

On Sampling Lower Bounds for Polynomials

Mohammad Mahdi Khodabandeh
mmk25@sfu.ca
Simon Fraser University

Igor Shinkar
ishinkar@sfu.ca
Simon Fraser University

May 1, 2026

Abstract

In this work, we continue the line of research on the complexity of distributions [Vio12], and study samplers defined by low degree polynomials. An n -tuple $\mathcal{P} = (P_1, \dots, P_n)$ of functions $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ defines a distribution over $\{0, 1\}^n$ in the natural way: draw X uniformly at random from \mathbb{F}_2^m and output $(P_1(X), \dots, P_n(X)) \in \{0, 1\}^n$.

We show that when \mathcal{P} is defined by polynomials of degree d , the total variation distance of \mathcal{P} from the product distribution $\text{Ber}(1/3)^{\otimes n}$ is $1 - o_n(1)$, where $o_n(1)$ is a vanishing function of n for any constant degree d . For small values of d , we show the following concrete bounds.

- For $d = 1$ we have $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\Omega(n))$.
- For $d = 2$ we have $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\Omega(\log(n)/\log \log(n)))$.
- For $d = 3$ we have $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\Omega(\sqrt{\log \log(n)}))$.

Our results extend the recent lower bound results for sampling distributions, which have mostly focused on local samplers, small depth decision trees, and small depth circuits [LV11, BIL12, Vio23, FLRS23, KOW24].

As part of our proof, we establish the following result, that may be of independent interest: for any degree- d polynomial $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ it holds that $\Pr_X[P(X) = 1]$ is bounded away from $1/3$ by some absolute constant $\delta = \delta_d > 0$. Although the statement may seem obvious, we are not aware of an elementary proof of this.

The proof techniques rely on the structural results for low degree polynomials [GT09, KL08, HS10], saying that any biased polynomial of degree d can be written as a function of a small number of polynomials of degree $d - 1$.

Contents

1	Introduction	3
2	Our results	6
3	Preliminaries	14
4	Chebyshev lemma	20
5	A sunflower pairwise regularization	24
6	One polynomial's distance from $\text{Ber}(1/3)$	27
7	Lower bound against linear distributions	30
8	Lower bound against bounded linear rank distributions	30
9	Lower bound against quadratic distributions	35
10	Lower bounds against distributions of bounded quadratic rank	37
11	Lower bounds against distributions of degree three	41
12	Lower bound against degree d distributions	42
13	Acknowledgments	43
A	One polynomial's distance from $\text{Ber}(\text{non-dyadic})$	47
B	Deferred proofs	49

1 Introduction

1.1 Background and previous work

The seminal work of Viola [Vio12] initiated the systematic study of a topic referred to as *the complexity of distributions*. In this framework, one has a target distribution \mathcal{T} over $\{0, 1\}^n$ that one wishes to sample from using uniformly random bits. Sampling is then the task of designing an algorithm A that, given a uniformly random input bit string $X \sim \{0, 1\}^m$, outputs a string $A(X) \in \{0, 1\}^n$ such that the resulting output distribution is close to \mathcal{T} in statistical distance. The class of algorithms A under consideration, i.e., the model of computation, and the target distribution determine the nature of the problem in this context. One is then led to both upper- and lower-bound questions: to either show the existence of an algorithm that (approximately) samples the distribution \mathcal{T} , or to show that every such algorithm fails to generate a distribution close to \mathcal{T} .

Over the past fifteen years this area has received considerable attention [Vio12, LV11, BIL12, DW12, Vio16, Vio20, GW20, CGZ22, Vio23, FLRS23, KOW24]. Viola [Vio12] started by showing the first impossibility result that local functions $f: \{0, 1\}^m \rightarrow \{0, 1\}^n$, where every output bit $f(X)_i$ depends on only a few input bits, are unable to sample a uniformly random string from a Hamming slice $H_k = \{x \in \{0, 1\}^n : \sum_i x_i = k\}$. The work also featured an application of such sampling lower bounds to lower bounds for succinct data structures which were extended in [KOW24]. Subsequent works [WP26, KOW24] explored applications of a different kind, namely separations between the sampling power of quantum and classical circuits. The hardness of sampling was also used to construct explicit codes [SS24].

The following typical example illuminates why sampling is easier than computing in general, and therefore, proving lower bounds is a more difficult task. For uniformly random $(X_1, X_2, \dots, X_n) \in \{0, 1\}^n$ the distribution $(X_1, X_2, \dots, X_n, X_1 \oplus \dots \oplus X_n)$ can be alternatively expressed as $(X_1 \oplus X_2, X_2 \oplus X_3, \dots, X_{n-1} \oplus X_n, X_n \oplus X_1)$ where \oplus denotes sum mod 2. Note that although the former view famously has no sub-exponential size circuit in \mathbf{AC}^0 ([FSS84, Ajt83, Has86]), the latter form can be implemented with a 2-local circuit.

Examples of this kind brought attention to the study of the sampling power of local maps or related models such as shallow decision forests. On this front, Filmus, Leigh, Riazanov, and Sokolov [FLRS23] show a lower bound against shallow decision forests for sampling from a Hamming slice $H_{o(n)}$. They also posed a conjecture characterizing the power of \mathbf{NC}^0 in sampling uniform distributions over a symmetric support, which was resolved by Kane, Ostuni, and Wu [KOW25a] in the affirmative. They later extended this result to characterize all symmetric distributions sampled by local maps [KOW25b]. Horacek, Lee, Shinkar, Viola, and Zhou [HLS⁺25] in a recent result showed that \mathbf{NC}^0 can sample a product of dyadic Bernoullies using nearly the information-theoretically minimum number of uniform random bits.

1.2 This work

We turn our focus to proving lower bounds for the class of degree d polynomials over \mathbb{F}_2 , a class that contains within itself the class of d -local maps and depth- d decision forests. A *degree d distribution* over $\{0, 1\}^n$ is defined by a parameter $m \in \mathbb{N}$ and n polynomials $P_1, \dots, P_n: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ of degree d . We sample from such a distribution $\mathcal{P} = (P_1, \dots, P_n)$ by sampling a uniformly random input $X \sim \mathbb{F}_2^m$ and outputting $(P_1(X), \dots, P_n(X))$. The main result of this paper shows that for any constant d any degree- d distribution is $1 - o_n(1)$ -far from $\text{Ber}(1/3)^{\otimes n}$ in total variation distance, where $o_n(1)$ is a vanishing function of n for any constant d . For small degrees $d = 1, 2, 3$ we show concrete bounds on the vanishing term $o_n(1)$.¹

We also study the class of distributions generated by *bounded rank* polynomials where each P_i is a function of bounded rank_d . Rank, as used here, comes from the work of Green and Tao [GT09] who defined it to study its relationship with pseudorandom properties of polynomials and Gowers norms. A function $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ is said to have $\text{rank}_d(P) \leq r$ if there exist r polynomials $Q_1(x), \dots, Q_r(x)$ of degree at most d and some $\Gamma: \mathbb{F}_2^r \rightarrow \mathbb{F}_2$ such that $P(x) = \Gamma(Q_1(x), \dots, Q_r(x))$ for all x . Distributions defined by functions of small rank_d essentially interpolate between degree d and $d + 1$ distributions. Indeed, our strategy for proving lower bounds against low degree polynomials will be to consider distributions defined by bounded rank functions. Specifically, we will first consider distributions (P_1, \dots, P_n) of bounded linear rank, i.e., where each P_i satisfies $\text{rank}_1(P_i) \leq r$ for which we show a lower bound of $1 - \exp(-2^{-O(r^2)}n)$ on the distance from $\text{Ber}(1/3)^{\otimes n}$. Note that these distributions already generalize r -local distributions. Then, we will consider functions of bounded quadratic rank for which we prove a $1 - \exp(-\log(n)^{\Omega(1/r^2)})$ lower bound. These will help us prove a $1 - \exp(-\Omega(\sqrt{\log \log(n)}))$ lower bound against degree three distributions.

Analogous results have been shown for d -local maps in [KOW24, Vio23], where the lower bound is against sampling Hamming slices. [Vio23] states the result against depth- d decision forests for sampling from the Hamming slice $H_{n/h}$ for h not a power of 2, while the result of [KOW24] extends it to any h . Many of our results can be translated from the setting of product distribution $\text{Ber}(1/3)^{\otimes n}$ to a sampling lower bound of $H_{n/3}$ against degree d distributions with essentially no change.

While it is natural to expect that for any $\rho \in (0, 1)$ similar lower bounds hold for $H_{\rho n}$ against degree d distributions, in the product distribution setting $\text{Ber}(\rho)^{\otimes n}$ it is essential that the parameter ρ is some non-dyadic, since for dyadic $\rho = a/2^d$ the distribution $\text{Ber}(\rho)^{\otimes n}$ can be perfectly sampled by polynomials of degree d .

An essential (and non-trivial) step in our results is to show that a single polynomial $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ cannot sample $\text{Ber}(1/3)$ better than some constant $\Omega_d(1)$ that does not depend on m . Put differently, $|\Pr_X[P(X) = 1] - 1/3|$ is bounded away from zero by some constant that only depends on d . In contrast, the analogous question for d -local distributions is trivial, since the probability that a d -local function outputs 1 is $a/2^d$ for some integer a which leaves a $\Omega(2^{-d})$ gap from $1/3$; however, a degree d polynomial $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ can genuinely depend on all m bits, and even for degree $d = 2$ we can

¹The parameter $1/3$ is fixed throughout the paper for simplicity, and can be replaced by any non-dyadic $\rho \in (0, 1)$.

construct a quadratic polynomial with $\Pr[P(x) = 1] = 3/8$, which clearly is not of the form $a/2^d$. To show this, we use structural results for low degree polynomials from the works of Green and Tao [GT09], Kaufman and Lovett [KL08], and Haramaty and Shpilka [HS10]. A related problem to this result is the weight distribution of binary Reed-Muller codes [KT70, KTA76, KLP12] which asks for given parameters $m, d \in \mathbb{N}, \eta \in (0, 1)$ how many polynomials $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ of degree d exist that satisfy $\Pr_X[P(X) = 1] \leq \eta$. Our result says that for a constant $\delta = \delta_d > 0$, the number of such polynomials does not change when η is in the δ -neighborhood of $1/3$, that is, $\eta \in [1/3 - \delta, 1/3 + \delta]$.

As mentioned above, the parameter $1/3$ is fixed throughout the paper for simplicity only, and analogous distance lower bound results can be shown for the distribution $\text{Ber}(\rho)^{\otimes n}$ for any non-dyadic parameter $\rho \in (0, 1)$. In fact, since for a single degree d polynomial $\Pr[P(x) = 1]$ is bounded away from, say $\rho = 1/2^{d+1}$, our distance lower bounds also extend to $\text{Ber}(1/2^{d+1})^{\otimes n}$.

At a high level, our results form a hierarchy where we show lower bounds against degree 1, bounded rank₁, degree 2, bounded rank₂, and so on, at the cost of progressively weaker distance guarantees. We conjecture that the correct lower bound for sampling $\text{Ber}(1/3)^{\otimes n}$ with distributions of bounded rank_d $\leq r$ should be $1 - \exp(-\Omega_{d,r}(n))$.

One of the main technical tools in establishing our results is a *Chebyshev lemma* stating the following. Suppose we are given a distribution $\mathcal{P} = (P_1, \dots, P_n)$, where each P_i can be written as $P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir})$ for some degree d polynomials Q_{i1}, \dots, Q_{ir} , that are *almost pairwise independent*. That is, for each $i \neq j$ the $2r$ -tuple $(Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr})$ is a collection of *almost independent* polynomials. Then, the distance of \mathcal{P} from the distribution $\text{Ber}(1/3)^{\otimes n}$ is $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - o_r(1) - o_n(1)$.

One issue is that, generally, polynomial distributions are not *almost pairwise independent*. However, we can hope for a relaxed scenario of finding a *subset* of the P_i 's so that the corresponding (Q_{i1}, \dots, Q_{ir}) satisfy the required *almost pairwise independence* for P_i 's in the subset. We will show a slight variant of that. Specifically, we show that any set of polynomials has a large subset of P_i 's such that their corresponding (Q_{i1}, \dots, Q_{ir}) are almost pairwise independent *up to a small collection of Q_1^*, \dots, Q_c^* that appear in all of them*. In other words, the tuples (Q_{i1}, \dots, Q_{ir}) form a *sunflower* whose petals are almost pairwise independent.

In order to quantify *almost independence* for a collection of polynomials, we use the notion of *regularity*, which has been studied in the context of structure versus randomness of polynomials [GT09, KL08]. Using similar machinery, we show that any collection of low degree polynomials contains a large subset that form a sunflower, whose petals are pairwise regular polynomials in the sense described above. This is of independent interest to us, and, to the best of our knowledge, has not appeared previously in the literature. While sunflowers have appeared before in related contexts (see, e.g., [FLRS23, YZ24]), our work appears to be the first to study such a sunflower structure for polynomials.

1.3 Organization of the paper

In [Section 2](#) we give a broad overview of our results and techniques. [Section 3](#) fixes our notation, introduces definitions, and presents several basic lemmas that will be used throughout the paper. In [Section 4](#) we prove our *Chebyshev lemma* and some of its variants, which will be used in the subsequent sections. In [Section 5](#) we prove our key combinatorial lemma on *sunflower pairwise regularization* that is used in proving our sampling lower bound for constant degree distributions. [Section 6](#) proves that a single polynomial cannot sample $\text{Ber}(1/3)$ better than a constant distance; this section is extended to $\text{Ber}(\rho)$ for any non-dyadic ρ in [Appendix A](#). The remaining sections, [Sections 7 to 12](#), talk about lower bounds of sampling $\text{Ber}(1/3)^{\otimes n}$ against degree 1, bounded rank_1 , degree 2, bounded rank_2 , degree 3, and general degree d distributions.

2 Our results

In this section we present our results and describe the ideas and techniques utilized to prove them. In some cases we present an informal statement in favor of a cleaner exposition. We close this section with some open problems.

2.1 A single polynomial cannot approximate $\text{Ber}(1/3)$ well

As discussed in the introduction, the following theorem lies at the heart of all of our lower bounds.

Theorem 1 (A gap for one polynomial; formal version in [Theorem 6.4](#)). *Let $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ be a polynomial of degree d . Then*

$$\left| \Pr_{X \sim \mathbb{F}_2^m} [P(X) = 1] - \frac{1}{3} \right| \geq \delta,$$

for some constant $\delta = \delta_d > 0$ that depends only on d .

Although the statement may appear almost self-evident at first glance, its proof is non-trivial even for degree as small as three. For quadratic polynomials, the proof can be obtained from the structural results of quadratic polynomials of Dickson (see, e.g., [[LN96](#), Theorem 6.30]). Thus, quite naturally, our argument ultimately relies on regularizing techniques from [[GT09](#), [KL08](#)] and the corresponding structural results. For degree three, we give an alternative proof using the specific structural results for cubic polynomials of Haramaty and Shpilka [[HS10](#)].

Remark 2.1. We remark that throughout the paper, the choice of $1/3$ is only a matter of simplicity and convenience. In [Appendix A](#), we show that $1/3$ in [Theorem 1](#) can be replaced with any non-dyadic² parameter ρ , and our sampling lower bounds for $\text{Ber}(\rho)^{\otimes n}$ hold for any $\rho \in (0, 1)$ satisfying [Theorem 1](#).

Before proceeding to other results, let us first formally define a polynomial distribution.

²A number of the form $a/2^b$ where a and b are integers is called *dyadic*.

Definition 1 (Degree- d distributions). An n -tuple $\mathcal{P} = (P_1, \dots, P_n)$ where each $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ is a polynomial of degree at most d , defines a distribution over $\{0, 1\}^n$ in the natural way: draw $X \sim \mathbb{F}_2^m$ uniformly at random, and output $(P_1(X), \dots, P_n(X)) \in \{0, 1\}^n$.

In our setting, the number of input variables m may be arbitrarily large. Observe that the polynomials P_1, \dots, P_n are fed the same input X which allows the n outputs to be correlated.

2.2 Lower bounds against linear distributions

We start with the following simple claim, showing a lower bound for degree-1 distributions. We include this result for completeness rather than for novelty.

Theorem 2 (Degree-1 distributions; formal version in [Section 7](#)). *Let $\mathcal{L} = (L_1, \dots, L_n)$ be a degree-1 distribution. Then $\|\mathcal{L} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-cn)$ for some absolute constant $c > 0$.*

The (simple) proof of this theorem illustrates a strategy that is used frequently in proving lower bounds for sampling: reducing to a win-win scenario. The argument for this theorem goes as follows. Let D be the maximum number of linearly independent L_i 's. The key is to analyze what happens when D is large and small respectively. If D is large, say, more than $n/2$, then the D linearly independent coordinates form a uniform distribution over $\{0, 1\}^D$, thus giving $1 - 2^{-\Omega(n)}$ distance from the target distribution $\text{Ber}(1/3)^{\otimes n}$. On the other hand, if D is small, then conditioning on the values of the D linearly independent L_i 's fixes the value of all other coordinates. In other words, \mathcal{L} is expressed as a convex combination of 2^D many distributions, each with singleton support. This case also gives a $1 - 2^{-\Omega(n)}$ lower bound, using a certain union bound claim for convex combinations of distributions.

Our first interesting result is the following theorem, showing a sampling lower bound for bounded rank₁ distributions. These are distributions $\mathcal{P} = (P_1, \dots, P_n)$ where each $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ can be described by a bounded number of linear/degree-1 polynomials. Already at this point, we are considering a more general class of distributions than local maps. For r -local maps the works of [\[Vio23, KOW24\]](#) essentially prove a $1 - \exp(-\Omega_r(n))$ lower bound for sampling $\text{Ber}(1/3)^{\otimes n}$. We show a similar bound for distributions whose rank₁ is upper bounded by r .

Theorem 3 (Bounded rank₁ distributions; formal version in [Theorem 8.1](#)). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be a distribution, where each $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ can be written in the form $P_i = \Gamma_i(L_{i1}, \dots, L_{ir})$, where $\Gamma_i: \{0, 1\}^r \rightarrow \{0, 1\}$ is some function and $L_{i1}, \dots, L_{ir}: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ are degree-1 polynomials. Then $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-2^{-Cr^2}n)$ for some absolute constant $C > 0$.*

The proof of the theorem is an induction on r , and has a similar flavor to the lower bound for r -local functions in [\[KOW24\]](#), generalized to account for r linear functions (rather than r input bits) using linear algebraic tools. We define a bipartite graph, where on the right vertex set we have our P_i 's, and on the left vertex set we have all linear functions. We then connect each polynomial to the subspace of dimension r that it depends on; in the notation of [Theorem 3](#), P_i is adjacent to all

$L \in \text{span}(L_{i1}, \dots, L_{ir})$. Then the argument considers a maximal subspace W (a subset of the left vertex set) whose dimension is not too large and which expands well relative to that dimension. If W is adjacent to many polynomials, then by conditioning/fixing the values of the linear functions in W we reduce the rank of its neighbors, which allows for applying induction. Otherwise, the neighborhood of W is rather small, and its maximality ensures that we can find a large subset of polynomials that depend on linearly independent subspaces, which reduces to the scenario where a large subset of P_i 's are *independent*, and hence the subset corresponds to a product distribution with each coordinate having $\Omega(2^{-r})$ distance from $\text{Ber}(1/3)$.

2.3 Lower bounds against quadratic distributions

We would like to extend our lower bounds in [Theorems 2](#) and [3](#) to quadratic distributions. Suppose we are given a distribution $\mathcal{Q} = (Q_1, \dots, Q_n)$ defined by quadratic polynomials. Note that if, say, at least $n/2$ polynomials Q_i have bounded $\text{rank}_1(Q_i)$, then we can simply apply [Theorem 3](#). Hence, it remains to handle the case, where many Q_i 's have high rank_1 .

Let us assume for simplicity that all Q_i 's have high rank_1 . Note first that for parameters r and s one of the following two cases must hold.

- (a) There is a subset $S \subseteq [n]$ of s polynomials such that $\text{rank}(Q_i + Q_j) \geq r$ for all $i \neq j \in S$.
- (b) There exists some i^* and a set $S \subseteq [n]$ of size $\geq t \geq n/s$ such that $\text{rank}(Q_{i^*} + Q_j) \leq r$ for all $j \in S$.

Indeed, this can be viewed as a graph-theoretic argument, where the vertices of the graph correspond to the n polynomials, and (i, j) is an edge if and only if $\text{rank}(Q_i + Q_j) \leq r$. Then, the first case corresponds to the graph having an independent set of size s and the second case corresponds to the graph having a vertex of degree at least n/s .



(a) **Large independent set.** The vertices form an independent set. The tuple (Q_1, \dots, Q_s) is *almost pairwise independent*.

(b) **Large degree vertex.** Each polynomial in the tuple (Q_1, \dots, Q_t) can be expressed in a new basis $Q_j = Q^* + \Gamma_j(L_{j1}, \dots, L_{jr})$, which is *essentially* low-rank up to the common polynomial Q^* . Intuitively, Q^* explains the (potentially) high-rank part of many polynomials simultaneously.

Figure 1: A win-win scenario for sampling lower bound.

Suppose first that $\text{rank}(Q_i + Q_j) \geq r$ for all $i \neq j \in S$. Then, the corresponding distribution $(Q_i)_{i \in S}$ is almost pairwise independent with each coordinate having expectation bounded away from $1/3$. Suppose for concreteness that at least $s/2$ of them have the expectation at least $1/3 + \delta$.³ In particular, on average at least $1/3 + \delta$ fraction of these polynomials output 1, and by pairwise almost independence, a standard Chebyshev-like argument implies that $(Q_i)_{i \in S}$ is $1 - o_s(1)$ -far from $\text{Ber}(1/3)^{\otimes s}$.

