

Entropy of weight distributions of small-bias spaces and pseudobinomiality ^{*†}

Louay Bazzi [‡]

July 13, 2015

Abstract

A classical bound in information theory asserts that small L_1 -distance between probability distributions implies small difference in Shannon entropy, but the converse need not be true. We show that if a probability distribution on $\{0, 1\}^n$ has small-bias, then the converse holds for its weight distribution in the proximity of the binomial distribution. Namely, we argue that if a probability distribution μ on $\{0, 1\}^n$ is δ -biased, then $\|\bar{\mu} - \text{bin}_n\|_1^2 \leq (2 \ln 2)(n\delta + H(\text{bin}_n) - H(\bar{\mu}))$, where $\bar{\mu}$ is the weight distribution of μ and bin_n is the binomial distribution on $\{0, \dots, n\}$. The key result behind this bound is a lemma which asserts the non-positivity of all the Fourier coefficients of the log-binomial function $L : \{0, 1\}^n \rightarrow \mathbb{R}$ given by $L(x) = \lg \text{bin}_n(|x|)$. The original question which motivated the work reported in this paper is the problem of explicitly constructing a small subset of $\{0, 1\}^n$ which is ϵ -pseudobinomial in the sense that the weight distribution of each of its restrictions and translations is ϵ -close to the binomial distribution. We study the notion of pseudobinomiality and we conclude that, for spaces with $n^{-\Theta(1)}$ -small bias, the pseudobinomiality error in the L_1 -sense is equivalent to that in the entropy-difference-sense, in the $n^{-\Theta(1)}$ -error regime. We also study the notion of average case pseudobinomiality, and we show that for spaces with $n^{-\Theta(1)}$ -small bias, the average entropy of the weight distribution of a random translation of the space is $n^{-\Theta(1)}$ -close to the entropy of the binomial distribution. We discuss resulting questions on the pseudobinomiality of sums of independent small-bias spaces. Using the above results, we show that the following conjectures are equivalent: (1) For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, the \mathbb{F}_2 -sum $X + Y$ is $O((n\delta)^{\Theta(1)})$ -pseudobinomial; (2) For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, the entropy of the weight of the sum $H(|X + Y|) \geq \min\{H(|X|), H(|Y|)\} - O((n\delta)^{\Theta(1)})$.

^{*}To appear in Chicago Journal of Theoretical Computer Science, 2015

[†]A conference version of this paper appeared in COCOON 2015, LNCS 9198 proceedings, pages 495-506.

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

1 Introduction

The ultimate goal of pseudorandomness is to construct low complexity PRGs which look random to all small circuits. Without hardness assumptions [24, 14], asymptotically optimal seed lengths are not known even for simple models such as depth-2 circuits and log-space computations. Among the simplest desirable properties of a PRG are the almost k -wise independence property and the stronger small-bias property [22]. Small-bias probability distributions have various applications in pseudorandomness (e.g., [27, 20, 10] and the references therein). A probability distribution on $\{0, 1\}^n$ has small bias if it looks like the uniform distribution to all parity functions on subsets of the n input variables. More formally, let $0 \leq \delta \leq 1$ and consider the characters $\{\chi_z\}_{z \in \{0, 1\}^n}$ of the abelian group structure \mathbb{Z}_2^n on $\{0, 1\}^n$: $\chi_z(x) \stackrel{\text{def}}{=} (-1)^{\sum_i x_i z_i}$. A probability distribution μ on $\{0, 1\}^n$ is δ -biased if $|E_\mu \chi_z| \leq \delta$ for each nonzero $z \in \{0, 1\}^n$ [22]. Probability distributions with the δ -bias property and support size $\left(\frac{n}{\delta}\right)^{\Theta(1)}$ can be explicitly constructed from linear codes [22, 1].

The original question which motivated the work reported in this paper is the problem of explicitly constructing small subsets $S \subset \{0, 1\}^n$ such that for each nonempty subset of indices $I \subset \{1, \dots, n\}$ and each translation vector $u \in \{0, 1\}^I$, the weight distribution of the translation (over \mathbb{F}_2) by u of the restriction of S to $\{0, 1\}^I$ looks like the binomial distribution on $\{0, \dots, |I|\}$. We call S ϵ -pseudobinomial if the distance from the binomial distribution is at most ϵ in the L_1 -sense, for each I and u .

