], [Baz07, Baz09] proved that any t-wise independent distribution [3] α-fools depth-2 circuits, where $t = O(\log^2(m/\alpha))$, which gives a PRG of seed length $O(\log n \log^2(m/\alpha))$. The proof was later simplified by Razborov [Raz09] and the result was extended by Braverman [Bra10] to constant-depth circuits of depth greater than 2. By building on [Baz07, Baz09, Raz09], De *et al.* [DETT10] proved that any δ-biased distribution α-fools depth-2 circuits, where $\delta = 2^{-O(\log^2(m/\alpha)\log\log(m/\alpha))}$, which reduced the seed length to $O(\log n + \log^2(m/\alpha)\log\log(m/\alpha))$. To date, this is the best known seed-length for general depth-2 circuits. For constant-depth circuits of depth $d \geq 3$, the best known seed-length is $\tilde{O}(\log^{d+4}(M/\alpha))$ due to Trevisan and Xue [TX12], where M is the circuit size.

[1] If μ is a probability distribution on $\{0,1\}^n$ and $f : \{0,1\}^n \to \{0,1\}$ is a boolean function, we say that μ *α-fools* f [BM82, Yao82] if $|Pr_\mu[f = 1] - Pr[f = 1]| \leq \alpha$, where the second probability is with respect to the uniform probability distribution on $\{0,1\}^n$.

[2] A probability distribution μ on $\{0,1\}^n$ is called *δ-biased* [NN93] if μ $\delta/2$-fools all parity functions on the n binary variables. See Section 2 for an equivalent character-based definition.

[3] We say that μ is *t-wise independent* (e.g., [Lub85, Vaz86]) if any t or less of the underlying n binary random variables are statistically independent and each is equally likely to be zero or one.

1.2 The read-once depth-2 case

Better bounds are known for the special case of read-once depth-2 circuits. The best currently known seed-length is $O(\log(m)\log\log(m)+\log(1/\alpha))$ [GMRTV12]. The underlying PRG uses an iterative construction involving small-bias spaces. Earlier results were based entirely on the limited-independence or small-bias properties. Using inclusion-exclusion, it was shown in [Baz03] that any t-wise independent distribution α-fools read-once depth-2 circuits, where $t = O(\log(m/\alpha)\log(1/\alpha))$, which gives $O(\log n \log(m/\alpha)\log(1/\alpha))$ seed length. De *et al.* [DETT10] improved the seed length to $O(\log n + \log m \log(1/\alpha))$ by analyzing the inclusion exclusion construction in the context of small bias spaces. Namely, they proved that any δ-biased distribution α-fools read-once depth-2 circuits, where $\delta = 2^{-O(\log m \log(1/\alpha))}$. Accordingly, we have the following in the context of hitting sets.

Theorem 1.1 *[DETT10] If $\delta = 2^{-O(\log m \log(1/\epsilon))}$, then any δ-biased distribution on $\{0,1\}^n$ is an ϵ-hitting distribution for read-once CNF formulas on n variables with m clauses.*

1.3 Limitations of small bias

The bias of a subset $S \subset \{0,1\}^n$ (i.e., the bias of the probability distribution supported by S) is polynomially related to the size of S. Namely, [NN93, AGHP92] show how to construct δ-biased probability spaces of support size $\text{poly}(n/\delta)$. On the other hand, Alon *et. al* [AGHP92] proved that a δ-biased subset of $\{0,1\}^n$ has size $\Omega(\min(\frac{n}{\delta^2 log(1/\delta)}, 2^n))$.

The wide applicability of small-bias spaces in pseudorandomness naturally leads to the study of their limitations.

(A) Is small bias sufficient to construct polynomial-complexity $o(1)$-PRGs for general CNF formulas?

(B) Is it sufficient for the weaker task of constructing polynomial-complexity $o(1)$-hitting sets for general CNF formulas?

(C) Is it sufficient to construct polynomial-complexity $o(1)$-PRGs for the weaker class of read-once CNF formulas?

(D) Is it sufficient for the weaker task of constructing polynomial-complexity $o(1)$-hitting sets for read-once CNF formulas?

De *et. al* [DETT10] answered Questions A, B, and C in the negative.

Theorem 1.2 *[DETT10] Assume that $2^{-m/2} < p < p_0$, where p_0 is absolute constant p_0. Then there exist a (read-many) CNF formula F on n variables with m clauses and probability of acceptance p, and a δ-biased distribution μ on $\{0,1\}^n$ such that $\delta = 2^{-\Omega(\log m \log(1/p))}$ and no element in the support of μ satisfies F.*

In the read-once case, they established a weaker negative result. They proved that the small bias property is not sufficient to obtain a polynomial complexity $o(1)$-PRG for a class of read-once CNF formulas whose probability of acceptance is $\Theta(1)$.

Theorem 1.3 *[DETT10] There exists a read-once CNF formula F on n variables with m clauses and probability of acceptance $p = \Theta(1)$, and a δ-biased distribution μ on $\{0,1\}^n$ such that $\delta = 2^{-\Omega(\log m \log(1/\alpha)/\log\log(1/\alpha))}$ and μ does not α-fool F.*

1.4 Contribution

We answer Question D in the negative. We derive a lower bound on bias needed to hit read-once CNFs which asymptotically meets inclusion-exclusion upper bound in Theorem 1.1, and accordingly is asymptotically tight.

Our approach is different from that of De *et. al.* While their approach is based on directly analyzing the properties of an explicit distribution, our proof is nonconstructive and it is based on approaching the problem as an optimization problem. The result of De *et. al* in Theorem 1.3 rules out the possibility that the task of constructing an $o(1)$-PRG for read-once CNF formulas is achievable using small bias, but since the probability of acceptance p is constant, it does not rule out the possibility that the weaker task of constructing an $o(1)$-hitting set for read-once CNF formulas is achievable using small bias. We show that this is not possible. It is worth noting that the problem of hitting a read-once CNF F of probability of acceptance p is easy for $p = \Theta(1)$: by inclusion-exclusion (Theorem 1.1), any δ-biased distribution contains a satisfying assignment of F provided that $\delta = 2^{-O(\log m \log(1/p))} = n^{O(1)}$. Hence, in the context of hitting sets, the interesting case corresponds to $o(1)$ values of the probability of acceptance.

We argue that to guarantee that any δ-biased distribution is an ϵ-hitting distribution for read-once CNF formulas, we need $\delta = 2^{-\Omega(\log m \log(1/\epsilon))}$.

Theorem 1.4 *Let m and c be positive integers and F a read-once CNF formulas with m clauses of c variables each. Let $n = mc$ and p be the probability of acceptance of F. Assume that $2^{-m^{0.3}} \leq p \leq p_0$, where $p_0 > 0$ is an absolute constant.*

Then there exists a δ-biased distribution μ on $\{0,1\}^n$ such that $\delta = 2^{-\Omega(\log m \log(1/p))}$ and no element in the support of μ satisfies F.

Asymptotically, our lower bound is tight since it meets the inclusion-exclusion upper bound of De *et. al* in Theorem 1.1 (assuming $2^{-m^{0.3}} \leq \epsilon \leq \epsilon_0$, where $\epsilon_0 > 0$ is an absolute constant). Moreover, it extends the negative result of De *et. al* in Theorem 1.2 from read-many to read-once CNF formulas.

The special case when $p = n^{-\Theta(1)}$ is of particular interest.

Corollary 1.5 *There exists a read-once CNF formula F on n variables with probability of acceptance $n^{-\Theta(1)}$ and a probability distribution μ on $\{0,1\}^n$ with bias $2^{-\Omega(\log^2 n)}$ such that no element in the support of μ satisfies F.*

Hence for probability of acceptance $p = n^{-\Theta(1)}$, the bound $\delta = 2^{-\Omega(\log^2 n)}$ requires a hitting set of size $2^{\Omega(\log^2 n)}$ [AGHP92].

The LP dual of Theorem 1.4 is a statement about the nonseparability of the boolean functions computed by a read-once CNF formulas by real-valued functions with low $L1$-norm in the Fourier domain.

Corollary 1.6 *Let F be a read-once CNF formula with m, c, n, and p as defined in Theorem 1.4. If $f_{low} : \{0,1\}^n \to \mathbb{R}$ is such that (i) $E_U f_{low} > 0$, where U is the uniform distribution on $\{0,1\}^n$, and (ii) $f_{low}(x) \leq 0$ for each $x \in \{0,1\}^n$ such that $F(x) = 0$, then $\|\hat{f}_{low}\|_1 \geq (E_U f_{low}) 2^{\Omega(\log m \log(1/p))}$.*

We extract form the proof of Theorem 1.4 the the following key technical lemma, which exhibits a non-hitting distribution for F with the property that all variables appearing in $O(\log \frac{1}{p})$ clauses of F are independent.

Theorem 1.7 *Let m and c be positive integers and F a read-once CNF formulas with m clauses of c variables each. Let $n = mc$ and p be the probability of acceptance of F. Assume that $2^{-o(\sqrt{m})} < p \leq p_0$, where $p_0 > 0$ is an absolute constant.*

Then there exists a probability distribution μ on $\{0,1\}^n$ such that:

*a) (*Limited independence with respect to the Formula*) For any set S of variables which appear in no more than $d_0 = \Theta(\log \frac{1}{p})$ clauses, the variables in S are statistically independent with respect to μ and each is equally likely to be 0 or 1.*

b) No element in the support of μ satisfies F.

The dual statement is the following.