For the second case for each $j \in S$ we can write $Q_j = Q_{i^*} + \Gamma_j(L_{j1}, \dots, L_{jr})$ for some function $\Gamma_j: \mathbb{F}_2^r \rightarrow \mathbb{F}_2$, some linear functions L_{j1}, \dots, L_{jr} , and the common quadratic polynomial Q_{i^*} . That almost reduces it to the setting of [Theorem 3](#), except for the common part Q_{i^*} . We overcome this slight complication by taking the union bound over the conditioning on the two possible output values of Q_{i^*} , which effectively corresponds to restricting the domain of the polynomials. See [Lemma 3.6](#) for details.

By choosing the appropriate values of r and s , we obtain the following result.

Theorem 4 (Quadratic distributions; formal version in [Theorem 9.1](#)). *Let $\mathcal{Q} = (Q_1, \dots, Q_n)$ be a distribution, where each $Q_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ is a quadratic polynomial. Then $\|\mathcal{Q} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - n^{-c/\log \log(n)}$ for some absolute constant $c > 0$.*

The next natural step is to show a lower bound for bounded rank_2 distributions. Specifically, we prove the following theorem.

Theorem 5 (Bounded rank_2 distributions; formal version in [Theorem 10.1](#)). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be a distribution, where each P_i satisfies $\text{rank}_2(P_i) \leq r$, i.e., $P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir})$ for some arbitrary function Γ_i and degree-2 polynomials Q_{i1}, \dots, Q_{ir} . Then $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\log(n)^{c/r^2})$ for some absolute constant $c > 0$.*

Below we describe the ideas that go into the proof of the theorem. The main idea is to use the technique known as *regularization* in higher order Fourier analysis [[GT09](#), [HHL19](#)]. We describe the special case of it for P_i 's of bounded rank_2 . Later, we will generalize it to arbitrary degree d .

Suppose we have a collection of quadratic polynomials (Q_1, \dots, Q_r) . We define the structural notion of *regularity* as follows. A collection or *factor* of polynomials (Q_1, \dots, Q_r) is said to be k -regular if all nonzero linear combinations $\tilde{Q} = \sum_i \lambda_i Q_i$ have rank greater than k . By the structure of quadratic polynomials (Dickson's Lemma [Theorem 3.17](#)), if the factor (Q_1, \dots, Q_r) is k -regular, then all nonzero linear combinations $\sum_i \lambda_i Q_i$ are almost balanced in the sense that $\Pr_X[\sum_i \lambda_i Q_i(X) = 1] = (1 \pm 2^{-k/2})/2$. A standard Fourier-analytic argument shows that if $k \gg r$, then all nonzero linear combinations $\sum_i \lambda_i Q_i$ are almost balanced, and the distribution of (Q_1, \dots, Q_r) must be close to uniform.

Suppose now that P has $\text{rank}_2(P) \leq r$, and hence can be written as $P = \Gamma(Q_1, \dots, Q_r)$ for some quadratic polynomials Q_i . Observe that if the factor of quadratics (Q_1, \dots, Q_r) is not k -regular,

³Otherwise, at least $s/2$ of them have the expectation at most $1/3 - \delta$, and the same argument applies.

then there exists some linear combination $\sum_i \lambda_i Q_i$ whose rank is at most k , and hence we can remove one of the Q_i 's, and replace it with at most k linear functions, and hence P can be written as $P = \Gamma(Q_1, \dots, Q_{r-1}, L_1, \dots, L_k)$, where all L_i 's are linearly independent. Continuing this process, we get a *refinement* $(Q_1, \dots, Q_{r'}, L_1, \dots, L_{s'})$ consisting of $r' \leq r$ quadratic polynomials and $s' \leq kr$ linear function such that $P = \Gamma(Q_1, \dots, Q_{r'}; L_1, \dots, L_{s'})$ and $(Q_1, \dots, Q_{r'}, L_1, \dots, L_{s'})$ is k -regular.

We can now return to the proof overview of the lower bound for distributions of bounded rank₂. The proof uses induction. Let $\mathcal{P} = (P_1, \dots, P_n)$ be such a distribution, where each P_i can be written as a function of r polynomials of degree 2, which we write as $P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir})$. Let us write \mathcal{Q}_i as a shorthand for (Q_{i1}, \dots, Q_{ir}) .

We may assume that all \mathcal{Q}_i 's are k -regular for some sufficiently large k , as otherwise, we can perform a refinement step; this step will reduce the number of actual quadratics and might introduce new linear functions into \mathcal{Q}_i . Now, we construct a graph, similar to Fig. 1, where the vertices correspond to \mathcal{Q}_i 's. We put an edge between \mathcal{Q}_i and \mathcal{Q}_j if the pair $(\mathcal{Q}_i; \mathcal{Q}_j)$ is not k -regular for some parameter k , i.e. if the $2r$ -tuple $\mathcal{Q}_{i,j} = (Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr})$ is not k -regular.

Once again, the win-win scenario of Fig. 1 is applicable. If a large number $S \subseteq [n]$ satisfies the pairwise regularity condition, namely, $\mathcal{Q}_{i,j} = (Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr})$ is k -regular for all $i \neq j \in S$, then a Chebyshev-like argument implies that the corresponding distribution is $1 - o(1)$ -far from $\text{Ber}(1/3)^n$.

Otherwise, there exists some \mathcal{Q}_{i^*} and a large set $S \subseteq [n]$ such that for all $j \in S$ the factor $\mathcal{Q}_{i^*,j}$ is not regular. We show that in this case, we can change the basis of \mathcal{Q}_j , possibly introducing a few linear functions along the way, so that for all $j \in S$ the function P_j can now be written in the form $P_j = \Gamma'_j(Q^*; Q_{j1}, \dots, Q_{j,r-1}; \mathcal{L}_j)$ where \mathcal{L}_j is a collection of linear functions, and Q^* is a quadratic polynomial common to all $j \in S$. This step generalizes Fig. 1b from a single polynomial to factors of polynomials. In this new basis, the polynomials P_1, \dots, P_i have *essentially* had their quadratic rank reduced to $r - 1$, which allows us to invoke the induction hypothesis after handling Q^* ; see the proof of Theorem 10.1 for the details.

2.4 Lower bounds against polynomial distributions of degree greater than 2

In the next step we lift Theorem 5 to a lower bound for cubic distributions by an argument analogous to the one used to derive the lower bound for quadratics from bounded rank₁. We obtain the following concrete bound for degree three.

Theorem 6 (Cubic distributions; formal version in Theorem 11.1). *Let $\mathcal{C} = (C_1, \dots, C_n)$ be a distribution, where each $C_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ is a polynomial of degree 3. Then $\|\mathcal{C} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-c\sqrt{\log(n)})$ for some absolute constant $c > 0$.*

In principle, one could continue in this manner and obtain concrete bounds for higher-degree distributions, but at this point deriving an explicit lower bound $1 - \varepsilon_d(n)$ requires a lengthy calculation, while the gain over the simpler bound $1 - o_n(1)$ is comparatively modest. We therefore content

ourselves with this form of the result and turn to the final conclusion of the paper: a general sampling lower bound for degree- d distributions.

Theorem 7 (Degree- d distributions; formal version in [Theorem 12.2](#)). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be a distribution, where each $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ is a polynomial of degree d . Then $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - o_d(1)$.*

We now sketch the main idea behind the proof of this [Theorem 7](#). In order to describe the proof we first need the following lemma, which we call the *Chebyshev lemma*.⁴

2.5 Chebyshev lemma

Before proceeding to this lemma, we recall the definition of degree- d rank. The $\text{rank}_d(P)$ of a function $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ is the smallest positive integer k for which k polynomials Q_1, \dots, Q_k of degree d exist so that P can be written as $P = \Gamma(Q_1, \dots, Q_k)$ for some $\Gamma: \mathbb{F}_2^k \rightarrow \mathbb{F}_2$. When we speak of the rank of a polynomial of degree exactly $d + 1$, we mean its rank_d .

Lemma 1 (Chebyshev lemma; formal version in [Lemma 4.1](#)). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be such that each P_i has bounded rank_d ; i.e., $P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir})$ for some Γ_i and degree- d polynomials Q_{i1}, \dots, Q_{ir} . Suppose further that for all $i \neq j$ the distribution defined by the tuple*

$$\mathcal{Q}_{ij} = (Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr}) \quad \text{is } 2^{-r}\text{-close to uniform.} \quad (\text{Pairwise Uniform})$$

Then, $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - 2^{-r} - o(1)$.

The requirement of being close to uniform can be replaced by the structural requirement of *regularity*. A collection or *factor* of degree- d polynomials (Q_1, \dots, Q_r) is k -regular if all nonzero linear combinations $\tilde{Q} = \sum_i \lambda_i Q_i$ have rank more than k . The celebrated rank-bias theorem of [[GT09](#), [KL08](#)] gives a tradeoff between k and the bias of \tilde{Q} . More precisely, if k is large enough as a function of δ and d , then $\Pr_X[\tilde{Q}(X) = 1] = (1 \pm \delta)/2$. Hence, if the factor of polynomials (Q_1, \dots, Q_r) is k -regular, then all nonzero linear combinations of are balanced and the joint distribution (Q_1, \dots, Q_r) is close to uniform by a standard Fourier-analytic argument. It follows that for some $k = k(d, \delta, r)$ the following implies [Eq. \(Pairwise Uniform\)](#).

$$\mathcal{Q}_{ij} = (Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr}) \quad \text{is } k\text{-regular.} \quad (\text{Pairwise Regularity})$$

Loosely speaking, our Chebyshev lemma enables us to prove a sampling lower bound by applying induction on the rank_d . We explained this idea to some extent in our lower bound for bounded quadratic rank. If we have a distribution $\mathcal{P} = (P_1, \dots, P_n)$ such that each P_i has bounded rank_d , then either

- (a) there exists a large subset of coordinates satisfying [Eq. \(Pairwise Regularity\)](#), which puts us in the setting of [Lemma 1](#); or

⁴The name is due to Chebyshev's inequality playing the main role in its proof.

- (b) there exists a low-rank structure, which allows us to reduce a certain parameter (roughly corresponding to r), and hence apply the induction hypothesis.

The main tool that makes the Chebyshev lemma useful is *regularization*. We have already discussed regularization in part when considering distributions of bounded quadratic rank. In full generality, regularizing a factor $\mathcal{Q} = (Q_1, \dots, Q_r)$ of degree- d polynomials $Q_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$, is the process of producing another factor of degree- d polynomials $\mathcal{Q}' = (Q'_1, \dots, Q'_{r'})$ that enjoys the following properties.

- (*Regularity*) Each linear combination $\sum_i \lambda_i Q'_i$ has high rank.
- (*Refinement*) Everything that was computable by \mathcal{Q} , is also computable by \mathcal{Q}' . Formally, if $P = \Gamma(Q_1, \dots, Q_r)$ for some Γ , then $P = \Gamma'(Q'_1, \dots, Q'_{r'})$ for some Γ' .
- (*Small size blow-up*) The number of polynomials in \mathcal{Q}' is not much larger than \mathcal{Q} . More precisely, r' depends only on r, d , and the regularity parameter k , and it is independent of the number of variables m .

We now sketch the main idea behind the proof of [Theorem 7](#). The proof will be similar to [Theorem 5](#) for distributions with bounded quadratic rank. Roughly speaking, the proof of [Theorem 7](#) works by formulating a generalized version of the idea depicted in [Fig. 1](#) as a separate structural theorem, which we call *sunflower pairwise regularization*. This structural result is of independent interest to us, and may be useful beyond the present argument.

Theorem 8 (Sunflower pairwise regularization; formal version in [Theorem 5.2](#)). *Let (P_1, \dots, P_n) be a collection of degree- d polynomials $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$. Then there exists a subset $S \subseteq [n]$ of polynomials of size $w = n^{\Omega(1)}$ such that for each $i \in S$ the following holds.*

- Every P_i can be written in the form $P_i = \Gamma_i(Q_1^*, \dots, Q_c^*; Q_{i1}, \dots, Q_{ir})$ where Q_1^*, \dots, Q_c^* appear in all of the P_i and Q_{i1}, \dots, Q_{ir} is unique to P_i . That is, the sets $\mathcal{G}_i = \{Q_1^*, \dots, Q_c^*, Q_{i1}, \dots, Q_{ir}\}$ form a sunflower with the core $\mathcal{Q}^* = (Q_1^*, \dots, Q_c^*)$ and petals $\mathcal{Q}_i = (Q_{i1}, \dots, Q_{ir})$.
- The number $r + c$ is small; in particular, it is independent of m and n .
- For all $i \neq j$ the tuple $\mathcal{Q}_{ij} = (Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr})$ satisfies [Eq. \(Pairwise Regularity\)](#).

With [Theorem 8](#) in hand, the only remaining issue is handling the common polynomials Q_1^*, \dots, Q_c^* without which we can apply our Chebyshev lemma ([Lemma 1](#)) to the pairwise regular set S . In a separate lemma we prove that because $c \ll w$, the effect of Q_i^* 's is not significant (see [Lemma 3.6](#)), which concludes the proof of our lower bound for degree- d distributions.

We close our discussion of the results by turning to the proof of [Theorem 8](#). The goal in this theorem is to write P_i in the form $P_i = \Gamma_i(\mathcal{Q}^*; \mathcal{Q}_i)$ satisfying the three conditions. At the beginning, such a representation exists trivially by letting \mathcal{Q}^* to be empty and \mathcal{Q}_i to be simply (P_i) ; however, [Eq. \(Pairwise Regularity\)](#) is not necessarily satisfied. At this point, we construct a graph similar to [Fig. 1](#) connecting \mathcal{Q}_i to \mathcal{Q}_j whenever the pair $(\mathcal{Q}_i; \mathcal{Q}_j)$ is not regular enough. Then if we find a large independent set in this graph ([Fig. 1a](#)) we simply stop and announce the corresponding independent set as our final subset of polynomials. Otherwise, a vertex of large degree exists as in

Fig. 1b, which allows for a change of basis, that introduces a new common polynomial into the core \mathcal{Q}^* at the expense of introducing a few polynomials into the petals \mathcal{Q}_i . We repeat this step until we get a subset of polynomials that are pairwise regular. To ensure that \mathcal{Q}_i are individually regular, we need to interweave such steps with usual refinement/regularization steps. The main concern is why this process must terminate. The reason is that at each step, a polynomial is replaced by a few polynomials of strictly smaller degree, so the process can continue only for finitely many steps.

2.6 Open problems

Below are some problems that arise naturally from this work.

- Improve the lower bounds on the statistical distance. In particular, is it true that $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\Omega_d(n))$ for all d ?
- What about samplers defined by polynomials of degree d , where d grows as a function of n , say $d = c \log(n)$. The results of this work seem to be useful only when d is a *very* slowly growing functions of n .
- Characterize all achievable values of $\Pr_X[P(X) = 1]$, where $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ is a polynomial of degree d and m is allowed to be any positive integer. A related problem is the weight distribution of binary Reed-Muller codes; see [KLP12, KTA76, KT70].
- For degree $d = 3$ how close can $\Pr_X[P(X) = 1]$ be to $1/3$? Below we provide several examples. First consider the following polynomials with $\Pr_X[P(X) = 1] < 1/3$.
 - The simplest example is $P(x) = x_1x_2$ with $\Pr[P(x) = 1] = 1/4$. The gap is $1/3 - 1/4 = 1/12$.
 - A more interesting example is the polynomial $P(x) = x_1x_2 + x_3x_4x_5$. The first term gives $\Pr[x_1x_2 = 1] = 1/4$, and the second term gives $\Pr[x_3x_4x_5 = 1] = 1/8$. Hence $\Pr[P(x) = 1] = 1/4 + 1/8 - 2 \cdot (1/4) \cdot (1/8) = 5/16$. The gap is $1/3 - 5/16 = 1/48 \approx 0.0208$.

Next, we describe several examples with $\Pr_X[P(X) = 1] > 1/3$.

- The polynomial $P(x) = x_1 \cdot (1 - x_2x_3)$ satisfies $\Pr_X[P(X) = 1] = 1/2 \cdot 3/4 = 3/8$. The gap is $3/8 - 1/3 = 1/24 \approx 0.0417$.
- The polynomial $P(x) = x_1x_2 + x_3(x_4x_5 + x_6x_7)$. The first term $\Pr[x_1x_2 = 1] = 1/4$, and the second term gives $\Pr[x_3(x_4x_5 + x_6x_7) = 1] = 3/16$. Hence $\Pr[P(x) = 1] = 1/4 + 3/16 - 2 \cdot (1/4) \cdot (3/16) = 11/32$. The gap is $11/32 - 1/3 = 1/96 \approx 0.0104$.
- Our last example is $P(x) = x_1x_2x_3 + x_1x_4 + x_4x_5x_6 + x_7x_8x_9$. In this example, we have $\Pr[P(x) = 1] = 43/128$, and the gap is $43/128 - 1/3 = 1/384 \approx 0.0026$.

It would be interesting to better understand such examples, and compare the gaps we obtain here to the lower bound δ_3 from Lemma 6.3.

We are grateful to Mark Kahn for providing the examples above. The examples were essentially achieved using (an LLM tool which ran) a random walk (P_0, P_1, P_2, \dots) on cubic polynomials, starting with, say, $P_0(x) = 0$, and generating P_{i+1} by adding or removing a random monomial from P_i . In each step we compute $\Pr[P_i(x) = 1]$ and check how close it is to $1/3$.

3 Preliminaries

3.1 Basic definitions and notation

Polynomials. We work with multilinear polynomials $P \in \mathbb{F}_2[x_1, \dots, x_m]$ of degree d , viewed as functions $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$. It is understood that a degree- d polynomial means a polynomial of degree *at most* d unless otherwise stated.

The definition of polynomials can be extended to functions of the form $P: V \rightarrow \mathbb{F}_2$ where V is an m -dimensional vector space over \mathbb{F}_2 . We use the fact that V is isomorphic to \mathbb{F}_2^m , so after fixing some linear isomorphism $\phi: \mathbb{F}_2^m \rightarrow V$ we say P is a polynomial of degree d if the function $P': \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ given by $P'(x) = P(\phi(x))$ is a polynomial of degree d . This can be further extended in the natural way to functions $Q: W + h \rightarrow \mathbb{F}_2$ where $W + h$ is an affine subspace of V . In this case, Q is defined to be a polynomial of degree d if $Q'(x) = Q(\alpha(x) + h)$ is a polynomial of degree d , where $\alpha: \mathbb{F}_2^{\dim(W)} \rightarrow W$ is a linear isomorphism. This definition is independent of the particular choice of isomorphisms ϕ and α . The class of polynomials is closed under affine restrictions.

We now define our main object of study, polynomial distributions, sometimes also referred to as polynomial sources in the context of randomness extractors.

Definition 3.1 (Degree- d polynomial distribution). A distribution \mathcal{P} over $\{0, 1\}^n$ is a *degree- d polynomial distribution* if for some $m \in \mathbb{N}$ there exist n polynomials $P_1, \dots, P_n: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ each of degree d such that the tuple sampled according to $(P_1(X), \dots, P_n(X))$ where X is drawn uniformly at random from \mathbb{F}_2^m has the same distribution as \mathcal{P} . We assert this by writing $\mathcal{P} = (P_1, \dots, P_n)$ usually suppressing the common random input X .

Often, we will identify P with the distribution $P(X)$ it defines on a random input X , and so we treat P both as a polynomial and a random variable.

The *bias* of a polynomial P is the quantity defined as $\text{bias}(P) := |\mathbb{E}_X[(-1)^{P(X)}]|$. This quantity without the absolute values is referred to as the *signed bias* $\text{bias}_{\pm}(P)$.

Distributions. Bernoulli distribution $\text{Ber}(\rho)$ with parameter ρ will be more compactly written as \mathcal{B}_ρ . The uniform distribution on $\{0, 1\}$ is denoted by \mathcal{U} . The total variation distance between two distributions $\mathcal{D}_1, \mathcal{D}_2$ defined over the same countable domain Ω is $\|\mathcal{D}_1 - \mathcal{D}_2\|_{\text{TV}} := \frac{1}{2} \sum_{x \in \Omega} |\mathcal{D}_1(x) - \mathcal{D}_2(x)|$. Equivalently, $\|\mathcal{D}_1 - \mathcal{D}_2\| = \max_{A \subseteq \Omega} (\mathcal{D}_1(A) - \mathcal{D}_2(A))$, where $\mathcal{D}_i(A)$ is the probability that the output of \mathcal{D}_i belongs to A . If \mathcal{J} is a distribution over a set S , we denote by $\mathcal{J}^{\otimes n}$ the distribution over S^n obtained by n independent copies of \mathcal{J} . Given a collection of distributions $\{\mathcal{D}_i : i \in I\}$ over a common domain, and a distribution \mathcal{C} over the index set I , their convex combination (weighted by \mathcal{C}) is denoted by $\mathbb{E}_{i \sim \mathcal{C}}[\mathcal{D}_i]$; a sample X from $\mathbb{E}_{i \sim \mathcal{C}}[\mathcal{D}_i]$ is obtained by first sampling an index $i \sim \mathcal{C}$ and the sampling $X \sim \mathcal{D}_i$.

Other conventions. We use the standard notation $[n]$ to refer to $\{1, \dots, n\}$. We include zero in the set of natural numbers, so $\mathbb{N} = \{0, 1, \dots\}$. Throughout, \log denotes the base-2 logarithm, and $\exp(x)$ denotes 2^x . For disjoint sets A and B , the union is sometimes written as $A \sqcup B$ to highlight

disjointness. For real numbers a, b, c , whenever we write $a = b \pm c$, that means there exists some $d \in [-c, c]$ that satisfies the equality $a = b + d$. For a function $f: A \rightarrow B$ its *restriction* to $A' \subseteq A$ is the function $f|_{A'}: A' \rightarrow B$ given by $f|_{A'}(x) = f(x)$. A *dyadic number* is a rational number q that can be written in the form $a/2^b$ for some $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. We call b the granularity of this dyadic representation.