The ϵ -pseudobinomiality property is a natural extension of the ϵ -bias property. The more general problem of constructing PRGs for combinatorial shapes was studied by Gopalan, Meka, Reingold and Zuckerman [13]. Without hardness assumptions, it is an open problem to construct ϵ -pseudobinomial spaces of size polynomial in n and $\frac{1}{\epsilon}$, i.e., of seed length $O(\log \frac{n}{\epsilon})$.

Nisan generator for log-space computations [23] gives n^{-c} -pseudobinomial spaces of seed length $O(\log^2 n)$, for each constant c . The work of Gopalan *et al.* [13] leads to an ϵ -pseudobinomial space of seed length $O(\log n + \log^2 \frac{1}{\epsilon})$, which matches the seed length of Nisan generator in the inverse polynomial error regime and improves on it in the subpolynomial regime. Independently and concurrently with the present paper, De [9] constructed a PRG for combinatorial sums, which results in an ϵ -pseudobinomial space of seed length $O(\log^{3/2} \frac{n}{\epsilon})$. Independently and concurrently also, Gopalan, Kane and Meka [12] constructed an ϵ -pseudobinomial space of seed length $O(\log \frac{n}{\epsilon} \log^{O(1)} \log \frac{n}{\epsilon})$.

A related problem was studied by Lovett, Reingold, Trevisan and Vadhan [20] and independently by Meka and Zuckerman [21] who constructed an $O(\log n)$ seed-length PRG which fools mod- M gates, where $M = O(1)$ is a power of a prime. Another related work is by Rabani and Shpilka [25], who constructed explicit polynomial complexity ϵ -nets for threshold functions with arbitrary coefficients.

1.1 Contribution

Motivated by the above question, we present in this paper a systematic study of pseudobinomial spaces. While we fall short of constructing pseudobinomial spaces of polynomial complexity, we establish some of their basic properties that contribute to a better understanding of such spaces.

To get started, we ignore translations and restrictions in the definition of pseudobinomiality, and in general we study the relation between different measures of similarity between the binomial distribution bin_n on $\{0, \dots, n\}$ and the weight distribution $\bar{\mu}$ of a probability distribution μ on $\{0, 1\}^n$, assuming that μ has small bias. The similarity between two probability distributions on

$\{0, \dots, n\}$ is captured by various measures, some of which are the relative entropy $D(\cdot||\cdot)$, the L_1 -distance, and the much weaker notions of distance in entropy. Pinsker's bound (Lemma 1.1.a) asserts that relative entropy is in general at least as strong distance L_1 -distance. Another classical bound in information theory asserts that small L_1 -distance between probability distributions implies small difference in Shannon entropy (Lemma 1.1.b), but the converse need not be true. We show that if μ has small bias, then the converse holds for its weight distribution in the proximity of the binomial distribution, hence it is enough to guarantee that the entropy $H(\bar{\mu})$ of its weight distribution $\bar{\mu}$ is close the entropy $H(\text{bin}_n)$ of the binomial in order to conclude that $\bar{\mu}$ is close to bin_n in the L_1 -sense. In particular, we show that

$$D(\bar{\mu}||\text{bin}_n) \leq n\delta + H(\text{bin}_n) - H(\bar{\mu}), \quad (1)$$

where δ is the bias of μ . Hence, it follows from Pinsker's bound that

$$\|\bar{\mu} - \text{bin}_n\|_1^2 \leq (2 \ln 2)(n\delta + H(\text{bin}_n) - H(\bar{\mu})). \quad (2)$$

The key result behind (1) is a lemma which asserts the non-positivity of all the Fourier coefficients of the log-binomial function $L : \{0, 1\}^n \rightarrow \mathbb{R}$ given by

$$L(x) = \lg \text{bin}_n(|x|).$$

A by-product of (1) is that, unlike arbitrary probability distribution on $\{0, 1\}^n$, $H(\bar{\mu})$ cannot be significantly larger than $H(\text{bin}_n)$ if μ has small bias. Note that $H(\text{bin}_n)$ is approximately half $\lg(n+1)$ – the maximum entropy of a probability distribution on $\{0, \dots, n\}$ [15] (Lemma 1.2).