Corollary 1.8 *Let F be a read-once CNF formula with $m, c, n,$ and p as defined in Theorem 1.7. If $p_{low}(x_1, \ldots, x_n) \in \mathbb{R}[x_1, \ldots, x_n]$ is a polynomial such that (i) $E_U p_{low} > 0$, where U is the uniform distribution on $\{0,1\}^n$, and (ii) $p_{low}(x) \leq 0$ for each $x \in \{0,1\}^n$ such that $F(x) = 0$, then p_{low} has a monomial whose variables appear in $\Omega(\log \frac{1}{p})$ clauses.*

1.5 Technique

After providing the basic definitions in Section 2, we outline the proof in Section 3. At a high level, our approach is the following. We formulate the problem of minimizing the bias of a distribution whose support does not contain any satisfying input of the formula as a linear program. We reduce the complexity of the LP by restricting our attention to distributions which are symmetric with respect to the formula in the sense that any two elements of $\{0,1\}^n$ that satisfy the same number of clauses have the same probability. Krawtchouk polynomials originate in the proof when studying the bias of such distribution. We reduce the problem of bounding the bias to that of showing that a certain polytope of distributions is not empty. The polytope consists of symmetric distributions which do not hit the formula and have the property that all variables appearing in at most a certain number of clauses are independent. By studying the dual, we reduce the problem to a minimax like problem on low degree univariate polynomials, which we solve using Chebyshev's alternating signs technique.

2 Basic definitions

2.1 CNF Formulas

A CNF formula is an AND of OR gates called *clauses*. It is called monotone if no variable is negated, and read-once if each variable appears in exactly one clause. Let the positive integers m, c, and $n = mc$ denote, respectively, the number of clauses, the clause size, and the number of variables. We study in this paper monotone read-once CNF formulas, where all the clauses have equal length. It is convenient to represent such a formula using a partition of $[n] \coloneqq \{1, \ldots, n\}$ which divides the variables according to the clauses they belong to. A *c-regular m-partition* π of $[n]$ is a collection $\{\pi(j)\}_{j=1}^m$ of subsets of $[n]$ such that $|\pi(j)| = c$, for all j, $\bigcup_{j=1}^m \pi(j) = [n]$, and $\pi(j) \cap \pi(j') = \emptyset$, for all $j \neq j'$. Given π, define the corresponding *CNF* formula $F_\pi : \{0,1\}^n \to \{0,1\}$ as $F_\pi(x) \coloneqq \bigwedge_{j=1}^m OR_{\pi(j)}(x)$, where $OR_{\pi(j)}(x) \coloneqq \bigvee_{i \in \pi(j)} x_j$. We say that clause $\pi(j)$ is satisfied by x if $OR_{\pi(j)}(x) = 1$. Thus $F_\pi(x) = 1$ iff all the m clauses are satisfied by x.

Define the *π-weight* function $W_\pi : \{0,1\}^n \to \mathbb{N}$ as $W_\pi(x) \coloneqq |\{j \in [m] : OR_{\pi(j)}(x) = 1\}|$, i.e., $W_\pi(x)$ is the number of clauses satisfied by x. We will call $W_\pi(x)$ the π-weight of x. Hence $F_\pi(x) = 1$ iff $W_\pi(x) = m$.

In the entire paper, we fix the definitions of m,$c \geq 1$, the partition π, and accordingly $n = mc$, F_π, and W_π.

2.2 Distributions

2.2.1 Bias

The bias of a probability distribution on $\{0,1\}^n$ measures how far the distribution is from the uniform distribution with respect to linear tests modulo 2 [NN93]. Those tests are captured by the characters $\{\mathcal{X}_z\}_{z \in \{0,1\}^n}$ of the abelian group $(\mathbb{Z}/2\mathbb{Z})^n$, where $\mathcal{X}_z : \{0,1\}^n \to \{0,1\}^n$ is given by $\mathcal{X}_z(x) = (-1)^{\langle x,z \rangle}$, where $\langle x, z \rangle = \sum_{i=1}^n x_i z_i$. They are orthogonal in the sense that $\sum_{x \in \{0,1\}^n} \mathcal{X}_z(x)\mathcal{X}_{z'}(x) = 0$ for all $z \neq z'$, hence in particular $\sum_{x \in \{0,1\}^n} \mathcal{X}_z(x) = 0$ for each $z \neq 0 \in \{0,1\}^n$. Let μ be a probability distribution on $\{0,1\}^n$. If $z \in \{0,1\}^n$, the *bias* of μ

at z is given by $bias_z(\mu) \coloneqq E_{x\sim\mu}\mathcal{X}_z(x) = \sum_{x\in\{0,1\}^n}\mu(x)\mathcal{X}_z(x)$. The bias of μ is $bias(\mu) \coloneqq \max_{z\neq 0\in\{0,1\}^n}|bias_z(\mu)|$. We call μ δ-*biased* if $bias(\mu)\leq\delta$

2.2.2 Limited independence with respect to clauses

We call a probability distribution μ on $\{0,1\}^n$ (d,π)-*wise independent* if for any set S of variables which appear in no more than d clauses, the variables in S are statistically independent and each is equally likely to be 0 or 1. In terms of bias, μ is (d,π)-wise independent, if $bias_z(\mu)=0$ for each $z\neq 0\in\{0,1\}^n$ such that $W_\pi(z)\leq d$.

2.2.3 Symmetric distributions

We call a probability distribution μ on $\{0,1\}^n$ π-*symmetric* if the probability $\mu(x)$ of x depends only on the π-weight of x (i.e., the number of clauses satisfied by x). That is, $\mu(x)=\mu(y)$, for all $x,y\in\{0,1\}^n$ such that $W_\pi(x)=W_\pi(y)$.

2.2.4 Induced distributions

If μ is a probability distribution on $\{0,1\}^n$, let $\underline{\mu}$ be the probability distribution *induced on* $[0:m]\coloneqq\{0,1,\ldots,m\}$ via π, i.e., $\underline{\mu}(w)\coloneqq\mu(x:W_\pi(x)=w)$ for $w=0,1,\ldots,m$. That is, $\underline{\mu}(w)$ is the probability with respect to μ that a random vector satisfies exactly w clauses of the CNF formula.

Hence if μ is π-symmetric, we have $\underline{\mu}(w)=\binom{m}{w}(2^c-1)^w\mu(x)$ for any $x\in\{0,1\}^n$ such that $W_\pi(x)=w$.

Finally, we denote by $\mu(F_\pi=1)$ the probability that F_π is 1 with respect to μ, i.e., $\mu(F_\pi=1)\coloneqq\mu(x\in\{0,1\}^n:F_\pi(x)=1)=\underline{\mu}(m)$.

2.2.5 Binomial distribution

For a uniformly random $x\in\{0,1\}^n$, the number $W_\pi(x)$ of satisfied clauses obeys the binomial distribution. We denote this *binomial distribution* by $bin_{m,c}\coloneqq\underline{\mu_{unif}}$, i.e., $bin_{m,c}$ is the distribution induced on $[0:m]$ by the uniform probability distribution μ_{unif} on $\{0,1\}^n$. Hence $bin_{m,c}(w)=\frac{1}{2^{cm}}\binom{m}{w}(2^c-1)^w$. Its mean is the expected number of satisfied clauses $E_{w\sim bin_{m,c}}w=m-m/2^c$.

Finally, note that the probability p that $F_\pi(x)=1$ over the uniform distribution of $x\in\{0,1\}^n$ is $(1-2^{-c})^m\approx e^{-m/2^c}$. Hence the term $m/2^c=\Theta(\log(1/p))$. Of particular interest to us is the case when $m/2^c=\Theta(\log n)$, which corresponds to $p=n^{-\Theta(1)}$.

3 Proof outline

In this section, we overview the proof and specialize it to the simple case when $m/2^c=c$, hence $m=c2^c$, $n=c^22^c$, $c=\log m-\log\Theta(\log m)$[4], and the probability of acceptance of the CNF is $(1-2^{-c})^m\approx e^{-c}=m^{-\Theta(1)}$.

Consider the optimization problem of estimating the smallest possible bias a distribution has while still not having in its support any satisfying input of the CNF formula. Define $\beta_{m,c}$ to be the minimum δ such that there exist a δ-biased probability distribution μ on $\{0,1\}^n$ such that $\mu(F_\pi=1)=0$. Recall that throughout the paper π is a c-regular m-partition of $[n]$. Our objective is to establish an upper bound on $\beta_{m,c}$. Due to the symmetry of the optimization

[4]Throughout the paper, log means $\log_2$.

problem, $\beta_{m,c}$ does not depend on the partition π, it depends only on m and c. Thus we have a linear program on the real variables $\{\mu(x)\}_{x\in\{0,1\}^n}$ and δ:

$$\begin{aligned}\beta_{m,c} \;\coloneqq\; & \min \delta \text{ subject to :} && (1)\\ & \mu(x) \geq 0 && \forall x \in \{0,1\}^n\\ & \mu(x) = 0 && \forall x \in \{0,1\}^n \text{ such that } F_\pi(x) = 1\\ & \textstyle\sum_{x\in\{0,1\}^n} \mu(x) = 1 &&\\ & -\delta \leq \textstyle\sum_{x\in\{0,1\}^n} \mu(x)\mathcal{X}_z(x) \leq \delta && \forall z \in \{0,1\}^n \setminus \{0\}.\end{aligned}$$

First, we note that the optimum is achieved by a π-symmetric distribution.

Lemma 3.1 *If μ is a probability distribution on $\{0,1\}^n$ such that $\mu(F_\pi = 1) = 0$, then there exists a π-symmetric probability distribution μ^* on $\{0,1\}^n$ such that $\mu^*(F_\pi = 1) = 0$ and $bias(\mu^*) \leq bias(\mu)$.*

This reduces the 2^n variables $\{\mu(x)\}_{x\in\{0,1\}^n}$ to $m+1$ variables $\{\underline{\mu}(w)\}_{w=0}^m$ ($\underline{\mu}$ is the induced distribution, as defined in Section 2.2).

Note that since our problem is about establishing an upper bound on the minimum bias, Lemma 3.1 is not used to establish other propositions in this paper. It is included as it justifies the focus on π-symmetric distributions. The proof of Lemma 3.1 is in Appendix A.