3.2 Some properties of total variation distance

A lemma we use extensively throughout the paper is the following XOR lemma, commonly attributed to Vazirani, which states that an ε -biased distribution is close to uniform whenever ε is exponentially small in the output length.

Lemma 3.2 (Vazirani's XOR lemma). *Let $\mathcal{X} = (X_1, \dots, X_n)$ be a distribution over \mathbb{F}_2^n . If for all nonzero linear functions $L: \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ it holds that*

$$\Pr[L(X_1, \dots, X_n) = 0] = \frac{1}{2} \pm \frac{\varepsilon}{2},$$

then, $\|\mathcal{X} - \mathcal{U}^{\otimes n}\|_{\text{TV}} \leq 2^{n/2}\varepsilon$.

(Proof in [Appendix B](#).)

The following lemma may be regarded as union bound for the distances between distributions.

Lemma 3.3 (Distance from convex combination). *Let \mathcal{D}, \mathcal{J} be two distributions over a finite domain X , and let $\mathcal{J} = \sum_i c_i \mathcal{J}_i$ be the convex combination of distributions $\mathcal{J}_1, \dots, \mathcal{J}_k$ according to the weights c_1, \dots, c_k where $c_i > 0$ and $\sum_i c_i = 1$. Suppose that $\|\mathcal{D} - \mathcal{J}\|_{\text{TV}} = 1 - \varepsilon_i$ for some $\varepsilon_i \in [0, 1/2]$. Then,*

$$\|\mathcal{D} - \mathcal{J}\|_{\text{TV}} \geq 1 - \sum_{i=1}^k (1 + c_i)\varepsilon_i.$$

Proof. Let $A_i = \{x \in X : \mathcal{D}(x) > \mathcal{J}_i(x)\}$. Then we have $\|\mathcal{D} - \mathcal{J}_i\|_{\text{TV}} = \mathcal{D}(A_i) - \mathcal{J}_i(A_i)$. In particular $\mathcal{D}(A_i) \geq 1 - \varepsilon_i$ and $\mathcal{J}_i(A_i) \leq \varepsilon_i$. Consider the set $A = \bigcap_{i=1}^k A_i$, and note that

$$\mathcal{D}(A) \geq 1 - \sum_{i=1}^k \mathcal{D}(X \setminus A_i) \geq 1 - \sum_{i=1}^k \varepsilon_i.$$

On the other hand,

$$\mathcal{J}(A) = \sum_{i=1}^k c_i \mathcal{J}_i(A) \leq \sum_{i=1}^k c_i \mathcal{J}_i(A_i) \leq \sum_{i=1}^k c_i \varepsilon_i.$$

Therefore,

$$\mathcal{D}(A) - \mathcal{J}(A) \geq \left(1 - \sum_{i=1}^k \varepsilon_i\right) - \sum_{i=1}^k c_i \varepsilon_i = 1 - \sum_{i=1}^k (1 + c_i)\varepsilon_i,$$

as required. \square

Lemma 3.4 (Distance between product distributions [HLS⁺25]). Let $\mathcal{D} = \mathcal{D}_1 \otimes \cdots \otimes \mathcal{D}_n$ and $\mathcal{J} = \mathcal{J}_1 \otimes \cdots \otimes \mathcal{J}_n$ be product distributions over $\{0, 1\}^n$ where all the marginals $i \in [n]$ satisfy $\|\mathcal{D}_i - \mathcal{J}_i\|_{\text{TV}} \geq \delta$. Then $\|\mathcal{D} - \mathcal{J}\|_{\text{TV}} \geq 1 - 2e^{-\delta^2 n/12}$.

Lemma 3.5 (Conditioning). Fix a mapping $f: \{0, 1\}^m \rightarrow \{0, 1\}^n$. Let $\mathbf{U}^{\otimes m}$ be the uniform distribution over $\{0, 1\}^m$ and let \mathbf{Y} be a distribution over $\{0, 1\}^n$ such that $\|f(\mathbf{U}^{\otimes m}) - \mathbf{Y}\|_{\text{TV}} \geq 1 - \varepsilon$. Let \mathbf{U}_S be the uniform distribution over $S \subseteq \{0, 1\}^m$ of size $|S| = \delta \cdot 2^m$. Then $\|f(\mathbf{U}_S) - \mathbf{Y}\|_{\text{TV}} \geq 1 - \varepsilon(1 + 1/\delta)$.

Proof. Let $A \subseteq \{0, 1\}^n$ be an event such that $\Pr[f(\mathbf{U}^{\otimes m}) \in A] - \Pr[\mathbf{Y} \in A] = \|f(\mathbf{U}^{\otimes m}) - \mathbf{Y}\|_{\text{TV}} \geq 1 - \varepsilon$. In particular, $\Pr[f(\mathbf{U}^{\otimes m}) \in A] \geq 1 - \varepsilon$ and $\Pr[\mathbf{Y} \in A] \leq \varepsilon$. By the assumption on S , we have $\Pr[f(\mathbf{U}_S) \notin A] \leq \Pr[f(\mathbf{U}^{\otimes m}) \notin A]/\delta$.

$$\begin{aligned} \|f(\mathbf{U}_S) - \mathbf{Y}\|_{\text{TV}} &\geq \Pr[f(\mathbf{U}_S) \in A] - \Pr[\mathbf{Y} \in A] \\ &= 1 - \Pr[f(\mathbf{U}_S) \notin A] - \Pr[\mathbf{Y} \in A] \\ &\geq 1 - \Pr[f(\mathbf{U}^{\otimes m}) \notin A]/\delta - \varepsilon \\ &\geq 1 - \varepsilon/\delta - \varepsilon \\ &= 1 - \varepsilon \cdot (1 + 1/\delta), \end{aligned}$$

as required. □

Lemma 3.6 (Fixing coordinates). Let $Q_1, \dots, Q_k: \{0, 1\}^m \rightarrow \{0, 1\}$ and $Q^*: \{0, 1\}^m \rightarrow S$ be $k + 1$ functions, and fix $\Gamma: \{0, 1\}^k \times S \rightarrow \{0, 1\}^n$. For $\sigma \in S$ define $\Gamma_\sigma: \{0, 1\}^k \rightarrow \{0, 1\}^n$ to be the restriction of Γ by setting the last variable to be σ , that is $\Gamma_\sigma(x_1, \dots, x_k) = \Gamma(x_1, \dots, x_k, \sigma)$.

Let X be the uniform distribution over $\{0, 1\}^m$, and let \mathbf{Y} be a distribution over $\{0, 1\}^n$ such that for all $\sigma \in S$ we have $\|\Gamma_\sigma(Q_1(X), \dots, Q_k(X)) - \mathbf{Y}\|_{\text{TV}} \geq 1 - \varepsilon$. Suppose that $\Pr[Q^*(X) = \sigma] \geq \tau$ for all $\sigma \in S$. Then

$$\|\Gamma(Q_1(X), \dots, Q_k(X), Q^*(X)) - \mathbf{Y}\|_{\text{TV}} \geq 1 - \frac{4\varepsilon|S|}{\tau}.$$

Proof. Note that

$$\Gamma(Q_1(x), \dots, Q_k(x), Q^*(x)) = \sum_{\sigma \in S} \mathbb{1}_{\{Q^*(x) = \sigma\}} \cdot \Gamma_\sigma(Q_1(x), \dots, Q_k(x)).$$

By [Lemma 3.5](#) if we condition on $Q^*(X) = \sigma$, then

$$\|\Gamma_\sigma(Q_1(X), \dots, Q_k(X)) \big|_{Q^*(X) = \sigma} - \mathbf{Y}\|_{\text{TV}} \geq 1 - \varepsilon(1 + 1/\tau).$$

Since $\Gamma(Q_1(X), \dots, Q_k(X), Q^*(X))$ is the weighted average of $\Gamma_\sigma(Q_1(X), \dots, Q_k(X))$, it follows by [Lemma 3.3](#) that

$$\begin{aligned} \|\Gamma(Q_1(X), \dots, Q_k(X), Q^*(X)) - \mathbf{Y}\|_{\text{TV}} &\geq 1 - \sum_{\sigma \in S} (1 + \Pr[Q^*(X) = \sigma]) \cdot \varepsilon(1 + 1/\tau) \\ &= 1 - (|S| + 1)\varepsilon(1 + 1/\tau) \end{aligned}$$

$$\geq 1 - 4\epsilon|S|/\tau,$$

as required. □

3.3 Rank, regularity, and factors of polynomials

Next, we introduce the standard notions relating polynomials which are also found in previous work [GT09, KL08, BHT14].

Definition 3.7 (Rank [GT09, KL08]). Let $P: \mathbb{F}^n \rightarrow \mathbb{F}$ be a function, not identically zero. The degree d rank of P , denoted by $\text{rank}_d(P)$, is the smallest positive integer k for which there exist k degree d polynomials Q_1, \dots, Q_k and a function $\Gamma: \mathbb{F}^k \rightarrow \mathbb{F}$ such that

$$P(x) = \Gamma(Q_1(x), \dots, Q_k(x)) \quad \text{for all } x.$$

We may use the terms *linear rank*, *quadratic rank*, and *cubic rank* to refer to $\text{rank}_1(P)$, $\text{rank}_2(P)$ and $\text{rank}_3(P)$ respectively. The most important notion of rank for a degree d polynomial P is $\text{rank}_{d-1}(P)$. As with degree, our convention is that a polynomial of rank r means a polynomial of rank *at most* r . We will state explicitly whenever an exact degree or rank is intended.

Fact 3.8. Let $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ be a polynomial of degree d and rank r . Then the restriction of P to any affine subspace $V + h$ is again a polynomial of degree d and rank r . In other words, upper bounds on degree and rank are preserved under affine restrictions.

Definition 3.9 (Factors). A *factor* \mathcal{F} of degree $d \in \mathbb{N}$ with *dimension vector* $\vec{M} = (M_1, \dots, M_d) \in \mathbb{N}^d$ is a tuple of polynomials $\mathcal{F} = (P_1, \dots, P_K)$ where $K = M_1 + \dots + M_d$ and for exactly M_ℓ many P_i 's we have that $\deg(P_i) = \ell$. The number K is called the *dimension* or *locality* of \mathcal{F} and is denoted by $\dim(\mathcal{F})$ or $|\mathcal{F}|$.

We equip \mathbb{N}^d with the *inverse lexicographic ordering*: for $\vec{N} = (N_1, \dots, N_d)$ and $\vec{M} = (M_1, \dots, M_d)$, we declare $\vec{N} < \vec{M}$ if, at the largest index j for which $N_j \neq M_j$, one has $N_j < M_j$.

Definition 3.10 (Refinement). Let $\mathcal{F} = (P_1, \dots, P_K)$ be a factor. We say a factor $\mathcal{G} = (Q_1, \dots, Q_{K'})$ is a *refinement* of \mathcal{F} (or that \mathcal{G} *refines* \mathcal{F}) if there exists some $\Gamma: \mathbb{F}_2^{K'} \rightarrow \mathbb{F}_2^K$ such that

$$(P_1, \dots, P_K) = \Gamma(Q_1, \dots, Q_{K'}).$$

Definition 3.11 (Rank & regularity of a factor). A factor $\mathcal{F} = (P_1, \dots, P_K)$ of degree d is said to be *r-regular* if for all nonzero linear combinations of P_i 's it holds that

$$\text{rank}_{\ell-1}(\lambda_1 P_1 + \dots + \lambda_K P_K) > r,$$

where $\ell = \max_i \deg(\lambda_i P_i)$ is the *degree of the linear combination*. Given a function $f: \mathbb{N} \rightarrow \mathbb{N}$, the factor \mathcal{F} is *f-regular* if it is $f(K)$ -regular.

The *rank of the factor* \mathcal{F} , denoted by $\text{rank}(\mathcal{F})$, is the maximal number r for which \mathcal{F} is r -regular.

Given a factor \mathcal{F} that is free of linear dependencies, for $P \in \text{span}(\mathcal{F})$ we denote by $\text{deg}_{\mathcal{F}}(P)$ the degree of the unique linear combination of polynomials in \mathcal{F} that defines P . Note that $\text{deg}(P) \leq \text{deg}_{\mathcal{F}}(P)$.

Sometimes, especially in [Section 5](#), we will view a factor \mathcal{F} as a set $\{P_1, \dots, P_K\}$. Accordingly, all the relevant definitions (e.g., dimension vector, refinement) then transfer naturally.

A *growth function* is a function $f: \mathbb{N} \rightarrow \mathbb{N}$ that is non-decreasing and satisfies $f(r) \geq r$ for all r .

Definition 3.12 (A rapidly growing function). Let $f: \mathbb{N} \rightarrow \mathbb{N}$ be a growth function and let d be a positive integer. Now define the function $\psi_{d,f}: \mathbb{N}^d \rightarrow \mathbb{N}$ recursively as follows. For the base case we have,

$$\psi_{d,f}(M_1, 0, \dots, 0) = M_1.$$

Otherwise, if $M_i > 0$ for some $i \geq 2$, then let $\ell \geq 2$ be the smallest of such indices. Then,

$$\psi_{d,f}(M_1, \dots, M_d) = \psi_{d,f}(M_1, \dots, M_{\ell-1} + f(\sum_i M_i), M_{\ell} - 1, \dots, M_d).$$

Now set $\psi_{d,f}^*(K) = \psi_{d,f}(0, \dots, 0, K)$.

Proposition 3.13. Let $\psi = \psi_{d,f}$ be defined as in [Definition 3.12](#). Then, $\psi(N_1, \dots, N_d) \leq \psi(M_1, \dots, M_d)$ whenever $\sum_i N_i \leq \sum_i M_i$ and $(N_1, \dots, N_d) \leq (M_1, \dots, M_d)$.

The proof of this proposition is deferred to the [Appendix B](#) as it is a rather tedious induction and is not a primary focus of ours. This function serves as an upper bound on some parameters in regularization processes that appear here.

Theorem 3.14 (Regularization [[GT09](#)]). Every factor \mathcal{F} of degree d admits an f -regular refinement \mathcal{G} with $\dim(\mathcal{G}) \leq \psi_{d,f}^*(\dim(\mathcal{F}))$, where $f: \mathbb{N} \rightarrow \mathbb{N}$ is a growth function.

Proof. The proof is by strong induction on the dimension vector $\vec{M} = (M_1, \dots, M_d)$ of \mathcal{F} , showing that for every \mathcal{F} there exists an f -regular refinement \mathcal{G} with $\dim(\mathcal{G}) \leq \psi_{d,f}(M_1, \dots, M_d)$. The bound $\dim(\mathcal{G}) \leq \psi_{d,f}^*(\dim(\mathcal{F}))$ follows immediately from [Proposition 3.13](#).

As the base case, the empty factor is vacuously f -regular and there is nothing to do. Now suppose that $M_i > 0$ for some i , and that the claim holds for all smaller dimensions.

We shall further assume that \mathcal{F} is free from linear dependencies, because otherwise we can discard some polynomials from \mathcal{F} and the resulting factor is smaller and still refines \mathcal{F} . If \mathcal{F} is already f -regular, then we are done. Otherwise, identify some linear combination $R = \sum_i \lambda_i P_i$ such that $\text{rank}_{\ell-1}(R) = r \leq f(K)$, where $\ell = \text{deg}_{\mathcal{F}}(R)$.

Then we replace a maximal degree P_{i^*} with $\lambda_{i^*} \neq 0$ by r many polynomials of degree at most $\ell - 1$ to obtain a new factor \mathcal{F}' . Doing so would change the dimension vector from $\vec{M} = (M_1, \dots, M_d)$ to $\vec{M}' = (M'_1, \dots, M'_d)$ such that

$$(M'_1, \dots, M'_d) < (M_1, \dots, M_{\ell-1} + f(K), M_{\ell} - 1, \dots, M_d).$$

We may now invoke the induction hypothesis on \mathcal{F}' to obtain an f -regular refinement \mathcal{G} of \mathcal{F}' . The factor \mathcal{G} is also a refinement of \mathcal{F} . Its dimension satisfies $\dim(\mathcal{G}) \leq \psi_{d,f}(\vec{M}')$ by induction. By [Proposition 3.13](#) it follows that $\psi_{d,f}(\vec{M}') \leq \psi(\vec{M})$. \square

3.4 The structure theorems and the bias-rank tradeoffs

Consider the beautiful theorem of bias implies low rank [[GT09](#), [KL08](#)].

Theorem 3.15 (Structure of Biased Polynomials [[KL08](#)]). *There exists a function $c_{\text{KL}}(d, \delta)$ such that for all polynomials P over \mathbb{F}_2 of degree at most d and $\text{bias}(P) \geq \delta > 0$ it holds that $\text{rank}_{d-1}(P) \leq c_{\text{KL}}(d, \delta)$.*

Contrapositively, this theorem says that if $\text{rank}_{d-1}(P) > c_{\text{KL}}(d, \delta)$, then $\text{bias}(P) < \delta$. We will use this fact extensively. This theorem is commonly known as *bias implies low rank* and gives a tradeoff between bias and rank. It was first proved by Green and Tao in the case of $d \leq |\mathbb{F}_p|$, but was later extended to all small prime fields by Kaufman and Lovett [[KL08](#)].

For polynomials of degree 1, 2, 3, explicit and efficient tradeoffs are known.

Fact 3.16 (Structure of Degree 1). *Given a degree-1 polynomial L , if $\text{bias}(L) > 0$ then L is constant. In other words, $c_{\text{KL}}(1, \delta) = 0$ for any positive δ .*

Theorem 3.17 (Dickson's lemma: Structure of quadratic polynomials [[LN96](#), Theorem 6.30]). *Given a quadratic polynomial Q there exists an even number r and L_1, \dots, L_{r+1} linear forms where Q can be written as*

$$Q = L_1 L_2 + \dots + L_{r-1} L_r + L_{r+1} + c,$$

where c is a field constant, and L_1, \dots, L_r are linearly independent.

We have $\text{rank}_1(Q) = r + 1$ if L_{r+1} is linearly independent from L_1, \dots, L_r , and $\text{rank}_1(Q) = r$ otherwise. In the former case $\text{bias}(Q) = 0$ and in the latter case $\text{bias}(Q) = 2^{-r/2}$. This implies that $c_{\text{KL}}(2, \delta) = 2 \log(1/\delta)$.

Theorem 3.18 (Structure of biased cubic polynomials (Haramaty–Shpilka [[HS10](#)])). *Let C be a cubic polynomial with $\text{bias}(C) \geq \delta > 0$. Then there exist linear functions $L_1, \dots, L_{r+r'}$, quadratic polynomials Q_1, \dots, Q_r , and a cubic polynomial Γ (in r' variables) such that*

$$C = \sum_{i=1}^r L_i Q_i + \Gamma(L_{r+1}, \dots, L_{r+r'}),$$

where $r = O(\log(1/\delta))$ and $r' = O(\log^4(1/\delta))$. This implies that $c_{\text{KL}}(3, \delta) = O(\log^4(1/\delta))$.

For $d \geq 4$ the dependence on δ is very weak, so we leave it unspecified. We summarize this in the following table.

	$d = 1$	$d = 2$	$d = 3$	$d \geq 4$
$c_{\text{KL}}(d, \delta)$	0	$2 \log(1/\delta)$	$O(\log^4(1/\delta))$	$O_{d,\delta}(1)$

3.5 Sunflowers for subspaces

Definition 3.19. A collection $S = \{V_1, \dots, V_s\}$ of (possibly repeated) subspaces in \mathbb{F}^m is said to be a sunflower if for $C = \bigcap_{j=1}^s V_j$ it holds that $V_i \cap V_{i'} = C$ for all $i \neq i'$.

The common intersection $C = \bigcap_{j=1}^s V_j$ is called the *core* of the sunflower.

Lemma 3.20 (Sunflower Lemma for subspaces). *Fix a finite field $\mathbb{F} = \mathbb{F}_p$ and $m > k \geq 1$. Let S be a collection of linear subspaces of \mathbb{F}^m of dimension k (the collection might have repetitions). If $|S| \geq s^{k+1} p^{\frac{k^2+k-2}{2}}$, then S contains a sunflower of size at least s .*

Remark 3.21. The proof is a straightforward adaptation of the standard proof of the lemma for set systems without a linear structure.

Note however that the bound on the size of S is better than a naive application of the sunflower lemma, as our sets are of size p^k , and so the naive bound would be double exponential in k .

Proof of Lemma 3.20. The proof is by induction on k . For the base case of $k = 1$ we either have at least s repetitions of the same subspace or we have at least $|S|/s \geq s$ different one dimensional subspaces, which all intersect only at the origin.

For the induction step, let $k \geq 2$. Let V_1, \dots, V_t be a maximal subset of S whose pairwise intersection is only $\{0\}$. If $t \geq s$, we are done, as the sets form a sunflowers with the trivial core $C = \{0\}$.

Otherwise, consider the set $X = \bigcup_{i=1}^t V_i \setminus \{0\}$, and note that $|X| = t \cdot (p^k - 1) < s \cdot p^k$. By the maximality of the subspaces V_1, \dots, V_t , it follows that $V \cap X \neq \emptyset$ for all $V \in S$. Therefore, there exists $x^* \in X \setminus \{0\}$ such that $|\{V \in S : x^* \in V\}| \geq \frac{|S|}{|X|}$.