Then we study the notion of pseudobinomiality and we conclude from (2) that, for spaces with $n^{-\Theta(1)}$ -small bias, the pseudobinomiality error in the L_1 -sense is equivalent to that in the entropy-difference-sense, in the $n^{-\Theta(1)}$ -error regime.

The notion of pseudobinomiality in entropy suggests studying the minimum weight entropy $H_{min}(\mu)$ of a probability distribution μ on $\{0, 1\}^n$, which we define as the minimum Shannon entropy of the weight distribution of a translation of μ . To compare $H_{min}(\mu)$ to $H(\text{bin}_n)$, we study the related notion of average weight entropy $H_{avg}(\mu)$, which we define as the average entropy of the weight distribution of a random translation of μ . Thus, $H_{min}(\mu) \leq H_{avg}(\mu)$. We show that $H_{avg}(\mu) \leq H(\text{bin}_n)$, where the inequality is strict unless μ is the uniform distribution. Then we study the notion of average case pseudobinomiality, and we show that for spaces with $n^{-\Theta(1)}$ -small bias, $H_{avg}(\mu)$ is $n^{-\Theta(1)}$ -close to $H(\text{bin}_n)$.

Finally, we discuss resulting questions on the pseudobinomiality of sums of independent small-bias spaces. Reingold and Vadhan asked whether the sum of two independent $n^{-O(1)}$ -biased spaces fools log-space [21]. Since any distribution which fools log-space must be pseudobinomial, a natural question is whether the sum of any two independent δ -biased spaces is $O((\delta n)^{\Theta(1)})$ -pseudobinomial. Using the above results, we show that this question is equivalent to the weaker question of whether for all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, the entropy of the weight of the sum

$$H(|X + Y|) \geq \min\{H(|X|), H(|Y|)\} - O((n\delta)^{\Theta(1)}).$$

In the remainder of this introductory section, we summarize in Section 1.2 our results in the context of the related literature and we outline the underlying proof techniques. Then we give in Section 1.3 Fourier analysis and information theory preliminaries used in this paper.

1.2 Paper overview and techniques

1.2.1 Entropy of weight distributions and small-bias

We start by studying the following basic question. Consider a probability distribution μ on $\{0, 1\}^n$ with small bias. What is the relation between different measures of similarity between the weight distribution of μ and the binomial distribution on $\{0, \dots, n\}$?

We need some preliminary notations and definitions. In what follows $n \geq 1$ is an integer. If $x \in \{0, 1\}^n$, the *weight* of x , which we denote by $|x|$, is the number of nonzero coordinates of x . If μ is a probability distributions on $\{0, 1\}^n$, we denote the *weight distribution* of μ by $\bar{\mu}$. That is, $\bar{\mu}$ is the probability distribution on $[0 : n] \stackrel{\text{def}}{=} \{0, 1, \dots, n\}$ given by $\bar{\mu}(w) = \mu(x \in \{0, 1\}^n : |x| = w)$. The *uniform distribution* on $\{0, 1\}^n$ is denoted by U_n and the *binomial distribution* on $[0 : n]$ is denoted by bin_n . Thus, $\text{bin}_n(w) = \frac{\binom{n}{w}}{2^n}$ for all $w \in [0 : n]$, and $\text{bin}_n = \overline{U_n}$.

If γ_1 and γ_2 are two probability distributions on a finite set \mathcal{X} (we are interested in $\mathcal{X} = [0 : n]$), the similarity between γ_1 and γ_2 is captured by various measures, some of which are the relative entropy, the L_1 distance, and the much weaker notions of distance in entropy. The L_1 distance is also called *total variation* since

$$\|\gamma_1 - \gamma_2\|_1 \stackrel{\text{def}}{=} \sum_{w \in \mathcal{X}} |\gamma_1(w) - \gamma_2(w)| = 2 \max_{A \subset \mathcal{X}} |\gamma_1(A) - \gamma_2(A)|. \quad (3)$$