In Section 4, we study the bias of π-symmetric distributions, which naturally leads to Krawtchouk polynomials. We note that if μ is π-symmetric, then $bias_z(\mu)$ depends only on the π-weight t of z and it is given by

$$bias_t(\mu) = E_{w\sim\underline{\mu}}\mathcal{K}_t^{(m,c)}(w), \tag{2}$$

where $\mathcal{K}_t^{(m,c)}$ is the Krawtchouk polynomial. Krawtchouk polynomials $\{\mathcal{K}_t^{(m,c)}\}_{t=0}^m$ originate in our context as

$$\begin{aligned}\mathcal{K}_t^{(m,c)}(w) &= E_{x:W_\pi(x)=w}\mathcal{X}_z(x), \text{ where } z \in \{0,1\}^n \text{ is any vector of } \pi\text{-weight } t && (3)\\ &= \frac{1}{\binom{m}{w}(2^c-1)^w}\sum_a \binom{t}{a}\binom{m-t}{w-a}(-1)^a(2^c-1)^{w-a} && (4)\end{aligned}$$

The Krawtchouk expression of the bias reduces the $2^n - 1$ constraints $|bias_z(\mu)| \leq \delta$, for $z \neq 0 \in \{0,1\}^n$, to m constraints $|bias_t(\mu)| \leq \delta$, for $t = 1, \ldots, m$. Therefore we have a linear program with $m+2$ variables and $O(m)$ constraints.

Estimating the optimum of the LP is a difficult problem due to the sensitivity of Krawtchouk polynomials. Accordingly, rather than directly working with the bias of π-symmetric distributions, we reduce the problem to estimating the maximum d such that there exists a π-symmetric (d,π)-wise independent distribution whose support contains no satisfying input for F_π, i.e., $\mu(F_\pi = 1) = 0$.

3.1 Reduction

The (d,π)-wise independence property alone is not enough to guarantee low bias. We argue in Section 5 that the (d,π)-wise independence property together with the π-symmetry property guarantee low bias. Namely, we show in Corollary 5.3 that if μ be a π-symmetric (d,π)-wise independent probability distribution on $\{0,1\}^n$, then

$$bias(\mu) \leq 2\left(\frac{8\sqrt{dm}}{2^c-1}\right)^d \tag{5}$$

assuming that $m \geq 2^c$ and $d \geq 1$. The bound improves as d increases, which reduces our problem to estimating the maximum d such that there exists a π-symmetric probability distribution μ on $\{0,1\}^n$ such that μ is (d,π)-wise independent and $\mu(F_\pi = 1) = 0$. For instance, when $m = c2^c$, the bound is $2\left(\frac{8\sqrt{dc2^c}}{2^c-1}\right)^d = \Theta\left(\frac{d\log m}{m}\right)^{d/2}$. Hence to obtain $\beta_{m,c} = 2^{-\Omega(\log^2 n)}$, we need $d = \Omega(\log m)$.

We establish (5) as follows. If $t \leq d$, then $bias_t(\mu) = 0$ since μ is (d,π)-wise independent. For $t > d$, it follows from (2) that

$$|bias_t(\mu)| \leq max^m_{w=m-m_0}|\mathcal{K}^{(m,c)}_t(w)| + \underline{\mu}(w : w < m - m_0), \tag{6}$$

where m_0 is a parameter we will optimize on.

The key observation is that $|\mathcal{K}^{(m,c)}_t(w)|$ decays quickly with t if w is large enough. Namely, we show in Lemma 5.1 that if $m \geq 2^c$ and m_0 is an integer such that $1 \leq m_0 \leq m/2$, then

$$|\mathcal{K}^{(m,c)}_t(w)| \leq 4^{m_0^2/m}\left(\frac{2m_0}{2^c-1}\right)^t \text{ for } w = m - m_0, \ldots, m. \tag{7}$$

The proof of (7) is based on expression (4) of Krawtchouk polynomials. Eventually we will set $m_0 = \lfloor\sqrt{dm}\rfloor$. For instance, when $m = c2^c$, we will set $m_0 = \Theta(\sqrt{m\log m})$, which reduces (7) to $m^{O(1)}\left(\frac{\Theta(\log^3 m)}{m}\right)^{t/2}$. We will use (7) for large values of $t > d$, e.g., $t = \Omega(\log m)$ when $m = c2^c$.

The second term $\underline{\mu}(w : w < m - m_0)$ is the probability that the number of satisfied clauses is less than $m - m_0$. Since μ is (d,π)-wise independent, this probability decays quickly with d if $m - m_0$ is small enough compared to the average $E_{w\sim\mu}w = m - m/2^c$ (In the special case when $m = c2^c$,$E_{w\sim\mu}w = m - \Theta(\log n)$). This follows easily from a d'th moment inequality. We show in Lemma 5.2 that for $m \geq 2^c$,

$$\underline{\mu}(w : w < m - m_0) \leq d\left(\frac{md}{2^c m_0}\right)^d \tag{8}$$

By setting $m_0 = \lfloor\sqrt{dm}\rfloor$ and replacing (7) and (8) in (6), we obtain (5).

3.2 Estimating the maximum d

Given $d \geq 1$, consider the convex polytope $P^{(m,c)}_d \subset \mathbb{R}^{\{0,1\}^n}$ of (d,π)-wise independent π-symmetric probability distributions μ on $\{0,1\}^n$ such that $\mu(F_\pi = 1) = 0$. We estimate in Section 6 the maximum d such that $P^{(m,c)}_d$ is nonempty.

We study the low dimensional version of $P^{(m,c)}_d$. The polytope $P^{(m,c)}_d$ is nonempty iff its low dimensional version $\underline{P}^{(m,c)}_d := \{\underline{\mu} : \mu \in P^{(m,c)}_d\} \subset \mathbb{R}^{m+1}$ is nonempty. We show in Lemma 6.3 that $\underline{P}^{(m,c)}_d \neq \emptyset$ for $d = \lfloor\frac{m}{4\times 2^c}\rfloor$ (assuming that $200 \times 2^c \leq m \leq 2^{2c}$ and m divisible by 2^{c+1}). For instance, when $m = c2^c$, we get $\underline{P}^{(m,c)}_d \neq \emptyset$ for $d = \lfloor c/4\rfloor = \Theta(\log m)$, hence (5) reduces to $\beta_{m,c} = 2^{-\Omega(\log^2 n)}$ as noted above. The argument in Section 6 is in two parts.

First, by a duality argument, we give in Lemma 6.2 a univariate low degree polynomial characterization of the dual. Then, using this characterization, we estimate in Lemma 6.3 the maximum d using Chebyshev's minimax technique. The proof uses the fact that for the binomial distribution $bin_{m,c}$, $bin_{m,c}(x)$ is much larger than $bin_{m,c}(m)$ when x is close to the mean $m - m/2^c$ of $bin_{m,c}$ (Lemma 6.6). The proof of Lemma 6.3 boils down to showing that if $0 < a < b$ are integers and $g(z) \in \mathbb{R}[z]$ is a polynomial of degree at most d such that $g(0) = 1$, then there exists an integer $a \leq x_0 \leq b$ such that $|g(x_0)|$ is not very small if a and b are not very close. We establish in Lemma 6.4 the lower bound $\Omega(\frac{1}{d}\left(\frac{b-a-2d}{2eb}\right)^d)$ for $b - a > 2d$ (and $d \geq 5$). The

proof idea is based on a construction of a suitable polynomial on which we apply Chebyshev's alternating signs technique. Our approach was motivated by the work Linial and Nisan [LN90] who used Chebyshev polynomials to solve a related optimization problem.

In Section 7, we complete the proof of Theorems 1.4 and 1.7 and we verify the dual propositions stated in Corollaries 1.6 and 1.8.

We conclude in Section 8 with an open question related to extending the small bias condition in Theorem 1.4 from mod 2 to mod M gates.

4 Krawtchouk polynomials and the bias of symmetric distributions

For a general reference on Krawtchouk polynomials, see for instance [Sze75]. Krawtchouk polynomials originate in the proof when studying the bias of π-symmetric distributions via the character sum $\sum_{x:W_\pi(x)=w} \mathcal{X}_z(x)$. We adopt the following definition.

Lemma 4.1 *Let $0 \le w \le m$ be an integer and $z \in \{0,1\}^n$. Consider the summation $S_z(w) = \sum_{x:W_\pi(x)=w} \mathcal{X}_z(x)$. Then $S_z(w)$ depends only on the π-weight t of z and it is given by*

$$S_t(w) = \sum_a \binom{t}{a}\binom{m-t}{w-a}(-1)^a(2^c-1)^{w-a}.$$

The proof is below. Accordingly, we define Krawtchouk polynomial as follows.

Definition 4.2 *For $t = 0, \ldots, m$, define the t'th* Krawtchouk polynomial *as*

$$\begin{aligned} \mathcal{K}_t^{(m,c)}(w) &\coloneqq E_{x:W_\pi(x)=w}\mathcal{X}_z(x), \text{ where } z \in \{0,1\}^n \text{ is any vector of } \pi\text{-weight } t \qquad (9)\\ &= \frac{1}{\binom{m}{w}(2^c-1)^w}\sum_a \binom{t}{a}\binom{m-t}{w-a}(-1)^a(2^c-1)^{w-a}. \qquad (10)\end{aligned}$$

It follows from Lemma 4.1 that if μ is a π-symmetric distribution, then $bias_z(\mu)$ depends only on the π-weight of z, and it can be expressed in terms of the Krawtchouk polynomials and the probability distribution $\underline{\mu}$ on $[0:m]$ induced by μ.