For all $V \in S$ such that $x^* \in V$ define $V_{-x^*} \subseteq V$ to be any $k - 1$ dimensional subspace of V such that $x^* \notin V_{-x^*}$. There might be several such subspaces, and we define V_{-x^*} to be one of them arbitrarily.

Given such x^* , consider now the collection S' of subspaces of dimension $k - 1$:

$$S' = \{V_{-x^*} : x^* \in V \text{ and } V \in S\}.$$

(Note that since S might contain the same subspace more than once, the same applies also to S' .)

The collection S' contains at least $\frac{|S|}{|X|} \geq \frac{s^{k+1} \cdot p^{\frac{k^2+k-2}{2}}}{s \cdot p^k} = s^{(k-1)+1} \cdot p^{\frac{(k-1)^2+(k-1)-2}{2}}$ subspaces of dimension $k - 1$. Therefore, by the induction hypothesis S' contains a sunflower of size s , which induces a sunflower of size s in S . \square

4 Chebyshev lemma

Lemma 3.4 talks about the distance between two product distributions. Indeed, our target $\mathcal{B}_{1/3}^{\otimes n}$ is a product distribution, but in general hoping for a product structure in polynomial distributions is too much. Therefore, we prove an analog of **Lemma 3.4** assuming that our polynomials are *pairwise*

almost independent, where our notion of independence is captured by regularity.

Lemma 4.1 (Chebyshev lemma). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be an n -tuple of functions $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ such that $\text{rank}_d(P_i) \leq r$, and suppose $\|P_i - \text{Ber}(1/3)\|_{\text{TV}} \geq \delta > 0$ for all $i \in [n]$.*

Write each P_i as $P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir})$, where $\Gamma_i: \mathbb{F}_2^r \rightarrow \mathbb{F}_2$ is an arbitrary function of r variables⁵ and Q_{ij} 's are polynomials of degree d . Suppose that all pairs $i \neq j$ satisfy the pairwise regularity assumption, namely

$$\text{rank}(Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr}) > c_{\text{KL}}(d, 2^{-r} \eta).$$

Then

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - O\left(\frac{\eta}{\delta^2}\right) - O\left(\frac{1}{\delta^2 n}\right),$$

where the hidden constants in big-O are absolute.

Proof. By pigeon-hole principle, at least half of the P_i 's satisfy either $\Pr_X[P_i(X) = 1] \geq \frac{1}{3} + \delta$ or $\Pr_X[P_i(X) = 1] \leq \frac{1}{3} - \delta$. We assume the former is true for all P_i for simplicity. (We will replace n with $n/2$ for the final result. The implied constants will be absorbed by big-O.)

Associate to each P_i , the indicator random variable $\mathbb{1}_{\langle P_i \rangle}$ (to emphasize that the arithmetic involving $\mathbb{1}_{\langle P_i \rangle}$ is over \mathbb{Z} rather than \mathbb{F}_2). The key step in the proof is the following claim.

Claim 4.2. *For all $i \neq j$ we have $\mathbf{Cov}[\mathbb{1}_{\langle P_i \rangle}, \mathbb{1}_{\langle P_j \rangle}] \leq 4\eta$.*

Given **Claim 4.2** we now show that the number of polynomials outputting 1, $\sum_i \mathbb{1}_{\langle P_i \rangle}$, is bounded away above $n/3$ with high probability. Specifically,

$$\begin{aligned} \Pr \left[\sum_i \mathbb{1}_{\langle P_i \rangle} < \left(\frac{1}{3} + \frac{\delta}{2}\right)n \right] &\leq \frac{\mathbf{Var}[\sum_i \mathbb{1}_{\langle P_i \rangle}]}{(\delta/2)^2 n^2} && \text{(Chebyshev's inequality)} \\ &= \frac{\sum_i \mathbf{Var}[\mathbb{1}_{\langle P_i \rangle}] + 2 \sum_{i < j} \mathbf{Cov}[\mathbb{1}_{\langle P_i \rangle}, \mathbb{1}_{\langle P_j \rangle}]}{(\delta/2)^2 n^2} \\ &\leq \frac{n + 4\eta n^2}{(\delta/2)^2 n^2} && \text{(Claim 4.2)} \\ &= O\left(\frac{1}{\delta^2 n}\right) + O\left(\frac{\eta}{\delta^2}\right). \end{aligned}$$

Now define the event $E = \{x \in \{0, 1\}^n : \text{wt}(x) < (\frac{1}{3} + \frac{\delta}{2})n\}$ to be all outcomes with Hamming weight less than $\frac{n}{3} + \frac{\delta n}{2}$. We get,

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq \Pr[\mathcal{B}_{1/3}^{\otimes n} \in E] - \Pr[\mathcal{P} \in E] \geq 1 - \frac{2}{2^{\Omega(\delta^2 n)}} - O\left(\frac{\eta}{\delta^2}\right) - O\left(\frac{1}{\delta^2 n}\right),$$

where in the last inequality we use a standard Chernoff-type concentration of measure for $\mathcal{B}_{1/3}^{\otimes n}$ to compute $\Pr[\mathcal{B}_{1/3}^{\otimes n} \in E]$. This completes the proof of the lemma. \square

⁵Strictly speaking, we mean that for each P_i there exists some $r_i \in \{0, \dots, r\}$ such that $P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir_i})$. Since treating every tuple as having length r does not affect the argument, we make that assumption for simplicity.

We now return to the proof of [Claim 4.2](#).

Proof of Claim 4.2. Let us use \mathcal{Q}_i to denote (Q_{i1}, \dots, Q_{ir}) . By [\[KL08\]](#), the pairwise regularity assumption implies that for any nonzero linear combination of polynomials in $(\mathcal{Q}_i; \mathcal{Q}_j)$ it holds that

$$\text{bias}(\lambda_{i1}Q_{i1} + \dots + \lambda_{ir}Q_{ir} + \lambda_{j1}Q_{j1} + \dots + \lambda_{jr}Q_{jr}) \leq 2^{-r}\eta.$$

Therefore, by Vazirani's XOR [Lemma 3.2](#), the total variation distance between $(\mathcal{Q}_i; \mathcal{Q}_j)$ and the uniform distribution over $\{0, 1\}^{2r}$ is upper bounded by η . Consequently, sampling the tuple $(X; Y)$ uniformly at random from $\{0, 1\}^r \times \{0, 1\}^r$, we get

$$\begin{aligned} \mathbf{Cov}[\mathbb{1}_{\langle P_i \rangle}, \mathbb{1}_{\langle P_j \rangle}] &= \mathbf{Cov}[\mathbb{1}_{\langle \Gamma_i(\mathcal{Q}_i) \rangle}, \mathbb{1}_{\langle \Gamma_j(\mathcal{Q}_j) \rangle}] \\ &= \mathbf{E}[\mathbb{1}_{\langle \Gamma_i(\mathcal{Q}_i) \rangle} \cdot \mathbb{1}_{\langle \Gamma_j(\mathcal{Q}_j) \rangle}] - \mathbf{E}[\mathbb{1}_{\langle \Gamma_i(\mathcal{Q}_i) \rangle}] \cdot \mathbf{E}[\mathbb{1}_{\langle \Gamma_j(\mathcal{Q}_j) \rangle}] \\ &\leq \mathbf{E}[\mathbb{1}_{\langle \Gamma_i(X) \rangle} \cdot \mathbb{1}_{\langle \Gamma_j(Y) \rangle}] - \mathbf{E}[\mathbb{1}_{\langle \Gamma_i(X) \rangle}] \cdot \mathbf{E}[\mathbb{1}_{\langle \Gamma_j(Y) \rangle}] + 4\eta \\ &= 4\eta. \end{aligned}$$

The inequality uses the fact that $\|(X; Y) - (\mathcal{Q}_i; \mathcal{Q}_j)\|_{\text{TV}} \leq \eta$ so \mathcal{Q}_i and \mathcal{Q}_j may be replaced by uniform random variables X and Y ; this affects the value of each expectation by at most η . \square

4.1 Special variants of Chebyshev lemma

Corollary 4.3 (Chebyshev lemma; bounded rank₂). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be an n -tuple of functions of $\text{rank}_2(P_i) \leq r$. Suppose that for all $i \in [n]$ we have $P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir})$, where $\Gamma_i: \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ is an arbitrary function of r variables and Q_{ij} 's are quadratic polynomials satisfying*

$$\text{rank}_1(Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr}) > c_{\text{KL}}(2, 2^{-(C+1)r}) = 2(C+1)r$$

for all $i \neq j$. Furthermore, suppose that for each $i \in [n]$, $|\Pr[P(X) = 1] - 1/3| \geq \delta$. Then

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - O(2^{-Cr}/\delta^2) - O(1/(\delta^2n)),$$

where the hidden constants in big-O are absolute.

Proof. Recall that $c_{\text{KL}}(2, \delta) = 2 \log(1/\delta)$ by [Theorem 3.17](#). The corollary follows by applying [Lemma 4.1](#) with $\eta = 2^{-Cr}$. \square

Next we extend the lemma above to the setting Γ_i depends on a small number of quadratics and linear function. For a function $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$, we write $\text{rank}_{2,1}(P) \leq (r, s)$ if there exist quadratic polynomials Q_1, \dots, Q_r , linear polynomials L_1, \dots, L_s , and a function $\Gamma: \{0, 1\}^{r+s} \rightarrow \{0, 1\}$ such that $P = \Gamma(Q_1, \dots, Q_r; L_1, \dots, L_s)$.

Corollary 4.4 (Chebyshev lemma; bounded rank_{2,1}). Let $\mathcal{P} = (P_1, \dots, P_n)$ be an n -tuple of functions $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ with $\text{rank}_{2,1}(P_i) \leq (r, s)$, where $r \geq 1$ and $s \geq 0$. Suppose that $|\Pr[P_i(X) = 1] - 1/3| \geq \delta$ for all $i \in [n]$. Furthermore, suppose that for all $i \in [n]$ we have $P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir}; L_{i1}, \dots, L_{is})$, where $\Gamma_i: \{0, 1\}^{r+s} \rightarrow \{0, 1\}$ is an arbitrary function of $r + s$ variables, L_{ik} 's are linear functions and Q_{ik} 's are quadratic polynomials satisfying

$$\text{rank}_1(Q_{i1}, \dots, Q_{ir}; Q_{j1}, \dots, Q_{jr}) > 2(C + 3)(r + s)$$

for all $i \neq j$ and integer $C \geq 1$. Then

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - O\left(\frac{2^{2s}}{\delta^2 n^{1/(s+1)}}\right) - O(2^{-C(r+s)}/\delta^2).$$

Remark 4.5. We emphasize that for $i \neq j$ the linear functions $L_{i1}(x), \dots, L_{is}(x); L_{j1}(x), \dots, L_{js}(x)$ might have arbitrary linear dependencies.

Proof. The proof strategy is to reduce the problem to the setting of [Corollary 4.3](#).

Apply the sunflower lemma [Lemma 3.20](#) on the collection V_1, \dots, V_n of size n , where the i^{th} subspace V_i is $\text{span}(L_{i1}, \dots, L_{is})$.

There exists a subset $S \subseteq [n]$ of the V_i , of size $t \geq \left(\frac{n}{2^{\frac{s^2+s-2}{2}}}\right)^{\frac{1}{s+1}} \geq 2^{-s/2} \cdot n^{\frac{1}{s+1}}$ that forms a sunflower.

Let $V_{\text{core}} = \bigcap_{i \in S} V_i$ be the core of our sunflower, where $\dim(V_{\text{core}}) = \ell \leq s$. At this point, by a change of basis we obtain a new rank_{2,1} $\leq (r, s)$ representation of P_i 's that belong to the sunflower.

$$\forall i \in S: \quad P_i = \Gamma_i(Q_{i1}, \dots, Q_{ir}; L_{i1}, \dots, L_{is}) = \Gamma'_i(Q_{i1}, \dots, Q_{ir}; L'_{i1}, \dots, L'_{i, s-\ell}; L_1^*, \dots, L_\ell^*),$$

where (L_1^*, \dots, L_ℓ^*) is a basis of the core.

Let $\mathcal{Q}_i, \mathcal{L}'_i, \mathcal{L}^*$ denote each block of $(Q_{i1}, \dots, Q_{ir}; L'_{i1}, \dots, L'_{i, s-\ell}; L_1^*, \dots, L_\ell^*)$ respectively. Note that \mathcal{L}'_i and \mathcal{L}'_j are linearly independent for two distinct $i, j \in S$. Therefore, for any two distinct $i, j \in S$ the regularity condition of [Corollary 4.3](#) holds in the sense

$$\text{rank}(\mathcal{Q}_i, \mathcal{L}'_i, \mathcal{Q}_j, \mathcal{L}'_j) \geq 2(C + 3)(r + s).$$

It follows from [Corollary 4.3](#) that for any $z \in \mathbb{F}_2^\ell$ the distribution $\mathcal{P}_{S,z} = (\Gamma'_i(\mathcal{Q}_i, \mathcal{L}'_i, z) : i \in S)$ is $1 - O\left(\frac{2^{(s-\ell)/2}}{\delta^2 n^{1/(s+1)}}\right) - O(2^{-(C+2)(r+s)}/\delta^2)$ far from $\mathcal{B}_{1/3}^{\otimes n}$.

As our final step, we apply [Lemma 3.6](#). Since $\Pr[\mathcal{L}^* = z] = 2^{-\ell}$, we get

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - O\left(\frac{2^{s/2+3\ell/2}}{\delta^2 n^{1/(s+1)}}\right) - O(2^{-(C+2)(r+s)+2\ell}/\delta^2) \geq 1 - O\left(\frac{2^{2s}}{\delta^2 n^{1/(s+1)}}\right) - O(2^{-C(r+s)}/\delta^2). \quad \square$$

5 A sunflower pairwise regularization

The classic sunflower lemma [ER60] states that every sufficiently large family S_1, \dots, S_n of k -element sets contains a large subfamily $S_{\pi(1)}, \dots, S_{\pi(w)}$ whose members all have the same pairwise intersection; that is, $\bigcap_{i=1}^w S_{\pi(i)} = S_{\pi(j)} \cap S_{\pi(j')}$ for all $j \neq j'$. In this section, we prove a theorem about a collection of polynomials that is in some sense an analogue of the sunflower lemma for polynomials. We show that for every sufficiently large family of degree- d polynomials P_1, \dots, P_n there exists a large subfamily $P_{\pi(1)}, \dots, P_{\pi(w)}$ whose polynomials can simultaneously be written in the form $P_{\pi(i)} = \Gamma_i(Q_1^*, \dots, Q_c^*; Q_{i1}, \dots, Q_{ir})$. Here, the sets of polynomials $\mathcal{G}_i = \{Q_1^*, \dots, Q_c^*; Q_{i1}, \dots, Q_{ir}\}$ that compute $P_{\pi(i)}$ form a sunflower, and additionally enjoy some regularity conditions. In this section we restrict our attention to the combinatorial result itself, and leave its consequences for sampling lower bounds to [Section 12](#).

We will view a factor \mathcal{F} as a set of polynomials instead of a tuple. This enables us to take intersection or set difference between two factors using familiar notation. As a matter of convenience, the empty factor \emptyset is considered to be infinitely regular.

The *locality* of a collection of factors $\mathcal{F}_1, \dots, \mathcal{F}_n$ is the maximum cardinality (i.e. dimension) of a factor.

Definition 5.1. Let $\mathcal{F}_1, \dots, \mathcal{F}_n$ be a collection of degree- d factors. Define the *core* $\mathcal{C} = \bigcap_i \mathcal{F}_i$. The *state* of \mathcal{F}_i is a vector $\vec{S}_i \in \mathbb{N}^{2d+1}$ defined as follows. Let $K = \max_i |\mathcal{F}_i|$ denote the locality of this collection, and set $\Delta_i = K - |\mathcal{F}_i|$ for each $i \in [n]$. Then the state of \mathcal{F}_i is the vector $\vec{S}_i = (\Delta_i, \vec{N}_0, \vec{N}_i)$, where $\vec{N}_0 \in \mathbb{N}^d$ is the dimension vector \mathcal{C} , and $\vec{N}_i \in \mathbb{N}^d$ is the dimension vector of $\mathcal{F}_i \setminus \mathcal{C}$ (so the dimension vector \mathcal{F}_i is $\vec{M}_i = \vec{N}_0 + \vec{N}_i$). The first component of each \vec{S}_i , Δ_i , will be commonly denoted as its *null component*.

The *weight* of a state \vec{S}_i , denoted by $\text{wt}(\vec{S}_i)$, is the sum of each component of \vec{S}_i .

The *state* of the collection $\mathcal{F}_1, \dots, \mathcal{F}_n$ is simply the maximum state $\vec{\mathcal{S}} = \max_i \vec{S}_i$ w.r.t. the inverse lexicographic ordering of \mathbb{N}^{2d+1} .

Note that the following equalities are forced by this definition $\text{wt}(\vec{\mathcal{S}}) = \text{wt}(\vec{S}_1) = \dots = \text{wt}(\vec{S}_n) = K$.

Recall that a collection of sets V_1, \dots, V_w (possibly repeated) form a *sunflower* if for $\mathcal{C} = \bigcap_i V_i$ it holds that $\mathcal{C} = V_i \cap V_j$ for all $i \neq j$. The set \mathcal{C} is the *core* of the sunflower. This definition allows for a collection of empty sets to form a sunflower.

Theorem 5.2 (Sunflower pairwise regularization). *Let $f: \mathbb{N} \rightarrow \mathbb{N}$ be a growth function, and let K and d be positive integers. Then there exist $\varepsilon > 0$ and K' such that the following hold.*

Let $\mathcal{F}_1, \dots, \mathcal{F}_n$ be a collection of degree- d factors with locality $|\mathcal{F}_i| \leq K$. Then for a sub-collection $\mathcal{F}_{\pi(1)}, \dots, \mathcal{F}_{\pi(w)}$ of size $w = n^\varepsilon$ there exists a collection of degree- d factors $\mathcal{G}_1, \dots, \mathcal{G}_w$ that satisfies the following.

1. (*Refinement*) For every i , \mathcal{G}_i is a refinement of $\mathcal{F}_{\pi(i)}$;
2. (*Locality*) For every i , $|\mathcal{G}_i| \leq K'$;

3. (Pairwise Regularity) For every i, j the factor $\mathcal{G}_i \cup \mathcal{G}_j$ is f -regular;
4. (Sunflower) The factors $\mathcal{G}_1, \dots, \mathcal{G}_w$ form a sunflower.

Proof. The proof goes by strong induction on the state $\vec{\mathcal{S}}$ of our collection $\mathcal{F}_1, \dots, \mathcal{F}_n$. As a base case, any state of the form $\vec{\mathcal{S}} = (K, \vec{0}, \vec{0})$ corresponds to a collection of empty factors, which vacuously satisfy all the four requirements. Now we assume that $\vec{\mathcal{S}} = (M_0, \dots, M_{2d})$ contains some nonzero component $M_i > 0$ for some $i \geq 1$.

Let $\mathcal{C} = \bigcap_i \mathcal{F}_i$ be the core, and let $\mathcal{Q}_i = \mathcal{F}_i \setminus \mathcal{C}$. We make explicit the elements of each factor, writing $\mathcal{C} = \{C_1, \dots, C_{k_0}\}$ for the core, and $\mathcal{Q}_i = \{P_{i1}, \dots, P_{ik_i}\}$ for the part outside of the core. Let $K = \max_i |\mathcal{F}_i|$ denote the current locality.

We will assume that the polynomials in each factor are linearly independent, since any linear dependencies can be discarded. This clearly does not affect the final claim.

We have three cases, in each of which after applying a refinement step the state $\vec{\mathcal{S}}$ decreases, allowing us to apply the induction hypothesis.

Case I (\mathcal{C} is not $(f(2K) + 2)$ -regular). We perform the following refinement step on all factors. Identify some nonzero linear combination $R_0 = \sum_{i=1}^{k_0} \lambda_i C_i$ whose $\text{rank}_{\ell-1}$ is $r \leq f(2K) + 2$, where $\ell = \deg_{\mathcal{C}}(R_0)$ is the degree of this linear combination. Replace a maximal degree C_i with $\lambda_i \neq 0$ with r many degree $\ell - 1$ polynomials. Note that this affects the state of all non-empty factors because \mathcal{C} is the common intersection of all of them. However, all the \vec{N}_i for $i \geq 1$ remain unaffected, and only \vec{N}_0 will be replaced by a smaller dimension vector.

In this case, *all* dimension vectors of non-empty factors decrease.

Case II (At least $n/2$ many \mathcal{F}_i are not $(f(2K) + 1)$ -regular). Let $I = \{i \in [n] : \text{rank}(\mathcal{F}_i) \leq f(2K) + 1\}$. Remove all \mathcal{F}_i with $i \in [n] \setminus I$ from the collection, and do the following for all \mathcal{F}_i with $i \in I$.

Identify some nonzero linear combination $R_i = \sum_{j=1}^{k_0} \lambda_j C_j + \sum_{j=1}^{k_i} \alpha_j P_{ij} \in \text{span}(\mathcal{F}_i)$ whose $\text{rank}_{\ell-1}$ is $r \leq f(2K) + 1$, where $\ell = \deg_{\mathcal{F}_i}(R_i)$ is the degree of this linear combination. Note that by case I, this linear combination cannot be supported entirely on \mathcal{C} ; hence there exists some P_{ij} with $\alpha_j \neq 0$. Furthermore, there must exist such a P_{ij} that has maximal degree ℓ in this linear combination. Replace a P_{ij} with $\alpha_j \neq 0$ and maximal degree with r many degree $\ell - 1$ many polynomials.

In this case, the states of all surviving factors, namely those in I , decrease.