The *relative entropy* of γ_1 with respect to γ_2 is given by

$$D(\gamma_1 || \gamma_2) \stackrel{\text{def}}{=} \sum_w \gamma_1(w) \lg \frac{\gamma_1(w)}{\gamma_2(w)}, \quad (4)$$

where $\lg = \log_2$ is the base-2 logarithm. The relative entropy is not symmetric but it satisfies the nonnegativity property $D(\gamma_1 || \gamma_2) \geq 0$ with equality iff $\gamma_1 = \gamma_2$. If γ is a probability distributions on \mathcal{X} , the *Shannon entropy* of γ is defined as

$$H(\gamma) \stackrel{\text{def}}{=} - \sum_w \gamma(w) \lg \gamma(w). \quad (5)$$

We have the following classical bounds.

Lemma 1.1 ([8]). *If γ_1 and γ_2 are two probability distributions on a finite set \mathcal{X} , then*

(a) **(Pinsker's bound)** $\|\gamma_1 - \gamma_2\|_1^2 \leq (2 \ln 2) D(\gamma_1 || \gamma_2)$

(b) **(Entropy-difference bound)** *If $\epsilon = \|\gamma_1 - \gamma_2\|_1 \leq \frac{1}{2}$, then $|H(\gamma_1) - H(\gamma_2)| \leq \epsilon \lg \frac{|\mathcal{X}|}{\epsilon}$.*

The entropy-difference bound asserts that small L_1 -distance between probability distributions implies small difference in Shannon entropy, but the converse need not be true.

In our context $\mathcal{X} = [0 : n]$. Thus, for $\epsilon = n^{-c}$, where $c > 0$ is constant, we have $\epsilon \lg \frac{|\mathcal{X}|}{\epsilon} = O(n^{-c} \lg n)$. We are interested in probability distributions on $[0 : n]$ which are weight distributions $\bar{\mu}$ of probability distributions μ on $\{0, 1\}^n$. We would like to study the conditions under which $\bar{\mu}$ is close to the binomial distribution bin_n .

The maximum entropy of a probability distribution on $[0 : n]$ is $\lg(n+1)$ and it is achieved by the uniform distribution $U_{[0:n]}$ on $[0 : n]$. The entropy $H(\text{bin}_n)$ of the binomial distribution is approximately half the maximum entropy:

Lemma 1.2 (Entropy of the binomial [15]). *The entropy of the binomial distribution is given by $H(\text{bin}_n) = \frac{1}{2} \lg \frac{\pi en}{2} + O(\frac{1}{n})$. Thus, $\frac{H(\text{bin}_n)}{H(U_{[0:n]})} = \frac{1}{2} + \Theta\left(\frac{1}{\lg n}\right)$.*

It is not hard to see that there are probability distributions μ on $\{0, 1\}^n$ such that $|H(\text{bin}_n) - H(\bar{\mu})|$ is as small as zero but $\|\text{bin}_n - \bar{\mu}\|_1 = \Theta(1)$.

In Section 2, we show that if a probability distribution μ on $\{0, 1\}^n$ has small bias, then the converse of the entropy-difference bound holds for its weight distribution $\bar{\mu}$ in the proximity of the binomial distribution. In particular, if μ has small bias, then it is enough to guarantee that $H(\text{bin}_n) - H(\bar{\mu})$ is small to conclude that $\|\bar{\mu} - \text{bin}_n\|_1$ is small.

Theorem (Entropy-difference converse bound). Let μ be a probability distributions on $\{0, 1\}^n$. If μ is δ -biased, then

$$D(\bar{\mu} || \text{bin}_n) \leq n\delta + H(\text{bin}_n) - H(\bar{\mu}).$$

Hence (by Pinsker's bound),

$$\|\bar{\mu} - \text{bin}_n\|_1^2 \leq (2 \ln 2)(n\delta + H(\text{bin}_n) - H(\bar{\mu})).$$

Accordingly, it follows from the nonnegativity of relative entropy that if μ is δ -biased, then

$$H(\bar{\mu}) \leq H(\text{bin}_n) + n\delta. \tag{6}$$

That is, unlike arbitrary probability distribution on $\{0, 1\}^n$, $H(\bar{\mu})$ cannot be significantly higher than the entropy of the binomial if it has small bias. For an arbitrary μ , the entropy of its weight distribution can be as large as $H(U_{[0:n]}) = \lg(n+1) \approx 2H(\text{bin}_n)$ (e.g., let μ be uniformly supported by a subset $S \subset \{0, 1\}^n$ such that for each $w \in [0 : n]$, S contains exactly one string of weight w).