Corollary 4.3 *Let μ be a π-symmetric probability distribution on $\{0,1\}^n$ and let $z \in \{0,1\}^n$. Then $bias_z(\mu)$ depends only on $t = W_\pi(z)$, i.e, $bias_\mu(z') = bias_\mu(z)$ for all $z' \in \{0,1\}^n$ such that $W_\pi(z') = W_\pi(z)$. Moreover, $bias_z(\mu)$ is given by $bias_t(\mu) = \sum_{w=0}^m \underline{\mu}(w)\mathcal{K}_t^{(m,c)}(w) = E_{\underline{\mu}}\mathcal{K}_t^{(m,c)}$.*

Proof: Since μ is π-symmetric, we have

$$bias_z(\mu) = \sum_{x\in\{0,1\}^n}\mu(x)\mathcal{X}_z(x) = \sum_{w=0}^m \underline{\mu}(w)\frac{1}{\binom{m}{w}(2^c-1)^w}\sum_{x:W_\pi(x)=w}\mathcal{X}_z(x) = \sum_{w=0}^m \underline{\mu}(w)\mathcal{K}_t^{(m,c)}(w).$$

■

Lemma 4.4 *For $t = 0, \ldots, m$, $\mathcal{K}_t^{(m,c)}(w)$ is a degree-t polynomial in the variable w.*

The reason is that $\mathcal{K}_t^{(m,c)}(w) = \mathcal{K}_w^{(m,c)}(t)$, which is verified below. The Krawtchouk polynomials $\{\mathcal{K}_t^{(m,c)}(x)\}_{t=0}^m$ form an orthogonal basis of the set of polynomials in $\mathbb{R}[x]$ of degree at most m. They are orthogonal with respect to the binomial probability distribution $bin_{m,c}$. In particular, we have

$$E_{w\sim bin_{m,c}}\mathcal{K}_t^{(m,c)}(w) = 0 \quad \text{for } t = 1, \ldots, m. \qquad (11)$$

We do not need the full orthogonality property in the proof of Theorem 1.4, we only need (11), which follows immediately from (9):

$$E_{w\sim bin_{m,c}}\mathcal{K}_t^{(m,c)}(w)=\sum_{w=0}^{m}bin_{m,c}(w)\frac{1}{\binom{m}{w}(2^c-1)^w}\sum_{x:W_\pi(x)=w}\mathcal{X}_z(x)=\frac{1}{2^{cm}}\sum_x\mathcal{X}_z(x)=0,$$

where $z\in\{0,1\}^n$ is any vector of π-weight t.

4.1 Proof of Lemma 4.1

Grouping terms, we get

$$S_z(w)=\sum_{S\subset[m]:|S|=w}\ \sum_{x\in\prod_{j\in S}(\{0,1\}^{\pi(j)}\setminus\{0\})}\mathcal{X}_z(\bar{x}),$$

where if $A\subset[n]$ and $x\in\{0,1\}^A$, $\bar{x}$ means the extension of x to $\{0,1\}^n$ by zeros.

Fix $S\subset[m]$ such that $|S|=w$. Using the multiplicativity of the character $\mathcal{X}_z$, and the fact that $\sum_{x\in\{0,1\}^{\pi(j)}}\mathcal{X}_z(\bar{x})$ is 0 if $z|_{\pi(j)}\neq 0$, and 2^c otherwise, we have

$$\sum_{x\in\prod_{j\in S}(\{0,1\}^{\pi(j)}\setminus\{0\})}\mathcal{X}_z(\bar{x})=\prod_{j\in S}\sum_{x\in\{0,1\}^{\pi(j)}\setminus\{0\}}\mathcal{X}_z(\bar{x})=(-1)^{|A_z\cap S|}(2^c-1)^{|A_z^c\cap S|},$$

where $A_z=\{j\in[m]:z|_{\pi(j)}\neq 0\}$. Therefore

$$S_z(w)=\sum_{S\subset[m]:|S|=w}(-1)^{|A_z\cap S|}(2^c-1)^{|A_z^c\cap S|},$$

and consequently $S_z(w)$ depends only on $|A_z|=t$. Grouping the terms which have the same value of $|A_z\cap S|$, we get

$$S_z(w)=\sum_{a=0}^{w}\sum_{S\subset[m]:|S|=w\&|A_z\cap S|=a}(-1)^a(2^c-1)^{w-a}=\sum_a\binom{t}{a}\binom{m-t}{w-a}(-1)^a(2^c-1)^{w-a}.$$

4.2 Proof of Lemma 4.4

Let $0\leq t,w\leq m$ and consider the identity

$$\sum_{z:W_\pi(z)=t}\left(\sum_{x:W_\pi(x)=w}\mathcal{X}_z(x)\right)=\sum_{x:W_\pi(x)=w}\left(\sum_{z:W_\pi(z)=t}\mathcal{X}_x(z)\right)$$

Note that $\mathcal{X}_z(x)=(-1)^{\langle x,z\rangle}=\mathcal{X}_x(z)$. Applying Lemma 4.1 to both sides, we get

$$\binom{m}{t}(2^c-1)^tS_t(w)=\binom{m}{w}(2^c-1)^wS_w(t).\tag{12}$$

It follows from (12) that

$$\begin{aligned}\mathcal{K}_t^{(m,c)}(w)&=\mathcal{K}_w^{(m,c)}(t)\\&=\frac{1}{\binom{m}{t}(2^c-1)^t}\sum_a\binom{w}{a}\binom{m-w}{t-a}(-1)^a(2^c-1)^{t-a}.\end{aligned}$$

Accordingly, $\mathcal{K}_t^{(m,c)}(w)$ is a degree-t polynomial in w.

5 Reduction

We derive in this section a bound on the bias of π-symmetric (d,π)-wise independent probability distributions on $\{0,1\}^n$. The bound is in Corollary 5.3 and it is based on Lemmas 5.1 and 5.2 below. See Section 3.1 for a detailed overview.

Lemma 5.1 *Assume that $m \geq 2^c$ and let m_0 be an integer such that $1 \leq m_0 \leq m/2$. If $w \geq m - m_0$, then $|\mathcal{K}_t^{(m,c)}(w)| \leq 4^{m_0^2/m}\left(\frac{2m_0}{2^c-1}\right)^t$ for $t = 1,\ldots,m$.*

Proof: The proof relies on upper bounding the $(-1)^a$ term in expression (10) of Krawtchouk polynomial by 1. Fix $0 \leq v \leq m_0$. We have

$$\begin{aligned}|\mathcal{K}_t^{(m,c)}(m-v)| &= \left|\frac{1}{\binom{m}{v}(2^c-1)^{m-v}}\sum_{a=max(0,t-v)}^{min(t,m-v)}\binom{t}{a}\binom{m-t}{m-v-a}(-1)^a(2^c-1)^{m-v-a}\right| \\ &= \left|\frac{1}{\binom{m}{v}}\sum_{a=max(0,t-v)}^{min(t,m-v)}\binom{t}{a}\binom{m-t}{v-t+a}(-1)^a\frac{1}{(2^c-1)^a}\right|\end{aligned}$$

since $m-v-a = (m-t)-(v-t+a)$. Upper bounding $(-1)^a$ by 1, we get

$$|\mathcal{K}_t^{(m,c)}(m-v)| \leq \frac{1}{\binom{m}{v}}\sum_{a=max(0,t-v)}^{min(t,m-v)}\binom{t}{a}\binom{m-t}{v-t+a}\frac{1}{(2^c-1)^a}$$

To simplify the above expression we use the bounds $\binom{m-t}{v-t+a} \leq \frac{m^{v-t+a}}{(v-t+a)!}$ and

$$\binom{m}{v} \geq 4^{-m_0^2/m}\frac{m^v}{v!}. \tag{13}$$

The proof of (13) is below. Thus

$$\begin{aligned}|\mathcal{K}_t^{(m,c)}(m-v)| &\leq 4^{\frac{m_0^2}{m}}\frac{v!}{m^v}\sum_{a=max(0,t-v)}^{min(t,m-v)}\binom{t}{a}\frac{m^{v-t+a}}{(v-t+a)!}\frac{1}{(2^c-1)^a} \\ &= 4^{\frac{m_0^2}{m}}\frac{1}{m^t}\sum_{a=max(0,t-v)}^{min(t,m-v)}\binom{t}{a}\frac{v!}{(v-t+a)!}\left(\frac{m}{2^c-1}\right)^a.\end{aligned}$$

Using the bounds $\frac{v!}{(v-t+a)!} \leq m_0^t$ and $\left(\frac{m}{2^c-1}\right)^a \leq \left(\frac{m}{2^c-1}\right)^t$ (since $m \geq 2^c$), we get

$$|\mathcal{K}_t^{(m,c)}(m-v)| \leq 4^{\frac{m_0^2}{m}}\left(\frac{m_0}{2^c-1}\right)^t\sum_{a=0}^{t}\binom{t}{a} = 4^{\frac{m_0^2}{m}}\left(\frac{2m_0}{2^c-1}\right)^t.$$

Proof of (13): If $v = 0$, the bound is trivial. If $v \neq 0$, we have

$$\binom{m}{v} \geq \frac{(m-v)^v}{v!} = \left(1-\frac{1}{m/v}\right)^{\frac{m}{v}\frac{v^2}{m}}\frac{m^v}{v!} \geq 4^{-\frac{v^2}{m}}\frac{m^v}{v!} \geq 4^{-\frac{m_0^2}{m}}\frac{m^v}{v!},$$

where the second inequality follows from the fact that $(1-1/x)^x \geq 1/4$ for $x \geq 2$ ($m/v \geq 2$ since $v \leq m_0 \leq m/2$). ■

Lemma 5.2 *Assume that $m \geq 2^c$ and let $m_0 \geq 1$. Let $d \geq 1$ and μ be a (d,π)-wise independent probability distribution on $\{0,1\}^n$. Then $\mu(x \in \{0,1\}^n : W_\pi(x) < m - m_0) \leq d\left(\frac{md}{2^c m_0}\right)^d$.*

Proof: The proof follows from a d'th moment inequality. Let $Z_\pi(x) = m - W_\pi(x)$ and $B = \{x \in \{0,1\}^n : Z_\pi(x) > m_0\}$. We are interested in upper-bounding $\mu(B)$. We have

$$Z_\pi(x) = \sum_{j=1}^{m} AND_{\pi(j)}(x \oplus \mathbf{1}),$$

where $AND_S(x) = \wedge_{i \in S}(x_i)$ for $S \subset [n]$, $\oplus$ denotes addition modulo 2 in $\{0,1\}^n$, and $\mathbf{1}$ denotes the all ones vector in $\{0,1\}^n$. Thus $E_{x\sim\mu} Z_\pi(x) = \frac{m}{2^c}$ if $d \geq 1$. By Markov Inequality

$$\mu(B) = Pr_{x\sim\mu}[Z_\pi(x) > m_0] = Pr_{x\sim\mu}[Z_\pi(x)^d > m_0^d] \leq \frac{E_{x\sim\mu} Z_\pi(x)^d}{m_0^d}. \tag{14}$$