Case III ($\mathcal{F}_i \cup \mathcal{F}_j$ is not $f(2K)$ -regular for some i, j). Because we apply case II prior to this case, there can be at most $n/2$ many \mathcal{F}_i that are not $f(2K) + 1$ regular. Remove all such \mathcal{F}_i 's from our collection, keeping at least $n' = n/2$ many $f(2K) + 1$ regular \mathcal{F}_i s.

Then, build a graph G that captures *pairwise irregularities*. More formally, G , defined on the vertex set $[n']$, has an edge $e = uv$ iff $\mathcal{F}_u \cup \mathcal{F}_v$ is *not* $f(2K)$ -regular. Now we label each edge $e = uv$ by a pair of polynomials (R_u^e, R_v^e) , where $R_u^e \in \text{span}(\mathcal{F}_u)$ and $R_v^e \in \text{span}(\mathcal{F}_v)$ such that $\text{rank}_{\ell-1}$ of $R_u^e + R_v^e$ is at most $f(2K)$, where $\ell = \max\{\deg_{\mathcal{F}_u}(R_u^e), \deg_{\mathcal{F}_v}(R_v^e)\}$ is the degree of the linear combination⁶

⁶Recall that the degree of a linear combination $\sum_i \lambda_i P_i$ is the maximum degree among the polynomials that participate

that defines $R_u^e + R_v^e$. Note that neither of these polynomials are constant, and neither belong to $\text{span}(\mathcal{C})$.

If G contains an independent set of size $\sqrt{n'}$, we declare the corresponding sub-collection of \mathcal{F}_i 's to be pairwise $f(2K)$ -regular, and we set all the remaining \mathcal{F}_i (those outside the independent set) to \emptyset . This is where the induction terminates.

Otherwise, G has a vertex of degree at least $\sqrt{n'}$. Up to relabeling, we assume this vertex is 1, and is adjacent to each of $2, 3, \dots, \sqrt{n'} + 1$. By the pigeonhole principle, in at least $t = \sqrt{n'}/2^K$ many edges $e = 1v$, the same polynomial $R_1^e =: R_1$ appears. Upon relabeling once again, assume that the first t neighbors $2, \dots, t + 1$ have that property. Now for each $v = 2, \dots, t + 1$ the edge $e = 1v$ is labeled by a pair (R_1, R_v) . Recalling how we defined the labels of edges, we write

$$R_i = \underbrace{\sum_{j=1}^{k_0} \lambda_{ij} C_j}_{=: A_i} + \underbrace{\sum_{j=1}^{k_i} \alpha_{ij} P_{ij}}_{=: B_i} \quad \text{for } i = 1, 2, \dots, t + 1.$$

Let $\ell_i = \deg_{\mathcal{F}_i}(R_i)$ denote the degree of the linear combination that defines R_i . For each $i = 1, \dots, t + 1$, we have that $A_1 + B_1 + A_i + B_i$ can be written as a bounded combination of the form,

$$A_1 + B_1 + A_i + B_i = \Gamma_i(Q_{i1}, \dots, Q_{ir_i}),$$

for some $r_i \leq f(2K)$, and Q_{i1}, \dots, Q_{ir_i} degree $\max\{\ell_1, \ell_i\} - 1$ polynomials. Both B_1 and B_i must contain a degree- $\max\{\ell_1, \ell_i\}$ polynomial: indeed, if, for example, B_i does not contain such a maximal-degree polynomial, then $A_1 + B_1 + A_i \in \text{span}(\mathcal{F}_1)$ has rank more than $f(2K) + 1$, and therefore, $A_1 + B_1 + A_i + B_i$ has rank more than $f(2K)$, a contradiction. Therefore, $\ell_1 = \ell_i =: \ell$.

For each $i = 1, \dots, t + 1$, let $H_i = P_{ij}$ for some j such that $\alpha_{ij} \neq 0$ and P_{ij} has maximal degree ($= \ell$). We will remove H_i from \mathcal{F}_i and then add polynomials $A_1 + B_1, Q_{i1}, \dots, Q_{ir_i}$. Therefore, deleting all factors \mathcal{F}_i with $i \notin [t + 1]$ from our collection, we refine the $t + 1$ factors as such,

$$\mathcal{F}_i \mapsto \mathcal{F}'_i = (\mathcal{F}_i \setminus \{H_i\}) \cup \{A_1 + B_1, Q_{i1}, \dots, Q_{ir_i}\} \quad \text{for } i = 1, 2, \dots, t + 1.$$

Combining cases I, II, and III. After performing one of the refinement steps to the collection $\mathcal{F}_1, \dots, \mathcal{F}_n$ with state $\vec{\mathcal{S}}$, we have obtained a refinement of a sub-collection collection $\mathcal{F}'_1, \dots, \mathcal{F}'_t$ with the new state $\vec{\mathcal{S}}'$, where $t \geq \sqrt{n}/2^{K+1}$.

We invoke the induction hypothesis to obtain $\mathcal{G}_1, \dots, \mathcal{G}_w$ with the state $\vec{\mathcal{J}}$, which satisfies $\text{wt}(\vec{\mathcal{J}}) \leq \psi(\vec{\mathcal{S}}')$, where $\psi = \psi_{2d+1, g}$ for $g(r) = f(2r) + 2$. Let j be such that $\vec{S}'_j = \max_i \vec{S}'_i = \vec{\mathcal{S}}'$. Then,

$$\text{wt}(\vec{\mathcal{J}}) \stackrel{\text{IH}}{\leq} \psi(\vec{\mathcal{S}}') = \psi(\vec{S}'_j) \stackrel{3.13}{\leq} \psi(\vec{S}_j) \stackrel{3.13}{\leq} \psi(\vec{\mathcal{S}}) \stackrel{3.13}{\leq} \psi^*(K).$$

in it, i.e., $\max_i \deg(\lambda_i P_i)$.

Because the induction terminates in at most $\psi^*(K)$ steps, $w \geq n^{\exp(-\psi^*(K))} 2^{-2(\psi^*(K)+1)} \geq n^\varepsilon$ for large enough n and some appropriate $\varepsilon > 0$. \square

6 One polynomial's distance from Ber(1/3)

In this section we show that a degree- d polynomial $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ satisfies

$$|\Pr_X[P(X) = 1] - 1/3| \geq \delta_d,$$

where $\delta_d > 0$ is a constant depending only on d . A more general claim is true: for any non-dyadic $\rho \in (0, 1)$ and degree- d polynomial P it holds that $|\Pr_X[P(X) = 1] - \rho| \geq \Omega_{d,\rho}(1)$. We show this in [Appendix A](#). A closely related problem is the weight distribution of binary Reed-Muller codes, which asks how many polynomials of degree d satisfy $\Pr_X[P(X) = 1] \leq \eta$ for some parameter η . This theorem shows there are no polynomials with $\Pr_X[P(X) = 1]$ close to $1/3$ exist.

A similar bound for a d -local function (trivially) holds, which has proved useful in showing lower bounds for distributions defined by local functions and bounded depth decision forests [[KOW24](#), [Vio23](#)] against $\mathcal{B}_{1/3}^{\otimes n}$ or the Hamming slice $H_{n/3} = \{x \in \{0, 1\}^n : \sum_i x_i = n/3\}$.

We divide the results of this section into two parts. First we deal with small degrees 1, 2, and 3. Then we move to a general constant d . These results rely on the structure theorems discussed in [Section 3.4](#).

6.1 One polynomial's gap: degree 1, 2, 3

A degree-1 polynomial L satisfies $\Pr_X[L(X) = 1] \in \{0, \frac{1}{2}, 1\}$. Therefore, $|\Pr_X[L(X) = 1] - \frac{1}{3}| \geq \delta_1$ with $\delta_1 = 1/6$.

The structure theorem [Theorem 3.17](#) implies that a quadratic polynomial Q satisfies $\text{bias}(Q) \in \{0\} \cup \{\pm 2^{-k} : k \in \mathbb{N}\}$. Therefore $\delta_2 = 1/24$.

Below we show that all polynomial of degree 3 we have $|\Pr_X[P(X) = 1] - \frac{1}{3}| > \delta_3$ for some absolute constant $\delta_3 > 0$. The key claim is the following statement.

Claim 6.1. *Let $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ be a function of $\text{rank}_2(P) \leq r$. That is, it can be written as $P(x) = \Gamma(Q_1(x), \dots, Q_r(x))$ for some boolean $\Gamma: \{0, 1\}^r \rightarrow \{0, 1\}$ and some quadratic polynomials $Q_1, \dots, Q_r: \{0, 1\}^m \rightarrow \{0, 1\}$. For every integer $t > 0$, there exists some $s \leq (r + t) \cdot 2^r + r$ such that $|\mathbf{E}_x[(-1)^{P(x)}] - \frac{A}{2^s}| \geq \frac{1}{2^t} \cdot \frac{1}{2^s}$ for all integers $A \in \mathbb{Z}$.*

Proof. We start by writing the Fourier expansion of Γ .

$$\Gamma(y_1, \dots, y_r) = \sum_{S \subseteq [r]} \widehat{\Gamma}(S) \cdot (-1)^{\sum_{i \in S} y_i},$$

where $\widehat{\Gamma}(S) = \frac{a_S}{2^r}$ for some integer $-2^r \leq a_S \leq 2^r$. Then, for $Q_S = \sum_{i \in S} Q_i$ we have

$$\mathbf{E}_x[(-1)^{\Gamma(Q_1(x), \dots, Q_r(x))}] = \sum_{S \subseteq [r]} \widehat{\Gamma}(S) \cdot \mathbf{E}_x[(-1)^{Q_S(x)}].$$

Denote $\varepsilon_S = \mathbf{E}_x[(-1)^{Q_S(x)}]$ for all $S \subseteq [r]$. Then, using [Theorem 3.17](#) we have $\varepsilon_S \in \{\pm 2^{-i} : i \in \mathbb{N}\}$. We claim that there is some $0 \leq i \leq 2^r$ such that the interval $(2^{-(r+t)(i+1)}, 2^{-(r+t)i}]$ does not contain any of $|\varepsilon_S|$. Indeed, consider the collection of disjoint intervals $\{(2^{-(r+t)(i+1)}, 2^{-(r+t)i}] : i = 0, \dots, 2^r\}$. Since there are $2^r + 1$ such intervals, and only 2^r different values of $|\varepsilon_S|$, one of the intervals does not contain any $|\varepsilon_S|$.

Next then write $\mathbf{E}[(-1)^{P(x)}]$ as follows.

$$\begin{aligned} \mathbf{E}_x[(-1)^{P(x)}] &= \mathbf{E}_x[(-1)^{\Gamma(Q_1(x), \dots, Q_r(x))}] = \sum_{S \subseteq [r]} \widehat{\Gamma}(S) \cdot \varepsilon_S \\ &= \sum_{S \subseteq [r]: \varepsilon_S > 2^{-(r+t)i}} \widehat{\Gamma}(S) \cdot \varepsilon_S + \sum_{S \subseteq [r]: \varepsilon_S < 2^{-(r+t)(i+1)}} \widehat{\Gamma}(S) \cdot \varepsilon_S \\ &= \frac{A}{2^{(r+t)i}} + \text{err}, \end{aligned}$$

where $A \in \mathbb{Z}$ is some integer, and $|\text{err}| = \sum_{S \subseteq [r]: \varepsilon_S \leq 2^{-(r+t)(i+1)}} \widehat{\Gamma}(S) \cdot \varepsilon_S \leq 2^r \cdot 2^{-(r+t)(i+1)} = \frac{1}{2^t} 2^{-(r+t)i}$.

Letting $s = (r+t)i + r \leq (r+t) \cdot 2^r + r$ we get

$$\left| \mathbf{E}_x[(-1)^{P(x)}] - \frac{A}{2^s} \right| \leq \frac{1}{2^t} \cdot \frac{1}{2^s},$$

as required. □

We will also need the following proposition.

Proposition 6.2. *We have $|\frac{1}{3} - \frac{A}{2^s}| \geq \frac{1}{3} 2^{-s}$ for any integer $s \geq 0$ and any integer $A \in \mathbb{Z}$.*

Proof. Note that it suffices to show that $|\frac{2^s}{3} - A| \geq \frac{1}{3}$, which obviously holds since 2^s is not divisible by 3 and A is an integer. □

We are now ready to prove our lemma for polynomials of degree 3.

- Lemma 6.3.**
1. *There is a constant $\delta_{2,r} > 0$ such that any function $P: \mathbb{F}_2^m \rightarrow \{0, 1\}$ of $\text{rank}_2(P) \leq r$ satisfies $|\Pr_x[P(x) = 1] - 1/3| > \delta_{2,r}$.*
 2. *There exists a constant $\delta_3 > 0$ such that for any polynomial $Q: \mathbb{F}_2^m \rightarrow \{0, 1\}$ of degree 3 satisfies $|\Pr_x[Q(x) = 1] - \frac{1}{3}| \geq \delta_3$.*

Proof. For the first item, by [Claim 6.1](#), there exists some integer $A \in \mathbb{Z}$ and s that depends only on r , such that $|\mathbf{E}_x[P(x)] - \frac{A}{2^s}| \leq \frac{1}{4} \cdot 2^{-s}$. On the other hand, using the fact that $|\frac{A}{2^s} - \frac{1}{3}| \geq \frac{1}{3} \cdot 2^{-s}$ for all

integers $s \geq 0$ and $A \in \mathbb{Z}$, it follows that

$$|\Pr_x[P(x) = 1] - \frac{1}{3}| \geq \frac{1}{12} \cdot 2^{-s}.$$

Since $s \leq (r+2) \cdot 2^r + r$, it follows that $|\Pr_x[P(x) = 1] - \frac{1}{3}| \geq \delta_{2,r}$ for $\delta_{2,r} = \frac{1}{12} \cdot 2^{-((r+2) \cdot 2^r + r)}$.

For the second item, if $|\Pr[Q(x) = 1] - \frac{1}{3}| \geq 0.06$ then we are done. Otherwise, we have $|\Pr[Q(x) = 1] - \frac{1}{2}| \geq 0.1$. By [Theorem 3.18](#) there exist some absolute constant r such that Q can be written as $Q(X) = \Gamma(Q_1(X), \dots, Q_r(X))$, where Q_1, \dots, Q_r are quadratic polynomials, and $\Gamma: \{0, 1\}^r \rightarrow \{0, 1\}$ is some function on r bits.⁷ Therefore, by the first item $\delta_3 \geq \delta_{2,r}$ for some absolute constant r , as required. \square

6.2 One polynomial's gap: degree d

Here we prove that a degree- d polynomial satisfies $|\Pr_X[P(X) = 1] - 1/3| \geq \delta_d$. In fact, we prove a more general statement.

Theorem 6.4. *Let $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ be a function with $\text{rank}_d(P) = r$. Then,*

$$|\Pr_X[P(X) = 1] - \frac{1}{3}| \geq \delta_{d,r},$$

where $\delta_{d,r} > 0$ is a constant that depends only on d and r , which can be taken to be $\exp(-\psi_{d,f}^*(r))/12$ for $f(k) = c_{\text{KL}}(d, \frac{1}{4}2^{-3k/2})$.

Remark 6.5. Specializing to $r = 1$, we obtain that the distance from $\text{Ber}(1/3)$ for degree- d polynomials is $\delta_d = \delta_{d,1}$.

Proof. Write P as a function of r degree- d polynomials: $P = \Gamma(R_1, \dots, R_r)$. Consider the degree d factor $\mathcal{F} = (R_1, \dots, R_d)$ with the dimension vector (M_1, \dots, M_d) where $\sum_i M_i = r$. By [Theorem 3.14](#), there exists a factor $\mathcal{Q} = (Q_1, \dots, Q_K)$ of dimension $K \leq \psi_{d,f}^*(r)$ that is f -regular for $f(r) = c_{\text{KL}}(d, \frac{1}{4}2^{-3r/2})$.

Every nonzero linear combination $\sum_i \lambda_i Q_i$ in \mathcal{Q} satisfies

$$\text{bias}(\lambda_1 Q_1 + \dots + \lambda_K Q_K) \leq \frac{1}{4} 2^{-3K/2}.$$

By [Lemma 3.2](#), we have that $\|\mathcal{Q} - \mathbf{U}^{\otimes K}\|_{\text{TV}} \leq \frac{1}{4} 2^{-K}$. Since $P = \Gamma(\mathcal{Q})$ for some $\Gamma: \mathbb{F}_2^K \rightarrow \mathbb{F}_2$, it follows

⁷In fact, [Theorem 3.18](#) gives a stronger structure theorem. Specifically, there exist constants c and r such that

$$P = \sum_{i=1}^{r-c} L_i Q_i + \Gamma(L_{r-c+1}, \dots, L_r),$$

where L_i 's are linear function and Q_i 's are quadratic polynomials, and Γ is some boolean function.

that

$$\Pr[P = 1] = \Pr[\Gamma(\Omega) = 1] = \Pr[\Gamma(\mathcal{U}^{\otimes K}) = 1] \pm \frac{1/4}{2^K} = \frac{a}{2^K} \pm \frac{1/4}{2^K} \quad \text{for some } a \in \mathbb{Z}.$$

On the other hand, $|\frac{a}{2^K} - \frac{1}{3}| \geq \frac{1}{3}2^{-K}$ by [Proposition 6.2](#), giving the final result

$$|\text{bias}(P) - \frac{1}{3}| \geq \frac{1/12}{2^K},$$

as required. □

7 Lower bound against linear distributions

We start the technical part with the following easy proof of [Theorem 2](#). We restate it for convenience.

Theorem 7.1 ([Theorem 2](#) restated). *Let $\mathcal{L} = (L_1, \dots, L_n)$ be a degree-1 distribution. Then*

$$\|\mathcal{L} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\Omega(n)),$$

where $\Omega()$ hides some absolute positive constant.

Proof. Let $\mathcal{L} = (L_1, \dots, L_n)$ be a distribution where each L_i has $\deg(L_i) \leq 1$, and let D be the dimension of the span of L_i 's. That is, D is the maximal number of linearly independent L_i 's.

Let $c \in (0, 1)$ be a parameter, and consider the following two cases.

Case 1 ($D \geq cn$). Assume for simplicity that L_1, \dots, L_D are linearly independent. In particular, the first D coordinates of \mathcal{L} are statistically independent, and individually their marginal distribution is either a constant or $\mathcal{B}_{1/2}$. In particular, the coordinate-wise distance is $\|L_i - \mathcal{B}_{1/3}\|_{\text{TV}} \geq 1/6$, and by [Lemma 3.4](#) we get $\|\mathcal{L} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - 2e^{-\frac{(1/6)^2 cn}{12}} \geq 1 - 2^{-\frac{(1/6)^2 cn}{12}}$, where last inequality holds for n that is larger than some absolute constant.

Case 2 ($D < cn$). Assume for simplicity that L_1, \dots, L_D are linearly independent. Hence, for any conditioning of the inputs so that $(L_1, \dots, L_D) = (a_1, \dots, a_D)$ the distribution \mathcal{L} becomes *constant*, and hence $\|\mathcal{L}|_{(L_1, \dots, L_D) = (a_1, \dots, a_D)} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - (2/3)^n$. By taking the union bound [Lemma 3.3](#) over all 2^D fixings of L_1, \dots, L_D , we get $\|\mathcal{L} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - 2^{cn} \cdot (2/3)^n$.

A straightforward calculation shows that $c \approx 0.58$ balances the error terms of the above cases $2^{-\frac{(1/6)^2 cn}{12}} = 2^{cn} \cdot (2/3)^n \leq 2^{-\Omega(n)}$ giving the desired bound. □

8 Lower bound against bounded linear rank distributions

In this section, we prove that the distance between $\mathcal{B}_{1/3}^{\otimes n}$ and any distribution $\mathcal{P} = (P_1, \dots, P_n)$ where each $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ is a function of at most r linear functions is at least $1 - \exp(-2^{-O(r^2)}n)$.

Closely related lower bounds have appeared in the literature. Viola [Vio23] shows that a depth- r decision forest cannot sample the Hamming slice $H_{n/3}$ to within total variation distance $1 - \exp(-n^{1/\exp(O(r))})$. The work of Kane, Ostuni, and Wu [KOW24] shows that r -local maps cannot sample $H_{n/3}$ or $\mathcal{B}_{1/3}^{\otimes n}$ any closer than $1 - \exp(-n \cdot 2^{-O(r^2)})$.

We show two lower bounds with different proof ideas. The first obtains an exponentially small error which is used in Section 9 to prove a lower bound for quadratic distributions. The second approach uses sunflowers and gives a weaker error bound; we nevertheless include it because it morally extends to higher-degree distributions.

8.1 A lower bound with exponentially small error

Consider a function $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ with linear rank r .⁸ Then there exist L_1, \dots, L_r linearly independent linear functions, and some function Γ such that $P = \Gamma(L_1, \dots, L_r)$. We use the notation $\mathcal{L}(P)$ to denote the r -dimensional linear space $\text{span}\{L_1, \dots, L_r\}$ on which P depends. The linear space $\mathcal{L}(P)$ is independent of the choice of L_1, \dots, L_r, Γ . Using some basic Fourier analysis, $\mathcal{L}(P)$ can be equivalently characterized as the span of all linear functions L with which P has a nonzero correlation $\mathbf{E}_X[(-1)^{P(X)-L(X)}] \neq 0$. We now set out to prove our sampling lower bound for maps of bounded linear rank.