The key behind the entropy-difference converse bound is the following lemma.

Lemma (Negative spectrum lemma). Let $L : \{0, 1\}^n \rightarrow \mathbb{R}$ be log-binomial function given by:

$$L(x) = \lg \text{bin}_n(|x|).$$

Then the Fourier transform \widehat{L} of L is non-positive: $\widehat{L}(z) \leq 0$ for each $z \in \{0, 1\}^n$. Moreover, $\widehat{L}(z) = 0$ if $|z|$ odd, and $\widehat{L}(z) < 0$ if $z \neq 0$ and $|z|$ even.

The origin of the log-binomial function in our context is the equation

$$D(\bar{\mu} || \text{bin}_n) = E_{U_n} L - E_{\mu} L + H(\text{bin}_n) - H(\bar{\mu}),$$

which holds for any probability distribution μ on $\{0, 1\}^n$ and follows from the definitions of D , H , and L . The negative spectrum lemma implies that $\|\widehat{L}\|_1$ is small, namely $\|\widehat{L}\|_1 = -L(0) = n$, which implies the entropy-difference converse bound since $|E_{U_n} L - E_{\mu} L| \leq \|\widehat{L}\|_1 \delta$.

The proof of the negative spectrum lemma is in Section 9 and it is analytical in nature. It boils down to showing that if $m \geq 2$ is even and $a, b \geq 0$ are integers, then

$$\beta_m(a, b) \stackrel{\text{def}}{=} \sum_{w=0}^m (-1)^w \binom{m}{w} \lg \binom{m+a+b}{w+a} < 0.$$

At a high level, we do this by first showing that $\lim_{c \rightarrow \infty} \beta_m(c, c) = 0$ using de Moivre-Laplace normal approximation of the binomial. Then we argue that β_m is strictly increasing in the sense that $\beta_m(a, b) < \beta_m(a + a', b + b')$ for each $a, b, a', b' \geq 0$ such that not both a' and b' are zero. We derive the latter from the inequality $\sum_w (-1)^w \binom{m}{w} \lg(w + i) < 0$, for all $i \geq 0$ and even $m \geq 2$, which we establish by examining the Taylor series of the logarithm and the moment generating function $g_s : \{0, 1\}^m \rightarrow \mathbb{R}$ given by $g_s(x) = (-1)^{|x|} e^{s(|x| - m/2)}$.

In Section 2.1 we give two specific applications of the negative spectrum lemma:

- If the Fourier transform of a probability distribution μ is nonnegative, then $H(\text{bin}_n) - H(\bar{\mu}) \leq D(\bar{\mu} || \text{bin}_n)$. This holds for instance if μ is uniformly supported by a linear code or if it is the convolution of a probability distribution on $\{0, 1\}^n$ with itself.
- The entropy of the weight distribution of the even weight strings in $\{0, 1\}^n$ is strictly larger than that of the odd weight strings if n is even.

1.2.2 Pseudobinominality

We study in Section 3 the notion of pseudobinominality in the L_1 -sense and the entropy-sense. We conclude from the entropy-difference bound and the entropy-difference converse bound that for spaces with $n^{-\Theta(1)}$ -small bias, the two notions are equivalent in the $n^{-\Theta(1)}$ -error regime.

Definition (Pseudobinominality in the L_1 -sense). *A probability distribution μ on $\{0, 1\}^n$ is called ϵ -pseudobinomial in the L_1 -sense if the weight distribution of each translation of a restriction of μ is ϵ -close to the binomial distribution in the L_1 -sense.*

That is, for each nonempty set of indices $I \subset [n]$ and each translation vector $u \in \{0, 1\}^I$, we have $\|\overline{\sigma_u \mu^I} - \text{bin}_{|I|}\|_1 \leq \epsilon$, where μ^I is the restriction of μ on $\{0, 1\}^I$ (i.e., $\mu^I(y) = \{x : x_I = y\}$), $\sigma_u \mu^I$ is the translation over \mathbb{F}_2 of μ^I by u (i.e., $(\sigma_u \mu^I)(y) = \mu^I(y + u)$), $\overline{\sigma_u \mu^I}$ is (as defined above) the weight distribution of $\sigma_u \mu^I$, and $\text{bin}_{|I|}$ is the binomial distribution on $[0 : |I|]$.