We have

$$Z_\pi(x)^d = \left(\sum_{j=1}^{m} AND_{\pi(j)}(x \oplus \mathbf{1})\right)^d = \sum_{(j_1,\ldots,j_d)\in[m]^d} AND_{\pi(j_1)\cup\ldots\cup\pi(j_d)}(x \oplus \mathbf{1}),$$

hence

$$E_{x\sim\mu} Z_\pi(x)^d = \sum_{(j_1,\ldots,j_d)\in[m]^d} 2^{-|\pi(j_1)\cup\ldots\cup\pi(j_d)|} = \sum_{(j_1,\ldots,j_d)\in[m]^d} 2^{-c|\{j_1,\ldots,j_d\})|}$$

since μ is (d,π)-wise independent. Let $U(t)$ denote the number of tuples $(j_1,\ldots,j_d)$ in $[m]^d$ such that $|\{j_1,\ldots,j_d\}| = t$. Thus $U(t) \leq \binom{m}{t} t^d$, and accordingly

$$E_{x\sim\mu} Z_\pi(x)^d = \sum_{t=1}^{d} 2^{-ct} U(t) \leq \sum_{t=1}^{d} 2^{-ct}\binom{m}{t} t^d \leq \sum_{t=1}^{d} 2^{-ct} m^t d^d = \sum_{t=1}^{d} \left(\frac{m}{2^c}\right)^t d^d \leq d\left(\frac{md}{2^c}\right)^d$$

since $m \geq 2^c$. Replacing in (14), we get $\mu(B) \leq d\left(\frac{md}{2^c m_0}\right)^d$. ■

Corollary 5.3 *Assume that $m \geq 2^c$ and $d \geq 1$. If μ be a π-symmetric (d,π)-wise independent probability distribution on $\{0,1\}^n$, then $bias(\mu) \leq 2\left(\frac{8\sqrt{dm}}{2^c-1}\right)^d$.*

Proof: Let $1 \leq t \leq m$. If $t \leq d$, then $bias_t(\mu) = 0$ since μ is (d,π)-wise independent. Assume that $t > d$. By Corollary 4.3, $bias_t(\mu) = E_{\underline{\mu}} \mathcal{K}_t^{(m,c)}$. Let m_0 be an integer such that $1 \leq m_0 \leq m/2$. Thus,

$$|bias_t(\mu)| \leq max_{w=m-m_0}^{m} |\mathcal{K}_t^{(m,c)}(w)| + \underline{\mu}(w : w < m - m_0),$$

since $|\mathcal{K}_t^{(m,c)}(w)| \leq 1$ for all $0 \leq w \leq m$. Applying Lemmas 5.1 and 5.2, we obtain

$$|bias_t(\mu)| \leq 4^{m_0^2/m}\left(\frac{2m_0}{2^c-1}\right)^t + d\left(\frac{md}{2^c m_0}\right)^d \leq 4^{m_0^2/m}\left(\frac{2m_0}{2^c-1}\right)^d + d\left(\frac{md}{2^c m_0}\right)^d$$

since $t > d$. Note that the bound trivially holds if $\frac{2m_0}{2^c-1} > 1$ since $4^{m_0^2/m} \geq 1$, and accordingly the RHS is larger than 1. While the first term $4^{m_0^2/m}\left(\frac{2m_0}{2^c-1}\right)^d$ improves as m_0 decreases, the second term $d\left(\frac{md}{2^c m_0}\right)^d$ improves as m_0 increases. Setting $m_0 = \lfloor\sqrt{dm}\rfloor$ makes both terms at

most $\left(\frac{8\sqrt{dm}}{2^c-1}\right)^d$, and hence $|bias_t(\mu)| \leq 2\left(\frac{8\sqrt{dm}}{2^c-1}\right)^d$. To verify this, use the inequalities $\sqrt{dm}/2 \leq m_0 \leq \sqrt{dm}$. We have $4^{m_0^2/m}\left(\frac{2m_0}{2^c-1}\right)^d \leq 4^{(\sqrt{dm})^2/m}\left(\frac{2\sqrt{dm}}{2^c-1}\right)^d = \left(\frac{8\sqrt{dm}}{2^c-1}\right)^d$. On the other hand, $d\left(\frac{md}{2^c m_0}\right)^d \leq d\left(\frac{md}{2^c\sqrt{dm}/2}\right)^d = d\left(\frac{2\sqrt{md}}{2^c}\right)^d \leq \left(\frac{8\sqrt{dm}}{2^c-1}\right)^d$ since $d \leq \left(\frac{4\times 2^c}{2^c-1}\right)^d$, for all $d \geq 1$.

Finally, we note that the conditions $m_0 \leq m/2$ can be ignored since if $m_0 > m/2$, then $\sqrt{md} > m/2$, hence $2\left(\frac{8\sqrt{dm}}{2^c-1}\right)^d > 2\left(\frac{4m}{2^c-1}\right)^d > 1$ (since $m \geq 2^c$), which makes the corollary claim trivial.

■

6 The limited independence polytope

Definition 6.1 *Given $d \geq 1$, let $P_d^{(m,c)} \subset \mathbb{R}^{\{0,1\}^n}$ be the convex polytope of (d,π)-wise independent π-symmetric probability distributions μ on $\{0,1\}^n$ such that $\mu(F_\pi = 1) = 0$.*

Note that the defining constraints of $P_d^{(m,c)}$ are:

- $\mu \geq 0$
- $\sum_x \mu(x) = 1$,
- $\mu(x) = 0$, for each $x \in \{0,1\}^n$ such that $W_\pi(x) = m$,
- $\sum_x \mu(x)\mathcal{X}_z(x) = 0$, for each nonzero $z \in \{0,1\}^n$ such that $W_\pi(z) \leq d$.

The bound on the bias in Corollary 5.3 improves as d increases. We estimate in this section the maximum d such that $P_d^{(m,c)}$ is nonempty. Since a π-symmetric distribution μ is uniquely determined by its corresponding distribution $\underline{\mu}$ on $[0:m]$, it is enough to study the low dimensional version $\underline{P}_d^{(m,c)} := \{\underline{\mu} : \mu \in P_d^{(m,c)}\} \subset \mathbb{R}^{m+1}$ of the polytope $P_d^{(m,c)}$.

We derive in Lemma 6.2 a low-degree polynomial characterization of the dual of $\underline{P}_d^{(m,c)}$. Using this characterization, we estimate in Lemma 6.3 the maximum d such that $P_d^{(m,c)}$ is nonempty.

Lemma 6.2 *Let $d \geq 0$ be an integer. Then $P_d^{(m,c)} \neq \emptyset$ if and only if for each polynomial $f(x) \in \mathbb{R}[x]$ such that*

- *$deg(f) \leq d$,*
- *$f(m) = 1$, and*
- *$f(w) \leq 0$, for $w = 0, \ldots, m-1$,*

we have $E_{w\sim bin_{m,c}} f(w) \leq 0$.

Proof: We will work with $\underline{P}_d^{(m,c)}$ since $P_d^{(m,c)} \neq \emptyset$ iff $\underline{P}_d^{(m,c)} \neq \emptyset$. By Corollary 4.3, $\underline{P}_d^{(m,c)}$ is the convex polytope consisting of $\gamma : [0:m] \to \mathbb{R}$ such that:

a) $\gamma \geq 0$,

b) $\sum_{w=0}^m \gamma(w) = 1$,

c) $\gamma(m) = 0$,

d) $E_\gamma \mathcal{K}_t^{(m,c)} = 0$, for $t = 1, \ldots, d$.

Recall from Lemma 4.4, that the Krawtchouk polynomial $\mathcal{K}_t^{(m,c)}$ has degree t. Hence the polynomials $\{\mathcal{K}_t^{(m,c)}(x)\}_{t=0}^d$ form a basis of the set of polynomials in $\mathbb{R}[x]$ of degree at most d. Recall also from (11) that $E_{bin_{m,c}} \mathcal{K}_t^{(m,c)} = 0$ for $t = 1, \ldots, m$. It follows that constraints (b) and (d)

are equivalent to: $E_{w\sim\gamma}w^t = E_{w\sim bin_{m,c}}w^t$ for $t = 0,\ldots,d$. Accordingly, $P_d^{(m,c)} \neq \emptyset$ is equivalent to $L = 0$, where L is the Linear Program given below with its dual D:

$L = \min \gamma(m)$ where $\gamma : [0:m] \to \mathbb{R}$ is subject to: $\gamma(w) \geq 0$, for $w = 0,1,\ldots,m$ $\sum_{w=0}^{m} \gamma(w)w^t = \sum_{w=0}^{m} bin_{m,c}(w)w^t$, for $t = 0,\ldots,d$	$D = \max \sum_{t=0}^{d} a_t(\sum_{w=0}^{m} bin_{m,c}(w)w^t)$ where $a_0, a_1, \ldots, a_d \in \mathbb{R}$ are subject to: $\sum_{t=0}^{d} a_t m^t \leq 1$ $\sum_{t=0}^{d} a_t w^t \leq 0$, for $w = 0,1,\ldots,m-1$

Since $L = D$ (because L and D are feasible and bounded), we get $P_d^{(m,c)} \neq \emptyset$ iff for each polynomial $f(x) \in \mathbb{R}[x]$ such that (i) $deg(f) \leq d$, (ii) $f(m) \leq 1$, and (iii) $f(w) \leq 0$, for $w = 0,\ldots,m-1$, we have (iv) $E_{w\sim bin_{m,c}}f(w) \leq 0$. To complete the proof, we note that (ii) can be replaced with the condition $f(m) = 1$. If $f(m) \leq 0$, then (iv) holds trivially. If $0 < f(m) < 1$, we can scale $f(x)$ by $1/f(m)$ while preserving (i), (iii), and the sign of $E_{w\sim bin_{m,c}}f(w)$. ■

Lemma 6.3 *Assume that* $200 \times 2^c \leq m \leq 2^{2c}$ *and* m *divisible by* 2^{c+1}. *Let* $d_0 = \lfloor \frac{1}{4}\frac{m}{2^c} \rfloor$, *then* $P_{d_0}^{(m,c)} \neq \emptyset$.