Theorem 8.1 (Bounded linear rank). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be a distribution where each $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ satisfies $\text{rank}_1(P_i) \leq r$ for some $r \geq 1$. Furthermore, suppose that over all affine subspaces $V + h$ of \mathbb{F}_2^m the coordinate-wise distances satisfy $\|P_i|_{V+h} - \text{Ber}(1/3)\|_{\text{TV}} \geq \delta$. Then for some absolute constant $c > 0$,*

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \exp\left(-\frac{c^r \delta^{2r}}{r!} \cdot n + r\right).$$

Proof. Let $\mathcal{P} = (P_1, \dots, P_n)$ be a distribution where each $P_i: V + h \rightarrow \mathbb{F}_2$ is defined over some affine subspace $V + h$ of \mathbb{F}_2^m and satisfies $\text{rank}_1(P_i) \leq r$. We show that for some absolute constant $c > 0$ the function $\tau: \mathbb{N} \rightarrow (0, 1]$ given by $\tau(r) = c^r \delta^{2r} / r!$ satisfies the following.

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\tau(r) \cdot n + r). \quad (\star)$$

The proof is by induction on r .

For the base case of induction consider $r = 0$. In this case each P_i is constant. Hence the support of \mathcal{P} is singleton. Consequently,

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - (2/3)^n \stackrel{\text{want}}{\geq} 1 - \exp(-\tau(0) \cdot n).$$

Thus we require

$$\tau(0) \leq \log(3/2). \quad (\text{Condition I})$$

⁸We also work with restrictions of P to affine subspaces $V + h$ of \mathbb{F}_2^m . As discussed in Section 3.1, affine restrictions preserve degree, and the same is true for rank_1 .

Fix $r \geq 1$ and suppose, by the induction hypothesis, the claim (\star) holds for distributions of rank up to $r - 1$. We now proceed to the inductive step.

Construct a bipartite graph G whose left vertex set \mathcal{L} is the set of all linear functions from V to \mathbb{F}_2 and whose right vertex set $[n]$ corresponds the functions P_1, \dots, P_n . For each $i \in [n]$ in the graph G add an edge between i and every $L \in \mathcal{L}(P_i)$ on the left. We denote the neighborhood of a subset S of vertices in G by $N_G(S)$.

Let $W \subseteq \mathcal{L}$ denote a subspace of maximum dimension that satisfies

$$|N_G(W)| \geq \frac{4}{\tau(r-1)} \cdot \dim(W). \quad (1)$$

We consider two cases separately.

Case 1 ($|N(W)| \geq n/2$). In this case, we restrict our attention to the corresponding polynomials in $I := N_G(W)$. Let n' be the size of I which we know to be at least $n/2$. In contrast, [Eq. \(1\)](#) implies that $\dim(W) \leq \frac{\tau(r-1)}{4} \cdot n$, so the dimension of W is relatively small compared to n' . This allows for conditioning on the values that W takes.

Consider some linear function $\alpha: W \rightarrow \mathbb{F}_2$ which we call an *assignment*. Because every $i \in I$ has some neighbor in W , the distribution $\mathcal{P}_I \upharpoonright_\alpha = (P_i \upharpoonright_\alpha : i \in I)$ is a distribution with linear rank $r - 1$. Thus for any assignment α , induction hypothesis provides the following.

$$\|\mathcal{P}_I \upharpoonright_\alpha - \mathcal{B}_{1/3}^{\otimes n'}\|_{\text{TV}} \geq 1 - \exp\left(-\tau(r-1) \cdot n' + r - 1\right) \geq 1 - \exp\left(-\frac{\tau(r-1)}{2} \cdot n + r - 1\right).$$

Note that $\mathcal{P}_I = \mathbf{E}_\alpha[\mathcal{P}_I \upharpoonright_\alpha]$. Applying [Lemma 3.3](#) we get that

$$\begin{aligned} \|\mathbf{E}_\alpha[\mathcal{P}_I \upharpoonright_\alpha] - \mathcal{B}_{1/3}^{\otimes n'}\|_{\text{TV}} &\geq 1 - \exp\left(-\left(\frac{\tau(r-1)}{2} - \frac{\tau(r-1)}{4}\right)n + r - 1 + 1\right) \\ &\stackrel{\text{want}}{\geq} 1 - \exp(-\tau(r) \cdot n + r). \end{aligned}$$

The last inequality holds if

$$\tau(r) \leq \frac{\tau(r-1)}{4}. \quad (\text{Condition II})$$

Case 2 ($|N(W)| < n/2$). In this case, consider the subgraph H obtained by removing the vertices in $N_G(W)$. Thus the right vertex set in H is $J := [n] \setminus N_G(W)$. Observe that for every $W' \subseteq \mathcal{L}$, the following holds in graph H .

$$|N_H(W')| < \frac{4}{\tau(r-1)} \cdot \dim(W'). \quad (2)$$

Otherwise, $W'' = \text{span}(W' \cup W)$ would be a subspace satisfying [Eq. \(1\)](#) with $\dim(W'') > \dim(W)$, contradicting the maximality of W . A useful consequence of [Eq. \(2\)](#) is that for each P_i we have $|N_H(\mathcal{L}(P_i))| < \frac{4r}{\tau(r-1)}$. Construct a set $I \subseteq [n]$ according to the following process.

```

1:  $I \leftarrow \emptyset$ 
2: while  $J \neq \emptyset$  do
3:    $i \leftarrow \min J$  ▷ In fact, any index from  $J$  would do.
4:    $J \leftarrow J \setminus \{i\}$ 
5:    $I \leftarrow I \cup \{i\}$ 
6:    $V \leftarrow \text{span}(\bigcup_{i \in I} \mathcal{L}(P_i))$ 
7:    $J \leftarrow J \setminus N_H(V)$  ▷ An index  $j \in J$  survives if  $\mathcal{L}(P_j)$  is linearly independent from  $V$ .
8: end while

```

We now argue that by the end of this process, $n' := |I|$ is greater than $\frac{\tau(r-1)}{8r}n$. Otherwise by Eq. (2) $|I \cup N_H(\text{span}(\bigcup_{i \in I} \mathcal{L}(P_i)))| \leq n/2$ contradicting the termination as J starts with a size bigger than $n/2$.

The polynomials $\mathcal{P}_I = (P_i : i \in I)$ form a product distribution since by construction the non-constant P_i 's depend on linearly independent $\mathcal{L}(P_i)$'s. We apply Lemma 3.4 and we get

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq \|\mathcal{P}_I - \mathcal{B}_{1/3}^{\otimes n'}\|_{\text{TV}} \geq 1 - \exp\left(-\Omega\left(\delta^2 \frac{\tau(r-1)}{8r}n\right) + 1\right) \stackrel{\text{want}}{\geq} 1 - \exp(-\tau(r) \cdot n + r).$$

Therefore, we require that

$$\tau(r) \leq \Omega\left(\delta^2 \frac{\tau(r-1)}{8r}\right). \quad (\text{Condition III})$$

Given conditions I, II, and III, solving for $\tau(r)$ we get $\tau(r) = c^r \delta^{2r} / r!$ for some constant $c > 0$, as required. \square

Remark 8.2. Theorem 8.1 treats the coordinate-wise distance δ as a free parameter. There are, however, only two cases of interest to us. First is the most general setting, where one poses no further assumptions on the P_i beyond having bounded rank_1 . In this case, one can take $\delta = \Omega(2^{-r})$. Second, when P_i are additionally quadratic (or degree- d) polynomials in which case δ can be taken to be an absolute constant (that depends only on d).

Corollary 8.3. Let $\mathcal{P} = (P_1, \dots, P_n)$ be a distribution where each P_i satisfies $\text{rank}_1(P_i) \leq r$. Then $\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-c2^{-Cr^2}n)$ for some absolute constants $C, c > 0$.

Proof. Write $P_i = \Gamma(L_{i1}, \dots, L_{ir})$ where the L_{ij} 's are linearly independent linear functions. Then $\Pr_X[P_i(X) = 1] = a/2^r$ for some $a \in \{0, \dots, 2^r\}$. Therefore, $|a/2^r - 1/3| \geq 2^{-r-2}$. The conclusion now follows by applying Theorem 8.1 with $\delta = 2^{-r-2}$. \square

Corollary 8.4. Let $\mathcal{Q} = (Q_1, \dots, Q_n)$ be a quadratic distribution where each Q_i satisfies $\text{rank}_1(Q_i) \leq r$. Then $\|\mathcal{Q} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-c2^{-Cr \log(r)}n)$ for some absolute constants $C, c > 0$.

Proof. Polynomials are closed under affine restrictions. Therefore, Theorem 6.4 asserts that $Q_i|_{V+h}$ has a distance of $\delta = \Omega(1)$ with $\mathcal{B}_{1/3}$. The corollary follows promptly. \square

8.2 A weaker bound using sunflowers for subspaces

Theorem 8.1 has our best lower bound $1 - \exp(-\Omega_r(n))$ for bounded rank_1 distributions. We prove a similar lower bound, albeit much weaker, using sunflowers for subspaces. The reason we include this weaker bound is twofold. First, its proof is more straightforward, and in particular, is not inductive. Second, the high-level idea of this version will be morally generalized to higher degrees while it is not clear how one can generalize the previous proof to higher degrees. One can work out a similar bound for d -local distributions using the classic sunflower lemma [ER60].

To prove our result, we derive a similar variant of **Lemma 4.1** (Chebyshev lemma) for bounded linear rank functions.

Corollary 8.5 (Chebyshev Lemma; bounded rank_1). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be an n -tuple where for all $i \in [n]$ we have $P_i = \Gamma_i(L_{i1}, \dots, L_{ir})$, where Γ_i is an arbitrary function of r variables and L_{ij} 's are linearly independent linear polynomials. Suppose further that,*

$$(\mathcal{L}_i; \mathcal{L}_j) = (L_{i1}, \dots, L_{ir}; L_{j1}, \dots, L_{jr})$$

is free from linear dependencies for all pairs $i \neq j$; i.e., $\text{rank}(\mathcal{L}_i; \mathcal{L}_j) > 0$. Then $\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - O(2^r/n)$.

Proof Sketch. The proof of this corollary is essentially the same as **Lemma 4.1**, with two differences: δ , the point-wise distance is now $O(2^{-r})$, and in the **Claim 4.2** we have $\text{Cov}[\mathbb{1}_{\langle P_i \rangle}, \mathbb{1}_{\langle P_j \rangle}] = 0$ for $i \neq j$. The result follows. \square

Theorem 8.6 (Bounded linear rank; using sunflowers). *Let $\mathcal{P} = (P_1, \dots, P_n)$ be a distribution where each P_i satisfies $\text{rank}_1(P_i) \leq r$, where $r \leq c\sqrt{\log n}$ for some constant $c > 0$. Then,*

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - O\left(\frac{2^{r/2}}{n^{1/(r+1)}}\right).$$

Proof. Define for P_i the subspace V_i of dimension $\leq r$ it depends on. By the pigeonhole principle, a subset $J \subseteq [n]$ of size at least $n/(r+1)$ subspaces have the same dimension.

By **Lemma 3.20**, the family of subspaces $S = \{V_i : i \in J\}$ contains a sunflower with at least

$$s = \left(\frac{n}{(r+1) \cdot \exp\left(\frac{r^2+r-2}{2}\right)} \right)^{1/(r+1)}$$

petals. Let $I \subseteq J$ denote the sub-collection that form a sunflower, where $|I| \geq s$.

Let $C = \bigcap_{i \in I} V_i$ denote the core of the sunflower. Note that C is itself a (linear) subspace. We call a linear map $\alpha : C \rightarrow \mathbb{F}_2$ an *assignment*. Let C^* denote the set of all assignments, where $|C^*| = 2^{\dim(C)}$.

Note that after fixing the core C by a given assignment, the distribution $\mathcal{P}_I \upharpoonright_\alpha = (P_i \upharpoonright_\alpha : i \in I)$ is now

pairwise independent, and by [Corollary 8.5](#)

$$\|\mathcal{P}_I \upharpoonright_\alpha - \mathcal{B}_{1/3}^{\otimes s}\|_{\text{TV}} \geq 1 - O(2^{r-\dim(C)}/s). \quad (3)$$

Since $\mathbf{E}_{\alpha \sim C^*}[\mathcal{P}_I \upharpoonright_\alpha] = \mathcal{P}_I$, we get

$$\begin{aligned} \|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} &\geq \|\mathcal{P}_I - \mathcal{B}_{1/3}^{\otimes s}\|_{\text{TV}} \\ &= \|\mathbf{E}_{\alpha \sim C^*}[\mathcal{P}_I \upharpoonright_\alpha] - \mathcal{B}_{1/3}^{\otimes s}\|_{\text{TV}} \\ &\geq 1 - 2 \sum_{\alpha \in C^*} (1 - \|\mathcal{P}_I \upharpoonright_\alpha - \mathcal{B}_{1/3}^{\otimes s}\|_{\text{TV}}) && \text{(By Lemma 3.3)} \\ &\geq 1 - 2^{\dim(C)+1} \cdot O(2^{r-\dim(C)}/s) && \text{(By Eq. (3))} \\ &\geq 1 - O(2^r/s) \\ &= 1 - O\left(\frac{(r+1)^{1/(r+1)} \exp\left(\frac{r^2+2r-2}{2r+2}\right)}{n^{1/(r+1)}}\right) \\ &\geq 1 - O\left(\frac{2^{r/2}}{n^{1/(r+1)}}\right), && (x^{1/x} \leq e^{1/e} \text{ for } x \geq 1) \end{aligned}$$

as required. (Note that the hidden constants in big- O are absolute.) \square

9 Lower bound against quadratic distributions

Theorem 9.1. *Let $\mathcal{Q} = (Q_1, \dots, Q_n)$ be a distribution, where each Q_i is a quadratic polynomial. Then*

$$\|\mathcal{Q} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - n^{-\Omega(1/\log \log(n))}.$$

Proof. Let $r > 0$ be a parameter to be determined later.

If at least $n/2$ many quadratics have $\text{rank}_1(Q_i) \leq 4r$, then apply [Corollary 8.4](#) to get a distance of $1 - \exp(-c2^{-Cr \log r} n)$ for some constants $C, c > 0$. (\star)

Otherwise, at least $n/2$ many Q_i 's, which we assume are $Q_1, \dots, Q_{n/2}$, satisfy $\text{rank}(Q_i) > 4r$. Construct a graph $G = (V, E)$ with vertex set $V = [n/2]$. We add an edge $ij \in E$ if the corresponding polynomials Q_i, Q_j have small-rank difference, that is, whenever $\text{rank}_1(Q_i - Q_j) \leq 4r$.

Let w be another parameter whose value will be chosen later.

We consider two cases for this graph.

Case 1 (large independent set). The graph G contains an independent set $I \subseteq V$ of size w . Then by [Corollary 4.3](#) we have that

$$\|\mathcal{Q} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq \|\mathcal{Q}_I - \mathcal{B}_{1/3}^{\otimes |I|}\|_{\text{TV}} \geq 1 - O\left(\frac{1}{w}\right) - O\left(\frac{1}{2^r}\right),$$

where big-O is hiding absolute constants.

Case 2 (large degree). There is some vertex $i^* \in V$ whose degree in G is at least $\lfloor \frac{n}{2w} \rfloor$. Let $S = N_G(i^*)$ be the set of neighbors of i^* . Let s be equal to $|S|$, which is $\Omega(n/w)$.

Since the difference of Q_{i^*} and any of its neighbors has small rank, we may write

$$Q_j = Q_{i^*} + \Gamma_j(L_{j1}, \dots, L_{j4r}),$$

for all $j \in S$ and some $4r$ linear functions.

We are now in the setting of [Lemma 3.6](#). Indeed, note that for any fixed $z \in \{0, 1\}$, the distribution

$$\mathcal{R}_z := (\Gamma_j(L_{j1}, \dots, L_{j4r}) + z : j \in S)$$

has rank₁ at most $4r$ in each coordinate. Hence, by our bounded rank₁ theorem ([Theorem 8.1](#), [Corollary 8.4](#)) we have the following for some constants $c, C > 0$.

$$\|\mathcal{R}_z - \mathcal{B}_{1/3}^{\otimes s}\|_{\text{TV}} \geq 1 - \exp\left(-c2^{-Cr \log r} n/w\right). \quad (4)$$

If Q_{i^*} is constant then [Eq. \(4\)](#) is already a lower bound for \mathcal{Q} . Else, Q_{i^*} has two possible outcomes, and $\Pr[Q_{i^*} = z] \geq \tau = 1/4$ for each $z \in \text{supp}(Q_{i^*})$, and by [Lemma 3.6](#) we have

$$\|(Q_i)_{i \in S} - \mathcal{B}_{1/3}^{\otimes s}\|_{\text{TV}} \geq 1 - \frac{8}{\tau} \cdot \exp\left(-c2^{-Cr \log r} n/w\right) \geq 1 - 32 \cdot \exp\left(-c2^{-Cr \log r} n/w\right).$$

Combining cases 1 & 2. For some absolute constant $c > 0$ we have,

$$\|\mathcal{Q} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \max \left\{ \underbrace{O\left(\frac{1}{w}\right) + O\left(\frac{1}{2^r}\right)}_{\text{Case 1}}; \underbrace{32 \exp\left(-c2^{-Cr \log r} n/w\right)}_{\text{Case 2}} \right\}.$$

Since the error from Case 2 subsumes the error at (\star) , we omit the latter. It remains to choose $r = r(n)$ and $w = w(n)$ so that the maximum error is minimized.

A straightforward calculation shows that for some constant $c' > 0$, setting $r = c' \log n / \log \log n$ and $w = 2^r$, the Case 1 error term eventually becomes larger than Case 2. It then follows that

$$\|\mathcal{Q} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - O\left(\frac{1}{2^r}\right) \geq 1 - \exp(-c' \log n / \log \log n + O(1)) \geq 1 - n^{-\Omega(1/\log \log n)},$$

as required. □

10 Lower bounds against distributions of bounded quadratic rank

Theorem 10.1. *Let $r \geq 2$, and let $\mathcal{P} = (P_1, \dots, P_n)$ be an n -tuple cubic polynomials $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$, such that $\text{rank}_2(P_i) \leq r \leq \sqrt{\log \log(n)}$ for all $i \in [n]$. Suppose that $|\Pr_X[P_i(X) = 1] - 1/3| \geq \delta$ for all $i \in [n]$. Then $\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \frac{\exp(-\log(n)^{c/r^2})}{\delta^2}$ for some absolute constant $c > 0$.*

Remark 10.2. By [Lemma 6.3](#) we have an explicit lower bound $\delta \geq \delta_{2,r} = \frac{1}{12} \cdot 2^{-((r+2) \cdot 2^r + r)} > 0$. In the special case when all P_i 's are polynomials of degree 3, we take $\delta = \delta_3 > 0$ to be an absolute constant.

We will use the following notation throughout this section. For a function $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$, we write $\text{rank}_{2,1}(P) \leq (r, s)$ if there exist quadratic polynomials Q_1, \dots, Q_r , linear polynomials L_1, \dots, L_s , and a function Γ such that $P = \Gamma(Q_1, \dots, Q_r, L_1, \dots, L_s)$.

Before proving the theorem, we need the following procedure, an explicit case of *regularization* for degree 2 factors, which is summarized in the following claim.

Claim 10.3. *Fix $C > 1$. Let $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ be a function of $\text{rank}_{2,1}(P) \leq (r, s)$. Then, there exists some $0 \leq k \leq r$ such that P can be written as*

$$P(x) = \Gamma'(\mathcal{Q}(x); \mathcal{L}(x)),$$

where

- either (i) $\mathcal{Q} = \emptyset$, and (ii) \mathcal{L} is a set of linear functions of size $|\mathcal{L}| \leq (C + 1)^r \cdot (r + s) - r$.
- or (i) \mathcal{Q} is a subset of $\{Q_1(x), \dots, Q_r(x)\}$ of size $r - k$, (ii) \mathcal{L} is a set of linear functions of size $|\mathcal{L}| \leq (C + 1)^k \cdot (r + s) - r$, and (iii) $\text{rank}_1(\mathcal{Q}) > C(C + 1)^k \cdot (r + s)$.

Proof. Write $P: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$ as

$$P(x) = \Gamma(Q_1(x), \dots, Q_r(x); L_1(x), \dots, L_s(x)),$$

where Q_i 's are quadratic polynomials and L_i 's are linear functions.

Define the collections \mathcal{Q} and \mathcal{L} as in [Algorithm 1](#).

Next we show that the obtained sets \mathcal{L} and $\mathcal{Q} = \{Q_i : i \in I\}$ satisfy the properties stated in the claim.

Consider the value of k at the end of the procedure, and note that k is equal to the number of iterations performed by the while loop.

Note first that we decrease the size of I by one in each iteration, and hence in the end \mathcal{Q} is a subset of $\{Q_1(x), \dots, Q_r(x)\}$ of size $|I| = r - k$.

We show an upper bound on the size of \mathcal{L} by induction on k . For $k = 0$ we have $|\mathcal{L}| \leq s = (C + 1)^k (r + s) - r$ by the assumption of the claim. Suppose the bound $|\mathcal{L}| \leq (C + 1)^{k-1} \cdot (r + s) - r$ holds just before the last iteration. Then, in the last iteration iteration we increase the size of \mathcal{L} by at

Algorithm 1 Regularizing \mathcal{Q}

- 1: Initialize $\mathcal{L} = \{L_1, \dots, L_s\}$.
 - 2: Initialize $I = \{1, \dots, r\}$. ▷ The set $I \subseteq \{1, \dots, r\}$ will correspond to $\mathcal{Q} = \{Q_i : i \in I\}$.
 - 3: Set $k = 0$. ▷ In each iteration we will have $|I| = r - k$ and $|\mathcal{L}| \leq (C + 1)^k \cdot (r + s)$.
 - 4: **while** $I \neq \emptyset$ and $\text{rank}_1(\mathcal{Q}) \leq C(C + 1)^k \cdot (r + s)$ **do**
 - 5: Let $(a_i)_{i \in I}$ be such that $\text{rank}_1(\sum_{i \in I} a_i Q_i) < C(C + 1)^k \cdot (r + s)$.
 - 6: Augment \mathcal{L} by adding to it at most $C(C + 1)^k \cdot (r + s)$ linear functions so that $\sum_{i \in I} a_i Q_i$ can be expressed using the functions in \mathcal{L} . If \mathcal{L} has linear dependencies, remove them.
 - 7: Remove from I one of the i 's for which $a_i = 1$.
 - 8: Redefine Γ to be a function that takes \mathcal{Q} and \mathcal{L} and computes P_i .
 - 9: Increment k by 1.
 - 10: **end while**
-

most $C(C + 1)^{k-1} \cdot (r + s)$, and hence the size of \mathcal{L} is now upper bounded by

$$((C + 1)^{k-1} \cdot (r + s) - r) + (C(C + 1)^{k-1} \cdot (r + s)) = (C + 1)^k \cdot (r + s) - r.$$

Finally, by the stopping condition of the while loop, if $k = r$, then $\mathcal{Q} = \emptyset$, and otherwise $\text{rank}_1(\mathcal{Q}) \geq C(C + 1)^k \cdot (r + s)$, as required. \square

We are now ready to prove [Theorem 10.1](#).

Proof of Theorem 10.1. Let $\varepsilon(n, r, s) > 0$ be such that for any n -tuple $\mathcal{P} = (P_1, \dots, P_n)$ of functions $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$, with $\text{rank}_{2,1}(P_i) \leq (r, s)$ satisfies $\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \varepsilon(n, r, s)$. Below we derive a bound on ε using induction.

For the base case, note that by [Theorem 8.1](#) we have $\varepsilon(n, r = 0, s) \leq \exp\left(-\frac{c^s \delta^{2s}}{s!} \cdot n + s\right)$ for some absolute constant $c > 0$.

Next we bound $\varepsilon(n, r, s)$ for $r \geq 1$ and arbitrary $s \geq 0$. Let $\mathcal{P} = (P_1, \dots, P_n)$ be an n -tuple of functions $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$, such that $\text{rank}_{2,1}(P_i) \leq (r, s)$ for all $i \in [n]$.

We start by regularizing each P_i as in [Claim 10.3](#) with some parameter $C \geq 8$ to be chosen later. After the regularization, for each P_i there is some $0 \leq k_i \leq r$ such that $P_i = \Gamma_i(\mathcal{Q}^{(i)}, \mathcal{L}^{(i)})$, where $\mathcal{Q}^{(i)}$ is a set of quadratic functions of size at most $r - k_i$, $\mathcal{L}^{(i)}$ is a set of linear functions, $|\mathcal{Q}^{(i)}| + |\mathcal{L}^{(i)}| \leq (C + 1)^{k_i} \cdot (r + s)$, and $\text{rank}_1(\mathcal{Q}^{(i)}) > C \cdot (C + 1)^{k_i} \cdot (r + s)$.

Let k^* be the most popular among k_i 's, and let $S_0 \subseteq [n]$ be a subset of size $\geq n/(r + 1)$ such that $|\mathcal{Q}^{(i)}| + |\mathcal{L}^{(i)}| \leq (C + 1)^{k^*} \cdot (r + s)$ and $\text{rank}_1(\mathcal{Q}^{(i)}) > C \cdot (C + 1)^{k^*} \cdot (r + s)$ for all $i \in S_0$.

We consider the following three cases.

- Suppose that $k^* \geq 1$. Then $\text{rank}_{2,1}(P_i) \leq (r - k^*, (C + 1)^{k^*} \cdot (r + s))$ for all $i \in S_0$. And by the induction hypothesis we get

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \varepsilon\left(\frac{n}{r + 1}, r - k^*, (C + 1)^{k^*} \cdot (r + s)\right). \quad (5)$$

- Assume now that $k^* = 0$. Let $w \in \mathbb{N}$ be a parameter to be chosen later, and suppose there is a subset $S \subseteq S_0$ of size $|S| = w$ such that for all $i \neq i'$ in S it holds that $\text{rank}_1(Q^{(i)} \cup Q^{(i')}) > C(r+s)$. Then, by applying [Corollary 4.4](#) we get

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \frac{4^s}{\delta^2 \cdot w^{1/(s+1)}} - \frac{2^{-(C-3)(r+s)/2}}{\delta^2} \geq 1 - \frac{4^s}{\delta^2 \cdot w^{1/(s+1)}} - \frac{2^{-C(r+s)/4}}{\delta^2}. \quad (6)$$

- Otherwise, we have $k^* = 0$ and there exists $i^* \in [n]$ and a subset $S \subseteq [n]$ of size $|S| \geq (1/2^r) \cdot (|S_0|/w)$ such that $\text{rank}_1(Q^{(i^*)} \cup Q^{(i)}) \leq C(r+s)$, and furthermore the bounded rank is realized with the *same linear combination* in $Q^{(i^*)}$. That is, there exists some linear combination $\sum_{j=1}^r \beta_j Q_j^{(i^*)}$ in such that for each $i \in S$ we have $\text{rank}_1(\sum_{j=1}^r \alpha^{(i)} Q_j^{(i)} + \sum_{j=1}^r \beta_j Q_j^{(i^*)}) < C(r+s)$. Hence, for each $i \in S$, we can remove one of the $Q_j^{(i)}$ from $\mathcal{Q}^{(i)}$, augment $\mathcal{L}^{(i)}$ by adding to it at most $C(r+s)$ linear functions, and rewrite Γ_i as a function that depends on the new $\mathcal{Q}^{(i)}$ (of size at most $r-1$), new $\mathcal{L}^{(i)}$ and $Q^* = \sum_{j=1}^r \beta_j Q_j^{(i^*)}$. We emphasize that Q^* is the same for all $i \in S$.

That is, for each $i \in S$, we can write P_i as

$$P_i = \Gamma_i(Q_1^{(i)}, \dots, Q_{r-1}^{(i)}; \mathcal{L}^{(i)}; Q^*).$$

For $z \in \{0, 1\}$ denote by $P_i^z = \Gamma_i(Q_1^{(i)}, \dots, Q_{r-1}^{(i)}; \mathcal{L}^{(i)}; z)$. Note that $\text{rank}_{2,1}(P_i^z) \leq (r-1, s + C(r+s))$. Hence, by the induction hypothesis, we have

$$\|(P_i^z)_{i \in S} - \mathcal{B}_{1/3}^{\otimes |S|}\|_{\text{TV}} \geq 1 - \varepsilon(|S|, r-1, s + C(r+s)).$$

Therefore, since $\deg(Q^*) \leq 2$, each $z \in \text{supp}(Q^*) \subseteq \{0, 1\}$ is obtained with probability at least $1/4$, and thus, by [Lemma 3.6](#) we get

$$\|(P_i)_{i \in S} - \mathcal{B}_{1/3}^{\otimes |S|}\|_{\text{TV}} \geq 1 - 32\varepsilon\left(\frac{n}{w \cdot (r+1) \cdot 2^r}, r-1, s + C(r+s)\right). \quad (7)$$

Combining [Eqs. \(5\) to \(7\)](#), we get

$$\varepsilon(n, r, s) \leq \max\left(\max_{k^*} \varepsilon\left(\frac{n}{r+1}, r-k^*, (C+1)^{k^*}(r+s)\right); \frac{4^s}{\delta^2 \cdot w^{1/(s+1)}} + \frac{2^{-C(r+s)/4}}{\delta^2}; 32\varepsilon\left(\frac{n}{w \cdot (r+1) \cdot 2^r}, r-1, s + C(r+s)\right)\right).$$

Using monotonicity, we can bound $\varepsilon(n, r, s)$ by

$$\varepsilon(n, r, s) \leq \max\left(\frac{4^s}{\delta^2 \cdot w^{1/(s+1)}} + \frac{2^{-C(r+s)/4}}{\delta^2}; 32\varepsilon\left(\frac{n}{w \cdot (r+1) \cdot 2^r}, r-1, (C+1)^r \cdot (r+s)\right)\right) \quad (8)$$

with the base case

$$\varepsilon(n, r = 0, s) \leq \exp\left(-\frac{c^s \delta^{2s}}{s!} \cdot n + s\right). \quad (9)$$

We have the following upper bound on ε .

Claim 10.4. Let $\varepsilon(n, r, s)$ be defined inductively by Eqs. (8) and (9). Then,

$$\varepsilon(n, r, s) \leq 32^r \cdot \max \left(\mathbf{K}; \exp \left(-\frac{(c\delta)^{h^*}}{h^*!} \cdot \frac{n}{w^r \cdot (r+1)! \cdot 2^{r(r+1)/2}} + h^* \right) \right)$$

for $h^* \leq 2(C+1)r^2 \cdot (2r+s)$ and $\mathbf{K} = \frac{4^s}{\delta^2 \cdot w^{1/(s+1)}} + \frac{2^{-C(r+s)/4}}{\delta^2}$.

Next, we set the parameters $w = n^{1/2r}$ and $C = \log(n)^{1/3r^2} - 1$. Hence, if $r, s < \sqrt{\log(n) \log(n)}$, then $h^* \leq (\log(n)^{1/3r^2})^{r^2} \cdot (2r+s) = \log(n)^{\frac{1}{3}} \cdot \sqrt{\log(n) \log(n)} < \sqrt{\log(n)}$. Therefore, for $r \geq 2$ we have

$$\begin{aligned} \varepsilon(n, r, s) &\leq 32^r \cdot \max \left(\mathbf{K}; \exp \left(-\frac{(c\delta)^{h^*}}{h^*!} \cdot \frac{n}{w^r \cdot (r+1)! \cdot 2^{r(r+1)/2}} + h^* \right) \right) \\ &\leq 32^r \cdot \max \left(\frac{4\sqrt{\log(n) \log(n)}}{\delta^2 \cdot n^{\Omega(1/rs)}} + \frac{\exp(-\log(n)^{\Omega(1/r^2)})}{\delta^2}; \exp \left(-\frac{(c\delta)^{\sqrt{\log(n)}}}{\sqrt{\log(n)}!} \cdot \frac{\sqrt{n}}{2r^2} + \log(n)^{1/6} \right) \right) \end{aligned}$$

for some absolute constant $c > 0$.

Recall, by Lemma 6.3 we have $\delta \geq \frac{1}{12} \cdot 2^{-((r+2) \cdot 2^r + r)} \geq 2^{-C\sqrt{\log \log(n)}}$ for some constant C (say, $C = 6$ works). Hence the dominant term in the displayed equation about is $\frac{\exp(-\log(n)^{\Omega(1/r^2)})}{\delta^2}$. Therefore

$$\|(\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n})\|_{\text{TV}} \geq \|(P_i)_{i \in S} - \mathcal{B}_{1/3}^{\otimes |S|}\|_{\text{TV}} \geq 1 - \varepsilon(n, r, s) \geq \exp(-\log(n)^{c/r^2})/\delta^2 \quad (10)$$

for some absolute constant $c > 0$, as required. \square

We now return to the proof of Claim 10.4.

Proof of Claim 10.4. Define $h_{r,s} = \sum_{i=1}^r (C+1)^{\sum_{j=1}^i j} \cdot i + (C+1)^{\sum_{j=1}^r j} \cdot s$. Note that

- $h_{0,s} = s$,
- $h_{r,s} = h_{r-1, (C+1)^r(r+s)}$,
- $h_{r,s} \leq (C+1)^{r^2} \cdot (2r+s)$.

We prove by induction that

$$\varepsilon(n, r, s) \leq 32^r \cdot \max \left(\mathbf{K}; \exp \left(-\frac{c^{h_{r,s}} \cdot \delta^{2h_{r,s}}}{h_{r,s}!} \cdot \frac{n}{w^r \cdot (r+1)! \cdot 2^{r(r+1)/2}} + h_{r,s} \right) \right) \quad (11)$$

The claim follow by setting $h^* = 2h_{r,s}$.

For the base case of $r = 0$ we have $h_{0,s} = s$, and hence $\varepsilon(n, r = 0, s) \leq \exp(-\frac{c^s \cdot \delta^{2s}}{s!} \cdot n + s) = \exp(-\frac{c^{h_{0,s}} \cdot \delta^{2h_{0,s}}}{h_{0,s}!} \cdot n + h_{0,s})$.

For the induction step, we assume that the bound on ε holds for $r-1$, and prove it for r . Indeed, by

Eq. (8) we have

$$\begin{aligned}\varepsilon(n, r, s) &\leq \max\left(\mathbf{K}; 32 \cdot \varepsilon\left(\frac{n}{w \cdot (r+1) \cdot 2^r}, r-1, (C+1)^r \cdot (r+s)\right)\right) \\ &= \max\left(\mathbf{K}; 32 \cdot \varepsilon(n', r-1, s')\right),\end{aligned}$$

where $n' = \frac{n}{w \cdot (r+1) \cdot 2^r}$ and $s' = (C+1)^r \cdot (r+s)$. By the induction hypothesis we have

$$\varepsilon(n', r-1, s') \leq 32^{r-1} \cdot \max\left(\mathbf{K}; \exp\left(-\frac{c^{h_{r-1,s'}} \cdot \delta^{2h_{r-1,s'}}}{h_{r-1,s'}!} \cdot \frac{n'}{w^{r-1} \cdot r! \cdot 2^{(r-1)r/2}} + h_{r-1,s'}\right)\right)$$

Plugging it into the bound on $\varepsilon(n, r, s)$ we get

$$\begin{aligned}\varepsilon(n, r, s) &\leq 32^r \cdot \max\left(\mathbf{K}; \exp\left(-\frac{c^{h_{r-1,s'}} \cdot \delta^{2h_{r-1,s'}}}{h_{r-1,s'}!} \cdot \frac{n'}{w^{r-1} \cdot r! \cdot 2^{(r-1)r/2}} + h_{r-1,s'}\right)\right) \\ &= 32^r \cdot \max\left(\mathbf{K}; \exp\left(-\frac{c^{h_{r,s}} \cdot \delta^{2h_{r,s}}}{h_{r,s}!} \cdot \frac{n}{w^r \cdot (r+1)! \cdot 2^{r(r+1)/2}} + h_{r,s}\right)\right),\end{aligned}$$

where the last equality uses the fact that $h_{r,s} = h_{r-1,s'}$. This completes the proof of Eq. (11). \square

11 Lower bounds against distributions of degree three

Theorem 11.1. *Let $\mathcal{P} = (P_1, \dots, P_n)$ be an n -tuple of cubic polynomials $P_i: \mathbb{F}_2^m \rightarrow \mathbb{F}_2$. Then*

$$\|\mathcal{P} - \mathfrak{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-c\sqrt{\log \log(n)}).$$

Proof. We have several cases.

- There is a subset $S_0 \subseteq [n]$ of size at least $n/2$ such that $\text{rank}_2(P_i) \leq r$ for all $i \in S_0$. This reduces to [Theorem 10.1](#), and we get $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\log(n)^{c/r^2})$.
- Otherwise, there is a subset $S_0 \subseteq [n]$ of size at least $n/2$ such that $\text{rank}_2(P_i) > r$ for all $i \in S_0$. Is there is $S \subseteq S_0$ of size at least w such that $\text{rank}_2(P_i + P_{i'}) > r$ for all $i \neq i' \in S$. The we apply Chebyshev, and get $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \frac{c}{w} - \frac{c}{\exp(r)}$.
- Otherwise, there exists some $i^* \in S_0$ and $S \subseteq S_0$ of size at least $|S_0|/w$ such that $\text{rank}_2(P_i + P_{i^*}) \leq r$ for all $i \in S$. This means for all $i \in S$ we can write $P_i = \Gamma_i(\mathbf{Q}^{(i)}) + P_{i^*}$ for some $\Gamma_i: \{0,1\}^r \rightarrow \{0,1\}$ and $\mathbf{Q}^{(i)}$ consisting of r quadratic polynomials. This gives the bound $\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} \geq 1 - \exp(-\log(n/w)^{c/r^2})$.

Taking $w = \sqrt{n}$ and $r = c\sqrt{\log \log(n)}$ for some $0 < c < 1/4$, we get that

$$\begin{aligned}\|\mathcal{P} - \text{Ber}(1/3)^{\otimes n}\|_{\text{TV}} &\geq 1 - \max\left\{\exp(\log(n)^{c/r^2}); \frac{c}{w} + \frac{c}{\exp(r)}; \exp(-\log(n/w)^{c/r^2})\right\} \\ &= 1 - \frac{c}{\exp(r)}\end{aligned}$$

$$= 1 - \exp(-c\sqrt{\log \log(n)})$$

for some absolute constant $c > 0$. □

12 Lower bound against degree d distributions

In this section we show that any polynomial distribution \mathcal{P} of degree d is $1 - o_n(1)$ far from $\mathcal{B}_{1/3}^{\otimes n}$, where $o_n(1)$ is a vanishing function of n for any constant d .

We begin with the following proposition, and then proceed to the proof of the main theorem of this section.

Proposition 12.1. *Let $\mathcal{F} = (P_1, \dots, P_K)$ be a degree d factor whose rank is $c_{\text{KL}}(d, 2^{-K}\eta)$ for some $\eta \in (0, 1]$, and let Z denote the set of common zeros of the P_i . Then the density of Z , i.e., $\Pr_X[X \in Z]$, is equal to $2^{-K} \pm \eta$.*

Proof. Observe that $\Pr[P_1(X) = \dots = P_K(X) = 0] = \mathbf{E}[(-1)^{Y_1 P_1(X) + \dots + Y_K P_K(X)}]$ where X is drawn uniformly at random and the $Y_i \sim \mathcal{U}$ are i.i.d. variables mutually independent from X . Then we have the following.

$$\begin{aligned} \mathbf{E}[(-1)^{Y_1 P_1 + \dots + Y_K P_K}] &= 2^{-K} \sum_{S \subseteq [K]} \mathbf{E} \left[(-1)^{\sum_{i \in S} P_i} \right] \\ &= 2^{-K} + \sum_{S \subseteq [K], S \neq \emptyset} (\pm 2^{-K} \eta) \\ &= 2^{-K} \pm \eta. \end{aligned} \quad \square$$

Theorem 12.2 (Degree- d). *For every $0 < \varepsilon \leq 1$ and $d \in \mathbb{N}$ there exists some $\varepsilon', d' > 0$ such that the following holds. Let $\mathcal{P} = (P_1, \dots, P_n)$ be a distribution generated by polynomials of degree d . Then,*

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - O(d'n^{-\varepsilon'}) - O(\varepsilon),$$

where the hidden constants in big O are absolute. It follows that,

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - o_n(1),$$

where $o_n(1)$ is a vanishing function of n for any constant d .

Proof. Define the collection of factors $\mathcal{F}_1, \dots, \mathcal{F}_n$ with $\mathcal{F}_i = \{P_i\}$. Let $f: \mathbb{N} \rightarrow \mathbb{N}$ be given by $f(r) = c_{\text{KL}}(d, 2^{-3r} \varepsilon \delta^2)$, where $\delta = \delta_d$ is from [Theorem 6.4](#).

We invoke [Theorem 5.2](#) to obtain a collection of factors $(\mathcal{G}_i : i \in I)$ where $I \subseteq [n]$ has size at least $|I| = s \geq n^{\varepsilon'}$, and \mathcal{G}_i are pairwise f -regular. For simplicity, assume that $I = [s]$. Note that [Theorem 5.2](#) guarantees $|\mathcal{G}_i| \leq r$ where r only depends on d, ε .

Let $\mathcal{C} = \bigcap_{i \in I} \mathcal{G}_i = \{C_1, \dots, C_\ell\}$ be the common intersection, and write each \mathcal{G}_i as the disjoint union $\mathcal{C} \sqcup \mathcal{Q}_i$. Now we can write the joint distribution of the set I as,

$$\mathcal{P}_I = (P_1, \dots, P_s) = (\Gamma_1(\mathcal{C}, \mathcal{Q}_1), \dots, \Gamma_s(\mathcal{C}, \mathcal{Q}_s)).$$

Let $\Gamma(\mathcal{Q}_1, \dots, \mathcal{Q}_s, \mathcal{C})$ denote $(\Gamma_1(\mathcal{C}, \mathcal{Q}_1), \dots, \Gamma_s(\mathcal{C}, \mathcal{Q}_s))$. Consider $\sigma: \mathcal{C} \rightarrow \mathbb{F}_2$ that assigns values to each polynomial in \mathcal{C} . Let us call σ an assignment, noting that there are at most 2^ℓ such assignments. Because of f -regularity and by [Proposition 12.1](#), $\Pr[\mathcal{C} = \sigma(\mathcal{C})] \geq 2^{-\ell} - 2^\ell 2^{-3r} \varepsilon \geq 2^{-r-1}$.

By [Lemma 4.1](#) (with $\eta = 2^{-2r} \varepsilon \delta^2$), we have that for any fixed assignment σ ,

$$\|\Gamma_\sigma(\mathcal{Q}_1, \dots, \mathcal{Q}_s) - \mathcal{B}_{1/3}^{\otimes s}\|_{\text{TV}} \geq 1 - O\left(\frac{1}{\delta^{2s}}\right) - O(2^{-2r} \varepsilon). \quad (12)$$

Now we apply [Lemma 3.6](#) (with $|S| = 2^\ell$, and $\tau = 2^{-r-1}$) to $\Gamma(\mathcal{Q}_1, \dots, \mathcal{Q}_s, \mathcal{C})$, where the common part is captured by \mathcal{C} .

$$\begin{aligned} \|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} &\geq \|\mathcal{P}_I - \mathcal{B}_{1/3}^{\otimes s}\|_{\text{TV}} \\ &\geq 1 - 4 \cdot 2^\ell \cdot 2^{r+1} \cdot \left(O\left(\frac{1}{\delta^{2s}}\right) + O(2^{-2r} \varepsilon)\right) && \text{(Lemma 3.6 and Eq. (12))} \\ &\geq 1 - O\left(\frac{2^{2r}}{\delta^2 n^{\varepsilon^r}}\right) - O(\varepsilon). && (2^\ell \leq 2^r) \end{aligned}$$

For constants d and ε , the term $O\left(\frac{2^{2r}}{\delta^2 n^{\varepsilon^r}}\right)$ vanishes as n grows. Since ε can be made arbitrarily small, it follows that

$$\|\mathcal{P} - \mathcal{B}_{1/3}^{\otimes n}\|_{\text{TV}} \geq 1 - o_n(1). \quad \square$$

13 Acknowledgments

We thank Mark Kahn for providing the examples of low degree polynomials with $\Pr[P(x) = 1]$ surprisingly close to $1/3$ described in [Section 2.6](#).