Proof: First note that $d_0 \geq 50$ since $200 \times 2^c \leq m$. Let $d \geq 50$ be an integer such that $m/2^c \geq 4d$ and let $f(x) \in \mathbb{R}[x]$ be a polynomial of degree at most d such that $f(m) = 1$ and $f(w) \leq 0$ for $w = 0,\ldots,m-1$. By Lemma 6.2, it is enough to show that $E_{bin_{m,c}}f \leq 0$.

The polynomial $f(x)$ takes positive values on the discrete interval $[0:m]$ only at $x = m$. Hence to establish the lemma, it is enough to find a single integer value $x_0 \in [0:m-1]$ such that

$$|f(x_0)| \times \frac{bin_{m,c}(x_0)}{bin_{m,c}(m)} \geq 1. \tag{15}$$

Since $f(m) = 1$, the existence of x_0 implies that

$$E_{bin_{m,c}}f \leq bin_{m,c}(m)f(m) - bin_{m,c}(x_0)|f(x_0)| \leq 0.$$

The intuition behind the existence of x_0 is that the binomial distribution $bin_{m,c}$ is concentrated around its mean $m - \frac{m}{2^c}$, hence $bin_{m,c}(m)$ is much smaller than $bin_{m,c}(x)$ when x is close to the mean. To ensure (15), will find x_0 such that x_0 is close enough to the mean and $|f(x_0)|$ is large enough. The key is the following lemma.

Lemma 6.4 *Let* $d \geq 5$ *be an integer and* $0 < a < b$ *be integers such that* $b - a > 2d$. *If* $g(z)$ *is a real polynomial of degree at most* d *such that* $g(0) = 1$, *then there exists an integer* $a \leq x_0 \leq b$ *such that* $|g(x_0)| \geq \frac{3}{d}\left(\frac{b-a-2d}{2eb}\right)^d$.

Proof of Lemma 6.4. The idea of this proof is based on Chebyshev alternating signs technique which is classically used in conjunction with Chebyshev polynomials [Che66]. Our approach was motivated by the work of Linial and Nisan [LN90] who used Chebyshev polynomials to solve a related optimization problem. Chebyshev polynomials are not directly applicable to our problem since we are interested in an integer point x_0. We construct below alternating polynomials which suit the discrete nature of our problem.

Lemma 6.5 *If* $d \geq 5$ *and* $0 < a < b$ *are integers such that* $b - a > 2d$, *then there exist* $d+1$ *integer points* $a \leq A_1 < A_2 < \ldots < A_{d+1} \leq b$ *and a degree-*d *polynomial* $g^*(z) \in \mathbb{R}[z]$ *such that:*

- $g^*(0) = 1$
- $sign(g^*(A_j)) = -sign(g^*(A_{j+1}))$, *for* $j = 1,\ldots,d$
- $|g^*(A_j)| \geq \frac{3}{d}\left(\frac{b-a-2d}{2eb}\right)^d$, *for* $j = 1,\ldots,d+1$.

The proof of Lemma 6.5 is in Section 6.1. Let $g(z)$ be a polynomial of degree at most d such that $g(0) = 1$. We argue by contradiction that $|g(A_j)| \geq \frac{3}{d}\left(\frac{b-a-2d}{2eb}\right)^d$ for some $1 \leq j \leq d+1$. Assume the opposite and consider the polynomial $e(z) = g^*(z) - g(z)$. Since $|g^*(A_j)| > |g(A_j)|$ for all $1 \leq j \leq d+1$, $e(A_j)$ has the same sign as $g^*(A_j)$. Therefore, between A_1 and A_{d+1}, e changes sign at least d times and hence has at least d zeros on this interval. But $e(0) = g^*(0) - g(0) = 0$, hence $e(z)$ has at least $d+1$ zeros. It follows that the degree of $e(z)$ is at least $d+1$, which which is not possible since it is the difference of two polynomials each of degree at most d. ▼

We apply Lemma 6.4 to the polynomial $q(z) = f(m-z)$ whose degree is at most d, hence $q(0) = 1$ since $f(m) = 1$. We would like to bound $|f(x)|$ on an interval centered around the mean $m - m/2^c$ of $bin_{m,c}$. This interval should be neither too narrow (otherwise the maximum of $|f(x)|$ might be too small) nor too wide (otherwise it might include points that are so far from the mean that their probability is too low). Let $\rho = m/2^{c+1}$, thus the mean of $bin_{m,c}$ is $m - 2\rho$. We choose the interval $[m-3\rho : m-\rho]$ of width $2\rho = m/2^c$, hence $a = \rho$ and $b = 3\rho$. Note that the requirements $b - a > 2d$ (i.e., $m/2^c > 2d$) and $d \geq 5$ in Lemma 6.4 follow respectively from the stronger conditions $m/2^c \geq 4d$ and $d \geq 50$. By Lemma 6.4, there exists an integer $x_0 \in [m-3\rho : m-\rho]$ such that

$$|f(x_0)| \geq \frac{3}{d}\left(\frac{\rho - d}{3e\rho}\right)^d. \tag{16}$$

To verify (15), we still have to establish a lower bound on $bin_{m,c}(x_0)/bin_{m,c}(m)$.

Lemma 6.6 *Assume that* $200 \times 2^c \leq m \leq 2^{2c}$ *and* m *divisible by* 2^{c+1} *and let* $\rho = m/2^{c+1}$. *Then for each integer* $m - 3\rho \leq x \leq m - \rho$, *we have* $bin_{m,c}(x)/bin_{m,c}(m) \geq 5^\rho$.

The proof of Lemma 6.6 is in Section 6.2. Combining with (16), we get that to guarantee (15), it is enough to satisfy

$$\frac{3}{d}\left(\frac{\rho - d}{3e\rho}\right)^d 5^\rho \geq 1.$$

Since $m/2^c \geq 4d$, i.e., $\rho \geq 2d$, we have

$$\frac{3}{d}\left(\frac{\rho - d}{3e\rho}\right)^d 5^\rho \geq \frac{3}{(\rho/2)}\left(\frac{1}{6e}\right)^{\rho/2} 5^\rho = \frac{3}{(\rho/2)}\left(\frac{25}{6e}\right)^{\rho/2} \geq 1$$

since $\frac{3}{x}\left(\frac{25}{6e}\right)^x \geq 1$ for all $x \geq 1$. ■

6.1 Proof of Lemma 6.5

Let $r = \lfloor \frac{b-a}{2d} \rfloor$ and consider the points $A_1 = a, A_2 = a + 2r, \ldots, A_{d+1} = a + 2dr$ and $B_1 = a + r, B_2 = a + 3r, \ldots, B_d = a + (2d-1)r$. Thus $r \geq 1$ since $b - a > 2d$ and $a = A_1 < B_1 < A_2 < B_2 \ldots < B_d < A_{d+1} \leq b$. Define the degree-$d$ polynomial

$$G^*(z) \coloneqq \prod_{i=1}^{d}(z - B_i).$$

The polynomial G^* changes signs d times in the interval $[a, b]$, namely, for $j = 1, \ldots, d$, $sign(G^*(A_j)) = -sign(G^*(A_{j+1}))$. Let $g^*(z) = \frac{G(z)}{G(0)}$ so that $g^*(0) = 1$. We will argue below that

$$|G^*(A_j)| \geq |G^*(A_{\lceil \frac{d+1}{2} \rceil})| \qquad \text{for } j = 1, \ldots, d+1, \tag{17}$$

and

$$|G^*(A_{\lceil \frac{d+1}{2} \rceil})| \geq \frac{3}{d}\left(\frac{b-a-2d}{2e}\right)^d \qquad \text{if } d \geq 5. \tag{18}$$

Since $|G^*(0)| = \prod_{i=1}^{d} |B_i| \leq b^d$, we get $|g^*(A_j)| \geq \frac{3}{d}\left(\frac{b-a-2d}{2eb}\right)^d$ for $j = 1, \ldots, d+1$. We still have to verify (17) and (18).

Proof of (17): We have $G^*(A_j) = \prod_{i=1}^{d}(A_j - B_i) = \prod_{i=1}^{j-1}(A_j - B_i)\prod_{i=j}^{d}(A_j - B_i)$. Since $A_j - B_i = (2(j-i)-1)r$, we get

$$|G^*(A_j)| = r^d \left(\prod_{i=1}^{j-1}(2i-1)\right)\left(\prod_{i=1}^{d-j+1}(2i-1)\right).$$

The ratio $\frac{|G^*(A_{j+1})|}{|G^*(A_j)|} = \frac{2j-1}{2d-2j+1}$ is strictly less that 1 for $1 \leq j < \frac{d+1}{2}$, and strictly greater than 1 for $\frac{d+1}{2} < j \leq d$. Hence $|G^*(A_j)|$ achieves its minimum at $j = \lceil \frac{d+1}{2} \rceil$.