References

- [Ajt83] Miklós Ajtai. Σ_1^1 -formulae on finite structures. *Annals of pure and applied logic*, 24(1):1–48, 1983.
- [BHT14] Arnab Bhattacharyya, Pooya Hatami, and Madhur Tulsiani. Algorithmic regularity for polynomials and applications. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1870–1889. SIAM, 2014.
- [BIL12] Chris Beck, Russell Impagliazzo, and Shachar Lovett. Large deviation bounds for decision trees and sampling lower bounds for ac0-circuits. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 101–110. IEEE, 2012.
- [CGZ22] Eshan Chattopadhyay, Jesse Goodman, and David Zuckerman. The space complexity of sampling. In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, pages 40–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2022.
- [DW12] Anindya De and Thomas Watson. Extractors and lower bounds for locally samplable sources. *ACM Transactions on Computation Theory (TOCT)*, 4(1):1–21, 2012.
- [ER60] Paul Erdős and Richard Rado. Intersection theorems for systems of sets. *Journal of the London Mathematical Society*, 1(1):85–90, 1960.
- [FLRS23] Yuval Filmus, Itai Leigh, Artur Riazanov, and Dmitry Sokolov. Sampling and Certifying Symmetric Functions. In Nicole Megow and Adam Smith, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023)*, volume 275 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 36:1–36:21, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [FSS84] Merrick Furst, James B Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical systems theory*, 17(1):13–27, 1984.
- [GT09] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the gowers norms. *Contributions to Discrete Mathematics*, 4(2):1–36, 2009.
- [GW20] Mika Göös and Thomas Watson. A lower bound for sampling disjoint sets. *ACM Transactions on Computation Theory (TOCT)*, 12(3):1–13, 2020.
- [Has86] John Hastad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the eighteenth annual ACM symposium on Theory of computing*, pages 6–20, 1986.
- [HHL19] Hamed Hatami, Pooya Hatami, and Shachar Lovett. Higher-order fourier analysis and applications. *Foundations and Trends® in Theoretical Computer Science*, 13(4):247–448, 2019.

- [HLS⁺25] Jordan Horacsek, Chin Ho Lee, Igor Shinkar, Emanuele Viola, and Renfei Zhou. Constant-time source decoding. *Electronic Colloquium on Computational Complexity (ECCC)*, 2025. TR25-164.
- [HS10] Elad Haramaty and Amir Shpilka. On the structure of cubic and quartic polynomials. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 331–340, 2010.
- [KL08] Tali Kaufman and Shachar Lovett. Worst case to average case reductions for polynomials. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 166–175. IEEE, 2008.
- [KLP12] Tali Kaufman, Shachar Lovett, and Ely Porat. Weight distribution and list-decoding size of reed–muller codes. *IEEE transactions on information theory*, 58(5):2689–2696, 2012.
- [KOW24] Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locality bounds for sampling hamming slices. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1279–1286, 2024.
- [KOW25a] Daniel M Kane, Anthony Ostuni, and Kewen Wu. Locally sampleable uniform symmetric distributions. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1807–1816, 2025.
- [KOW25b] Daniel M Kane, Anthony Ostuni, and Kewen Wu. Symmetric distributions from shallow circuits. *arXiv preprint arXiv:2511.14127*, 2025.
- [KT70] Tadao Kasami and Nobuki Tokura. On the weight structure of reed-muller codes. *IEEE Transactions on Information Theory*, 16(6):752–759, 1970.
- [KTA76] Tadao Kasami, Nobuki Tokura, and Saburo Azumi. On the weight enumeration of weights less than $2.5d$ of reed—muller codes. *Information and control*, 30(4):380–395, 1976.
- [LN96] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 1996.
- [LV11] Shachar Lovett and Emanuele Viola. Bounded-depth circuits cannot sample good codes. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 243–251. IEEE, 2011.
- [SS24] Ronen Shaltiel and Jad Silbak. Explicit codes for poly-size circuits and functions that are hard to sample on low entropy distributions. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 2028–2038, 2024.
- [Vio12] Emanuele Viola. The complexity of distributions. *SIAM Journal on Computing*, 41(1):191–218, 2012.

- [Vio16] Emanuele Viola. Quadratic maps are hard to sample. *ACM Transactions on Computation Theory (TOCT)*, 8(4):1–4, 2016.
- [Vio20] Emanuele Viola. Sampling lower bounds: boolean average-case and permutations. *SIAM Journal on Computing*, 49(1):119–137, 2020.
- [Vio23] Emanuele Viola. New sampling lower bounds via the separator. In *38th Computational Complexity Conference (CCC 2023)*, pages 26–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.
- [WP26] Adam Bene Watts and Natalie Parham. Unconditional quantum advantage for sampling with shallow circuits. In Shubhangi Saraf, editor, *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, volume 362 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 17:1–17:12, Dagstuhl, Germany, 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [YZ24] Huacheng Yu and Wei Zhan. Sampling, flowers and communication. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, pages 100–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024.

A One polynomial's distance from Ber(non-dyadic)

In this section we prove [Theorem A.8](#) which states that for any non-dyadic number $\rho \in (0, 1)$ and degree d polynomial P

$$|\Pr_X[P(X) = 1] - \rho| \geq \Omega_{d,\rho}(1).$$

We prove this by showing that if one has a sequence of polynomials $P^{(n)}$ converging to a distribution D , then the limit distribution D must have probability weights that are dyadic numbers. This means that there is no sequence of polynomials that samples $\text{Ber}(\rho)$ better than a constant distance as the probability weight ρ is not a dyadic number. Alas, the proof is non-constructive and does not elucidate what the constant $\Omega_{d,\rho}(1)$ should be. The proof relies on a key regularization lemma that uses similar techniques to [\[GT09, KL08\]](#) but instead is applied to a *sequence* of factors. Then the bias-rank theorem of Kaufman and Lovett ([Theorem 3.15](#)) and Vazirani's XOR lemma ([Lemma 3.2](#)) are used to conclude the result.

We will be using the standard concepts defined in [Section 3.3](#), and we will extend those definitions to sequences of factors.

Sequences. A sequence of objects will be written $\{A^{(n)}\}_{n \in \mathbb{N}}$, and we often suppress the braces and index set, writing simply $A^{(n)}$. Given two sequences $A^{(n)}$ and $B^{(n)}$, we say that $B^{(n)}$ is a *subsequence* of $A^{(n)}$ if there exists a strictly increasing map $f: \mathbb{N} \rightarrow \mathbb{N}$ such that

$$B^{(n)} = A^{(f(n))} \quad \text{for all } n \in \mathbb{N}.$$

In this case we write $B^{(n)} < A^{(n)}$.

Sequences of factors. We will frequently consider sequences of factors $\mathcal{F}^{(n)}$. Unless stated otherwise, it is understood that all factors in a given sequence have the same degree bound d and the same dimension vector (M_1, \dots, M_d) . Accordingly, we decorate by a superscript (n) all data that may vary with n (while d and (M_1, \dots, M_d) remain fixed); for instance,

$$\mathcal{F}^{(n)} = (P_1^{(n)}, \dots, P_K^{(n)}).$$

Definition A.1 (Refinement of a Sequence of Factors). We say a sequence of factors $\mathcal{G}^{(n)}$ is a refinement of $\mathcal{F}^{(n)}$ (or that $\mathcal{G}^{(n)}$ *refines* $\mathcal{F}^{(n)}$) if there is a sub-sequence $\mathcal{F}'^{(n)} < \mathcal{F}^{(n)}$ and a $\Gamma: \mathbb{F}_2^L \rightarrow \mathbb{F}_2^K$ such that $\Gamma(\mathcal{G}^{(n)}) = \mathcal{F}'^{(n)}$ for all n , where $L = \dim \mathcal{G}^{(n)}$ and $K = \dim \mathcal{F}^{(n)}$ (observe that Γ, L , and K are the same for all n). Refining is transitive: if $\mathcal{H}^{(n)}$ refines $\mathcal{G}^{(n)}$ and $\mathcal{G}^{(n)}$ refines $\mathcal{F}^{(n)}$, then $\mathcal{H}^{(n)}$ refines $\mathcal{F}^{(n)}$.

Definition A.2 (Regularity of a Sequence of Factors). Let $\mathcal{F}^{(n)}$ denote a sequence of factors. We say this sequence is r -regular if there exists some $N_0 \in \mathbb{N}$ such that for all $n \geq N_0$ the factor $\mathcal{F}^{(n)}$ is r -regular. If $\mathcal{F}^{(n)}$ is r -regular for every $r \in \mathbb{N}$ then we say this sequence is *fully regular*.

Lemma A.3 (Key Lemma). Let $\mathcal{F}^{(n)}$ be a sequence of degree d factors with dimension vector (M_1, \dots, M_d) . Then there exists a fully regular sequence of degree d factors $\mathcal{H}^{(n)}$ that is a refinement of $\mathcal{F}^{(n)}$.

Proof. The proof goes by strong induction on the dimension vector $(M_1, \dots, M_d) \in \mathbb{N}^d$. The set \mathbb{N}^d is a well-ordered set by the inverse lexicographic order⁹. As the base case of induction we have that any factor with dimension vector $(M_1, \dots, M_d) = (1, 0, \dots, 0)$ is evidently fully regular.

Let $\mathcal{F}^{(n)} = (P_1^{(n)}, \dots, P_K^{(n)})$ where $K = M_1 + \dots + M_d$. If this sequence is fully regular then we are done. Otherwise, there exists some $r \in \mathbb{N}$ such that for infinitely many n the factor $\mathcal{F}^{(n)}$ is not r -regular. Let $\mathcal{F}'^{(n)} < \mathcal{F}^{(n)}$ denote this subsequence so that every factor in $\mathcal{F}'^{(n)}$ is not r -regular.

Now, for simplicity of notation assume n as fixed, and consider some specific factor $\mathcal{F}' = (P'_1, \dots, P'_K)$ in the sequence $\mathcal{F}'^{(n)}$. Since \mathcal{F}' is not r -regular, there exist $\lambda_1, \dots, \lambda_K \in \mathbb{F}_2$ not all zeros, $\ell \in [d]$, $j \in [K]$, $\Gamma: \mathbb{F}_2^r \rightarrow \mathbb{F}_2$ and polynomials Q_1, \dots, Q_r of degree at most $\ell - 1$ such that

$$\lambda_1 P'_1 + \dots + \lambda_K P'_K = \Gamma(Q_1, \dots, Q_r),$$

where $\ell = \max_i \deg(\lambda_i P'_i)$ and j is such that $\deg(\lambda_j P'_j) = \ell$.

Let us refer to the tuple $(\lambda_1, \dots, \lambda_K, \ell, j, \Gamma, \deg(Q_1), \dots, \deg(Q_r))$ as a *certificate of non-regularity*. Since the number of possible certificates is bounded by a function of r, d , and K , it follows that infinitely many factors in the sequence $\mathcal{F}'^{(n)}$ share the same certificate. Let $\mathcal{F}''^{(n)} < \mathcal{F}'^{(n)}$ be a sub-sequence for which the certificate of non-regularity is the same.

Now by removing the j^{th} polynomial in $\mathcal{F}''^{(n)}$ and adding $Q_i^{(n)}$'s we then refine

$$\mathcal{F}''^{(n)} = (P''_1, \dots, P''_K) \longmapsto \mathcal{G}^{(n)} = (P''_1, \dots, P''_{j-1}, P''_{j+1}, \dots, P''_K, Q_1, \dots, Q_r).$$

(If $\ell = 1$, we simply remove P_j and add no Q_i 's because Q_i 's would be constant in that case). Since the certificate was chosen in such a way that it is the same for all factors in $\mathcal{F}''^{(n)}$, there exists a function $\Psi: \mathbb{F}_2^{K+r-1} \rightarrow \mathbb{F}_2^K$ such that $\Psi(\mathcal{G}^{(n)}) = \mathcal{F}''^{(n)}$ for all n . Note that the dimension vector of $\mathcal{G}^{(n)}$ is at most $(M_1, \dots, M_{\ell-1} + r, M_\ell - 1, M_{\ell+1}, \dots, M_d)$ which is smaller than (M_1, \dots, M_d) in our ordering of \mathbb{N}^d . By the induction hypothesis there exists some fully regular $\mathcal{H}^{(n)}$ that refines $\mathcal{G}^{(n)}$. Finally, by transitivity of refinement, $\mathcal{H}^{(n)}$ also refines $\mathcal{F}^{(n)}$. \square

Lemma A.4. *Let $\mathcal{F}^{(n)} = (P_1^{(n)}, \dots, P_K^{(n)})$ be a fully regular sequence of factors of degree d . Then the joint distribution of polynomials $(P_1^{(n)}, \dots, P_K^{(n)})$ converges to the uniform distribution U_K .*

Proof. We show that for every positive ε there exists some N_0 such that for all $n \geq N_0$ we have $\|\mathcal{F}^{(n)} - U_K\|_{\text{TV}} \leq \varepsilon$, which implies our claim.

Let $\varepsilon > 0$ be fixed. Then define $r = c_{\text{KL}}(d, \varepsilon 2^{-K/2})$, where c_{KL} comes from [Theorem 3.15](#). Since $\mathcal{F}^{(n)}$ is fully regular there exists some N_0 such that $\mathcal{F}^{(n)}$ is r -regular for all $n \geq N_0$. Then by [Theorem 3.15](#)

$$\text{bias}(\lambda_1 P_1^{(n)} + \dots + \lambda_K P_K^{(n)}) \leq \varepsilon 2^{-K/2}, \quad \text{for all } n \geq N_0 \text{ and } \lambda_i \text{ not all zeros.}$$

Thus, by Vazirani's XOR lemma, we have that $\|\mathcal{F}^{(n)} - U_K\|_{\text{TV}} \leq \varepsilon$ for all $n \geq N_0$. \square

⁹We have $(M'_1, \dots, M'_d) < (M_1, \dots, M_d)$ if there exists some $i \in [d]$ such that $M'_j = M_j$ for all $j > i$ and $M'_i < M_i$.

Definition A.5 (Computable distribution). We say a distribution D over $\{0, 1\}^s$ is *computable* if there exists some $K \in \mathbb{N}$ and function $\Gamma: \{0, 1\}^K \rightarrow \{0, 1\}^s$ such that $D = \Gamma(U_K)$.

It is an easy observation that a distribution D is computable if and only if the probability of every outcome is a dyadic number.

Lemma A.6. *Let $P^{(n)}$ be a sequence of polynomials of degree d that converges to the distribution D in TV-distance. Then, there exists a positive integer K and a function $\Gamma: \mathbb{F}_2^K \rightarrow \mathbb{F}_2^s$ such that $D = \Gamma(U_K)$. In other words, the limit distribution D is computable.*

Proof. Let us describe $P^{(n)}$ by the trivial sequence of factors that computes it; that is to say $\mathcal{F}^{(n)} = (P^{(n)})$ with the dimension vector $(0, \dots, 0, 1) \in \mathbb{N}^d$.

By [Lemma A.3](#) there exists K and a fully regular degree d sequence $\mathcal{H}^{(n)} = (Q_1^{(n)}, \dots, Q_K^{(n)})$ of dimension K that is a refinement of $\mathcal{F}^{(n)}$. That is, for a subsequence $P'^{(n)}$ of $P^{(n)}$ there exists some Γ such that for all n , $P'^{(n)} = \Gamma(Q_1^{(n)}, \dots, Q_K^{(n)})$.

Since $\mathcal{H}^{(n)}$ is fully regular, by [Lemma A.4](#) the joint distribution $(Q_1^{(n)}, \dots, Q_K^{(n)})$ converges to the uniform distribution, and therefore, $P'^{(n)} = \Gamma(Q_1^{(n)}, \dots, Q_K^{(n)})$ converges to $\Gamma(U_K)$. \square

Remark A.7. In fact, [Lemma A.6](#) extends with essentially no additional effort to the setting of joint distributions of low-degree polynomials. We have presented it in the present form purely for the sake of slightly cleaner notation.

Theorem A.8. *Let $0 < \rho < 1$ be a non-dyadic number, and let d be a positive integer. Then for any degree d polynomial $P: \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ it holds that $\|P - \text{Ber}(\rho)\|_{\text{TV}} \geq \Omega_{d,\rho}(1)$.*

Proof. Toward contradiction, suppose this were not the case. That is, suppose there exists a sequence $P^{(n)}$ of polynomials of degree d such that

$$\|P^{(n)} - \text{Ber}(\rho)\|_{\text{TV}} \leq o_{d,\rho}(1) \quad \text{for } n \rightarrow \infty.$$

Then this sequence converges to $\text{Ber}(\rho)$ in total variation distance. On the other hand, by [Lemma A.6](#) there exist K and Γ such that the limit of $P^{(n)}$ is the distribution $\Gamma(U_K)$. This is a contradiction as ρ is non-dyadic. \square

B Deferred proofs

Proof of Vazirani's XOR lemma ([Lemma 3.2](#)). Let $\mu, u: \mathbb{F}_2^n \rightarrow [0, 1]$ be the probability mass functions of \mathcal{X} and $\mathcal{U}^{\otimes n}$ respectively. We will bound $\|\mathcal{X} - \mathcal{U}^{\otimes n}\|_2$ from above. We have

$$\begin{aligned} \|\mathcal{X} - \mathcal{U}^{\otimes n}\|_2^2 &= \sum_x (\mu(x) - u(x))^2 \\ &= 2^n \mathbf{E}_x [(\mu(x) - u(x))^2] \end{aligned}$$

$$\begin{aligned}
&= 2^n \mathbf{E} [\mu(x)^2] + 2^n \mathbf{E} [u(x)^2] - 2^{n+1} \mathbf{E} [\mu(x)u(x)] \\
&= 2^n \sum_{\gamma} \widehat{\mu}(\gamma)^2 + 2^{-n} - 2\widehat{\mu}(0) \\
&= 2^n \sum_{\gamma \neq 0} \widehat{\mu}(\gamma)^2. \qquad \text{(because } \widehat{\mu}(0) = 2^{-n} \text{)}
\end{aligned}$$

Consider some non-trivial character $\chi_{\gamma} = (-1)^L$ where L is some nonzero linear form. Then, $|\widehat{\mu}(\gamma)| = |\mathbf{E}_x[\chi_{\gamma}(x)\mu(x)]| = |2^{-n} \text{bias}(L(X))| \leq \varepsilon \cdot 2^{-n}$. Therefore,

$$\|\mathcal{X} - \mathbf{u}^{\otimes n}\|_2^2 \leq 2^{2n-1}(2^{-2n} \varepsilon^2) \leq \varepsilon^2.$$

Then $\|\mathcal{X} - \mathbf{u}^{\otimes n}\|_1 \leq 2^{n/2} \varepsilon$ is immediate by Cauchy-Schwarz. \square

Proof of Proposition 3.13. We prove this claim by induction on $\vec{M} := (M_1, \dots, M_d)$, where the tuples in \mathbb{N}^d are ordered by inverse lexicographic ordering. Our base case covers all cases where $N_2 = \dots = N_d = 0$. We have,

$$\psi(N_1, 0, \dots, 0) = N_1 \leq \sum_i M_i \leq \psi(M_1, \dots, M_d).$$

From now on we assume that $N_i > 0$ for some $i \geq 2$. If $\vec{N} = \vec{M}$, then the statement clearly holds. Thus we further assume that $\vec{N} < \vec{M}$.

Let j be the right-most index such that $M_j > N_j$ (i.e. j is the index that shows \vec{M} is greater than \vec{N} in inverse lexicographic order). In particular, the entry M_j is nonzero.

Define $\ell \geq 2$ (resp. $k \geq 2$) to be the first index of \vec{M} (resp. \vec{N}) that is nonzero. Note that $\ell \leq j$.

Now by definition of ψ , we have $\psi(\vec{M}) = \psi(\vec{M}')$ and $\psi(\vec{N}) = \psi(\vec{N}')$ where,

$$\begin{aligned}
\vec{M}' &= (M_1, \dots, M_{\ell-1} + f(\sum_i M_i), M_{\ell} - 1, \dots, M_d) \\
\vec{N}' &= (N_1, \dots, N_{k-1} + f(\sum_i N_i), N_k - 1, \dots, N_d)
\end{aligned}$$

Clearly, $\sum_i \vec{N}'_i \leq \sum_i \vec{M}'_i$ still holds because of monotonicity of f . To show that $\vec{N}' < \vec{M}'$, we consider three cases.

- ($k \geq j$) If $k \geq j$ then $N'_k < M'_k$ and therefore $\vec{N}' < \vec{M}'$.
- ($k < j$ and $\ell < j$) Then $N'_j = N_j < M_j = M'_j$ and $\vec{N}' < \vec{M}'$.
- ($k < j$ and $\ell = j$) Then $M'_j \geq N'_j$, and if this inequality is strict, then we are done. Otherwise, $M'_{j-1} \geq f(\sum_i M_i) \geq \sum_i M_i > N'_{j-1}$.

The three cases show that $\vec{M}' > \vec{N}'$ and the claim follows by the induction hypothesis. \square