Proof of (18): First note that for all $t \geq 1$, $\prod_{i=1}^{t}(2i-1) \geq \prod_{i=1}^{t-1} 2i = 2^{t-1}(t-1)!$. We consider two cases according to the parity of d. If d is even, $\lceil \frac{d+1}{2} \rceil - 1 = \frac{d}{2}$ and $d - \lceil \frac{d+1}{2} \rceil + 1 = \frac{d}{2}$, hence

$$|G^*(A_{\lceil \frac{d+1}{2} \rceil})| = \left(\prod_{i=1}^{d/2}(2i-1)\right)^2 r^d \geq \left(2^{d/2-1}(d/2-1)!\right)^2 r^d.$$

We use Stirling approximation: $x! \geq \sqrt{2\pi x}(x/e)^x$, for $x \geq 1$ ($d/2 - 1 \geq 1$ since $d \geq 4$).

$$|G^*(A_{\lceil \frac{d+1}{2} \rceil})| \geq \left(2^{\frac{d-2}{2}}\sqrt{2\pi(d/2-1)}\left(\frac{d-2}{2e}\right)^{\frac{d-2}{2}}\right)^2 r^d = \frac{\pi e^2}{d-2}\left(1-\frac{2}{d}\right)^d\left(\frac{d}{e}\right)^d r^d \geq \frac{3}{d}\left(\frac{d}{e}\right)^d r^d$$

for all $d \geq 5$. If d is odd, $\lceil \frac{d+1}{2} \rceil - 1 = \frac{d-1}{2}$ and $d - \lceil \frac{d+1}{2} \rceil + 1 = \frac{d+1}{2}$, hence

$$|G^*(A_{\lceil \frac{d+1}{2} \rceil})| = \left(\prod_{i=1}^{\frac{d-1}{2}}(2i-1)\right)^2 dr^d \geq \left(2^{\frac{d-3}{2}}\left(\frac{d-3}{2}\right)!\right)^2 dr^d.$$

Using Stirling approximation (with $x = (d-3)/2 \geq 1$ for $d \geq 5$), we get

$$|G^*(A_{\lceil \frac{d+1}{2} \rceil})| \geq \left(2^{\frac{d-3}{2}}\sqrt{2\pi\frac{d-3}{2}}\left(\frac{d-3}{2e}\right)^{\frac{d-3}{2}}\right)^2 dr^d = \frac{\pi e^3 d}{(d-3)^2}\left(1-\frac{3}{d}\right)^d\left(\frac{d}{e}\right)^d r^d \geq \frac{3}{d}\left(\frac{d}{e}\right)^d r^d$$

for all $d \geq 5$. It follows that in both cases

$$|G^*(A_{\lceil \frac{d+1}{2} \rceil})| \geq \frac{3}{d}\left(\frac{d}{e}\right)^d \left\lfloor \frac{b-a}{2d} \right\rfloor^d \geq \frac{3}{d}\left(\frac{d}{e}\right)^d \left(\frac{b-a}{2d} - 1\right)^d = \frac{3}{d}\left(\frac{b-a-2d}{2e}\right)^d.$$

6.2 Proof of Lemma 6.6

The binomial distribution $bin_{m,c}$ achieves its minimum on the interval $[m-3\rho : m-\rho]$ at one of its extremities since its mean $m - 2\rho$ belongs to the interval. Thus it is enough to establish a lower bound on $bin_{m,c}(m-\beta\rho)/bin_{m,c}(m)$ for $\beta = 1, 3$. We have

$$\frac{bin_{m,c}(m-\beta\rho)}{bin_{m,c}(m)} = \frac{\frac{1}{2^{cm}}\binom{m}{\beta\rho}(2^c-1)^{m-\beta\rho}}{\frac{1}{2^{cm}}(2^c-1)^m} = \frac{1}{(2^c-1)^{\beta\rho}}\binom{m}{\beta\rho} \geq \frac{1}{2^{c\beta\rho}}\frac{(m-\beta\rho)^{\beta\rho}}{(\beta\rho)!}.$$

Using Stirling Approximation ($x! \leq ex^{x+1/2}e^{-x}$ for $x \geq 1$), we get

$$\frac{bin_{m,c}(m-\beta\rho)}{bin_{m,c}(m)} \geq \frac{\left(\frac{m-\beta\rho}{2^c}\right)^{\beta\rho}}{e\sqrt{\beta\rho}\,(\beta\rho)^{\beta\rho}\,e^{-\beta\rho}} = \frac{(1-\frac{\beta}{2^{c+1}})^{\beta\rho}(2e/\beta)^{\beta\rho}}{e\sqrt{\beta\rho}} \geq \frac{(1-\frac{3}{2^{c+1}})^{3\times 2^{c-1}}(2e)^{\rho}}{e\sqrt{3\rho}}$$

because $\rho = m/2^{c+1} \leq 2^{c-1}$ (since $m \leq 2^{2c}$) and $(2e/3)^3 > 2e$. Since $(1 - \frac{3}{2^{c+1}})^{3\times 2^{c-1}} \geq 1/64$ for $c \geq 1$ and since $2e > 5$, we obtain

$$\frac{bin_{m,c}(m - \beta\rho)}{bin_{m,c}(m)} \geq \frac{(2e)^\rho}{64e\sqrt{3\rho}} \geq 5^\rho$$

for $\rho \geq 100$, i.e, $m/2^c \geq 200$.

7 Concluding the proof

In this section, we conclude the proof of Theorems 1.4 and 1.7 and we verify the dual propositions stated in Corollaries 1.6 and 1.8.

7.1 Proof of Theorem 1.4

The following follows from Lemma 6.3 and Corollary 5.3.

Corollary 7.1 *Let $0 < \beta < 1/2$ be an absolute constant. Assume that $200 \leq \frac{m}{2^c} \leq 2^{\beta c}$ and 2^{c+1} divides m. Then there exists a probability distribution μ on $\{0,1\}^n$ such that $\mu(F_\pi = 1) = 0$ and $bias(\mu) = 2^{-\Omega(\log(m)\frac{m}{2^c})}$.*

Proof: By Lemma 6.3, $P_{d_0}^{(m,c)} \neq \emptyset$, where $d_0 = \lfloor \frac{1}{4}\frac{m}{2^c} \rfloor$. Let $\mu \in P_{d_0}^{(m,c)}$. Thus μ is π-symmetric, (d_0, π)-wise independent, and $\mu(F_\pi = 1) = 0$. By Corollary 5.3, $bias(\mu) \leq 2\left(\frac{8\sqrt{d_0 m}}{2^c - 1}\right)^{d_0} = 2^{-\Omega((\frac{c}{2} - \log\frac{m}{2^c})\frac{m}{2^c})}$. Since $\frac{m}{2^c} \leq 2^{\beta c}$, we have $\frac{c}{2} - \log\frac{m}{2^c} \geq (\frac{1}{2} - \beta)c \geq \frac{1/2-\beta}{1+\beta}\log m = \Omega(\log m)$. ■

We have $p = (1 - 2^{-c})^m = 2^{-\Theta(m/2^c)}$, hence $\frac{m}{2^c} = \Theta(\log\frac{1}{p})$. The condition $\frac{m}{2^c} \leq 2^{\beta c}$ is equivalent to $\frac{m}{2^c} \leq m^{\beta/(1+\beta)}$. Set $\beta = 0.43$, thus $\beta/(1+\beta) > 0.3006$, and hence the condition $\frac{m}{2^c} \leq 2^{\beta c}$ is guaranteed by $p > 2^{-m^{0.3006}}$ for m large enough. The condition $200 \leq \frac{m}{2^c}$ is equivalent to p less than some positive constant for m large enough.

Finally, we get rid of the requirement that 2^{c+1} divides m. Let G be the formula resulting from F by removing clauses to guarantee the divisibility requirement. Let m' be the number of clauses of G and p' be its probability of acceptance under the uniform distribution. Apply Corollary 7.1 to G, extend μ to the variables of F by adding uniformly random bits, and let μ' be the resulting distribution. We have $bias(\mu') = bias(\mu) = 2^{-\Omega(\log(m')\frac{m'}{2^c})}$ and $Pr_{\mu'}[F = 1] = 0$ since $Pr_\mu[G = 1] = 0$. The number of removed clauses is at most $2^{c+1} - 1$, thus $m = \Theta(m')$ (since $m \geq 200 \times 2^c$) and $p = \Theta(p')$ ($p(1 - 2^{-c})^{2^{c+1}-1} \leq p' \leq p$), which do not affect the asymptotic statements.

7.2 Proof of Theorem 1.7

Theorem 1.7 follows from Corollary 5.3. Assume that $200 \times 2^c \leq m \leq 2^{2c}$ and 2^{c+1} divides m. Then there exists a π-symmetric (d_0, π)-wise independent probability distribution μ on $\{0,1\}^n$ such that $\mu(F_\pi = 1) = 0$ and $d_0 = \lfloor \frac{1}{4}\frac{m}{2^c} \rfloor$. In terms of $p = 2^{-\Theta(m/2^c)}$, we have $d_0 = \Theta(\log\frac{1}{p})$. Moreover, the condition $m \leq 2^{2c}$ is equivalent to $\frac{m}{2^c} \leq \sqrt{m}$, hence it is guaranteed by the requirement $p = 2^{-o(\sqrt{m})}$ for m large enough. The condition $200 \leq \frac{m}{2^c}$ is equivalent to p less than some positive constant for m large enough. Finally, we get rid of the requirement that 2^{c+1} divides m by arguing as above. Construct the formula G as above and define m' and p' accordingly. Apply Corollary 5.3 to G and construct μ' from μ as above by adding uniformly random bits. Thus μ' is $(\lfloor \frac{1}{4}\frac{m'}{2^c} \rfloor, \pi)$-wise independent, $Pr_{\mu'}[F = 1] = 0$, $m = \Theta(m')$, and $p = \Theta(p')$.

7.3 Proof of Corollary 1.6

Let μ be the the probability distribution exhibited in Theorem 1.4. Assume without loss of generality that $E_U f_{low} = 1$. We have $\mu(x) = 0$ for each $x \in \{0,1\}^n$ such that $F(x) = 1$, and $f_{low}(x) \leq 0$ for each $x \in \{0,1\}^n$ such that $F(x) = 0$. Thus $E_\mu f_{low} \leq 0$. Consider the Fourier expansion of f_{low}: $f_{low}(x) = 1 + \sum_{z \neq 0} \hat{f}_{low}(z)\mathcal{X}_z(x)$ ($E_U f_{low} = 1$). Thus $0 \geq E_\mu f_{low} \geq 1 - \delta \sum_{z \neq 0} |\hat{f}_{low}(z)|$ since $|E_\mu \mathcal{X}_z| \leq \delta$ for each $z \neq 0$ because μ is δ-biased. It follows that $\sum_{z \neq 0} |\hat{f}_{low}(z)| \geq 1/\delta$, and hence $\|\hat{f}_{low}\|_1 \geq 1/\delta + 1$.

7.4 Proof of Corollary 1.8

Let μ be the the probability distribution exhibited in Corollary 1.7. By arguing as above, we have $E_\mu p_{low} \leq 0$. Consider the Fourier expansion of p_{low} on $\{0,1\}^n$: $p_{low}(x) = E_U p_{low} + \sum_{z \neq 0} \hat{p}_{low}(z)\mathcal{X}_z(x)$. Thus $E_\mu p_{low} = E_U p_{low} + \sum_{z \neq 0} \hat{p}_{low}(z) E_\mu \mathcal{X}_z(x)$. Since $E_\mu p_{low} \leq 0$ and $E_U p_{low} > 0$, we must have $\sum_{z \neq 0} \hat{p}_{low}(z) E_\mu \mathcal{X}_z(x) \neq 0$. Since μ is (d_0, π)-wise independent, $E_\mu \mathcal{X}_z(x) = 0$ for each nonzero z of π-weight less than or equal to d_0. Thus $\hat{p}_{low}(z) \neq 0$ for some z of π-weight greater than d_0, i.e., p_{low} has a monomial whose variables appear in more than $d_0 = \Theta(\log \frac{1}{p})$ clauses.

8 Open problem

A natural extension of the small bias property is from mod 2 to mod M gates. Let $M \geq 2$ be an integer. Call a probability distribution μ on $\{0,1\}^n$ (δ, M)-biased [MZ09, LRTV09] if $|E_{x \sim \mu} \xi_M^{\langle x,a \rangle} - E_{x \in \{0,1\}^n} \xi_M^{\langle x,a \rangle}| \leq \delta$, for each nonzero $a \in \mathbb{Z}_M^n$, where $\mathbb{Z}_M$ is the additive group modulo M, $\xi_M = e^{2\pi i/M}$, and $\langle x, a \rangle = \sum_{i=1}^n x_i a_i$.

Is the above extension of small bias enough to construct hitting sets for read-once CNF formulas? We believe that the answer is no.

One way to verify this claim is to extend Corollary 5.3. Let F be a read-once CNF formula with m clauses each of size c. Consider the simple case when the probability of acceptance of the read-once CNF formula F is $n^{-\Theta(1)}$, thus $c = \log m - \log \Theta(\log m)$. We established in Lemma 6.3 the existence of a π-symmetric (d, π)-wise independence distribution μ such that $d = \Theta(\log n)$ and no element in the support of μ satisfies F [5]. In Corollary 5.3, we argued that since (i) μ is π-symmetric and (ii) μ is (d, π)-wise independent, then μ is $2^{-\Omega(\log^2 n)}$-biased.

Conditions (i) and (ii) are not particularly related to mod 2 gates. Is μ a $(n^{-\omega(1)}, M)$-biased? for $M = O(1)$? for larger values of M? We believe that the answer is yes.

References

[ACR98] A.E Andreev, A.E.F Clementi, and J.D.P Rolim. A new general derandomization method. *Journal of the association for computing machinery* 45(1): 179-213, 1998

[AGHP92] N. Alon, O. Goldreich, J. Hastad, and R. Peralta. Simple Constructions of Almost k-wise Independent Random Variables. *Random Structures and Algorithms*, 3(3):289-304, 1992.

[Baz03] Louay Bazzi. Minimum Distance of Error Correcting Codes versus Encoding Complexity, Symmetry, and Pseudorandomness. Ph.D. dissertation, MIT, Cambridge, Mass., 2003.

[5] Lemma 6.3 assumes that the number m of clauses is divisible by 2^{c+1}. We can avoid this requirement by first using the trick in Section 7.2, and then the symmetrization procedure in Appendix A to restore the π-symmetry condition.

[Baz07] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 63-73, 2007.

[Baz09] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM journal on Computing*, Volume 38, Issue 6, pages 2220-2272, 2009.

[Bra10] Mark Braverman, "Poly-logarithmic independence fools AC0 circuits", *Journal of the ACM* 57(5), 2010.

[BM82] M. Blum and S. Micali. How to generate Cryptographically Strong Sequences of Pseudo-Random Bits. *SIAM journal on Computing*, 13(4):850-864, 1984.

[Che66] E.W Cheney, Approximation Theory, McGraw Hill Co,1966

[DETT10] A. De, O. Etesami, L. Trevisan, and M. Tulsiani, *Improved Pseudorandom Generators for Depth 2 Circuits.*, APPROX-RANDOM 2010:504-517

[GMRTV12] P. Gopalan, R. Meka, O. Reingold, L. Trevisan, and S. Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. *Proceedings of the 53rd IEEE Symposium on Foundations of Computer Science*, 2012.

[IW97] R. Impagliazzo and A. Wigderson. P = BPP if E Requires Exponential Circuits: Derandomizing the XOR Lemma. In *Proc. 29th Annual ACM Symposium on the Theory of Computing*, pages 220-229, 1997.

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

[LRTV09] S. Lovett, O. Reingold, L. Trevisan, and S. Vadhan. Pseudorandom bit generators that fool modular sums. APPROX-RANDOM 2009:615 - 630.

[Lub85] Michael Luby. A simple parallel algorithm for the maximal independent set problem. In *Proc. 17th Annual ACM Symposium on the Theory of Computing*, pages 1-10, 1985.

[LV96] M. Luby and B. Velickovic, On Deterministic Approximation of DNF. *Algorithmica*, 16(4/5):415-433, 1996.

[LVW93] M. Luby, B. Velickovic, and A. Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the 2nd ISTCS*, pages 18-24, 1993.

[MZ09] R. Meka and D. Zuckerman. Small-bias spaces for group products. APPROX-RANDOM 2009:658 - 672.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 12(4):63-70, 1991.

[NN93] J. Naor and M. Naor. Small bias probability spaces: efficient constructions and applications. *SIAM J. on Computing*, 22(4):838-856, 1993.

[NW88] N. Nisan and A. Wigderson. Hardness vs. Randomness. In *Proc. 29th IEEE Symposium on Foundations of Computer Science*, pages 2-11, 1988.

[Raz09] Alexander Razborov, "A simple proof of Bazzi's theorem", *ACM Trans. Comput. Theory*, 1(1):1-5, 2009.

[SZ11] J. Sima and S. Zak, Almost k-wise independent sets establish hitting sets for width 3 1-branching programs, *6th International Computer Science Symposium in Russia*, pages 120-133, 2011.

[Sze75] G. Szego, "Orthogonal Polynomials", Fourth edition, Colloquium Publications, Vol. 23, Amer. Math. Soc. Providence R.I., 1975.

[TX12] L. Trevisan and T. Xue, "A Derandomized Switching Lemma and an Improved Derandomization of AC0", TR12-116, Electronic Colloquium on Computational Complexity, 2012.

[Vaz86] Umesh Vazirani. Randomness, adversaries, and computation. Ph.D. dissertation, University of California, Berkeley, 1986.

[Yao82] Andrew C. Yao. Theory and application of Trapdoor functions. In *Proc. 23rd IEEE Annual Symposium on Foundations of Computer Science*, pages 80-91, 1982.

Appendix

A Proof of Lemma 3.1

Lemma 3.1 *If μ is a probability distribution on $\{0,1\}^n$ such that $\mu(F_\pi = 1) = 0$, then there exists a π-symmetric probability distribution μ^* on $\{0,1\}^n$ such that $\mu^*(F_\pi = 1) = 0$ and $bias(\mu^*) \leq bias(\mu)$.*

Proof: Let $G \subset GL_n(\mathbb{F}_2)$ be the group of $n \times n$ invertible π-block permutation matrices over $\mathbb{F}_2$, i.e., G consists of the invertible $n \times n$ matrices T over $\mathbb{F}_2$ such that: $\forall j_1 \in [m]$, $\exists$ a unique $j_2 \in [m]$ such that $\forall j_3 \neq j_2 \in [m]$, we have $T_{i_1,i_3} = 0$, $\forall i_1 \in \pi(j_1)$ and $i_3 \in \pi(j_3)$.

For $T \in G$, define the probability distribution μ_T on $\{0,1\}^n$ as $\mu_T(x) \coloneqq \mu(Tx)$. Symmetrize μ by averaging: define the probability distribution μ^* on $\{0,1\}^n$ as $\mu^*(x) \coloneqq E_{T \in G}\mu_T(x)$. The key points are:

i) $W_\pi(x) = W_\pi(Tx)$, $\forall x \in \mathbb{F}_2^n$ and $\forall T \in G$

ii) Conversely, $\forall x, y \in \mathbb{F}_2^n$ such that $W_\pi(x) = W_\pi(Tx)$, $\exists T \in G$ such that $y = Tx$

iii) $bias(\mu_T) = bias(\mu)$ for each $T \in G$ since the matrices in G are invertible. This follows from the fact that $bias_z(\mu_T) = bias_{T^{-1^*}z}(\mu)$, where * is the transpose operator [6].

It follows from (i) that $\mu_T(F_\pi = 1) = 0$ since $\mu(F_\pi = 1) = 0$ for each $T \in G$. Hence $\mu^*(F_\pi = 1) = 0$. The fact that μ^* is π-symmetric follows from (ii).

Finally, for each $z \in \{0,1\}^n$, we have $bias_z(\mu^*) = E_{T \in G} bias_z(\mu_T)$, hence $|bias_z(\mu^*)| \leq \max_{T \in G} |bias_z(\mu_T)|$. Therfore, it follows from (iii) that $bias(\mu^*) \leq \max_{T \in G} bias(\mu_T) = bias(\mu)$. ■

[6] Since $\langle T^{-1}x, z\rangle = \langle x, T^{-1^*}z\rangle$, we have $\sum_x \mu(Tx)(-1)^{\langle x,z\rangle} = \sum_x \mu(x)(-1)^{\langle T^{-1}x,z\rangle} = \sum_x \mu(x)(-1)^{\langle x,T^{-1^*}z\rangle}$.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de