It is not hard to see that ϵ -pseudobinominality in the L_1 -sense implies ϵ -bias. It is a natural extension of the ϵ -bias property, which is also invariant under translations and preserved by restrictions.

Similarly, we can define pseudobinominality in the L_∞ -sense and the L_2 -sense. They are both equivalent to pseudobinominality in the L_1 sense in the $n^{-\Theta(1)}$ -error regime. In more classical pseudorandomness terms, ϵ -pseudobinominality in the L_∞ -sense is equivalent to ϵ -fooling¹ all functions $f_{a,b} : \{0, 1\}^n \rightarrow \{0, 1\}$ given by $f_{a,b}(x) = 1$ iff $\sum_i a_i x_i = b$, where $a \in \{0, \pm 1\}^n$ and $b \in \mathbb{Z}$. Thus, constructing ϵ -pseudobinomial spaces is a special case of the problem of constructing PRGs for combinatorial shapes considered by Gopalan, Meka, Reingold and Zuckerman [13].

A related work is that of Lovett *et al.* [20] and Meka and Zuckerman [21] who based on the work of Viola [7, 19, 27], gave an $O(\log n)$ seed-length PRG which fools mod- M gates, where $M = O(1)$ is a power of a prime. The framework of Mod- M gates allows the coefficients $\{a_i\}_i$ to take arbitrary values mod M , but the restriction $M = O(1)$ makes pseudobinominality more difficult to achieve. Another related work is by Rabani and Shpilka [25], who constructed explicit polynomial complexity ϵ -nets for threshold functions. A threshold function $t_{a,b} : \{0, 1\}^n \rightarrow \{0, 1\}$ given by $t_{a,b}(x) = 1$ iff $\sum_i a_i x_i \leq b$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The result of [25] is an explicit construction of a subset $S \subset \{0, 1\}^n$ of size polynomial in n and $\frac{1}{\epsilon}$ such that for all $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, if $E_{U_n} t_{a,b} > \epsilon$,

¹A probability distribution μ on $\{0, 1\}^n$ ϵ -fools a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ if $|E_\mu f - E_{U_n} f| \leq \epsilon$.

then $E_{St_{a,b}} \neq 0$. In the context of threshold functions, $n^{-\Theta(1)}$ -pseudobinomiality is equivalent to $n^{-\Theta(1)}$ -fooling binary threshold functions $t_{a,b}$ corresponding to the case when $a \in \{0, \pm 1\}^n$ and $b \in \mathbb{Z}^2$.

We define below pseudobinomiality in the entropy-sense.

Definition (Minimum weight entropy). If μ is a probability distribution on $\{0, 1\}^n$, define the min-weight entropy of μ :

$$H_{min}(\mu) = \min_{u \in \{0, 1\}^n} H(\overline{\sigma_u \mu}),$$

i.e., $H_{min}(\mu)$ is the minimum Shannon entropy of the weight distribution of a translation of μ .

Definition (Pseudobinomiality in the entropy-sense). A probability distribution μ on $\{0, 1\}^n$ is called ϵ -pseudobinomial in the entropy-sense if for each nonempty index subset $I \subset [n]$, we have $H_{min}(\mu^I) \geq H(\text{bin}_{|I|}) - \epsilon$.

It follows from the entropy-difference bound and the entropy-difference converse bound that for spaces with $n^{-\Theta(1)}$ -small bias, pseudobinomiality in the L_1 -sense is equivalent to pseudobinomiality in the entropy-sense in the $n^{-\Theta(1)}$ -error regime. Namely, let μ be a δ -biased probability distribution on $\{0, 1\}^n$ and $\epsilon > 0$. Then:

- a) If $\epsilon \leq 1/2$ and μ is ϵ -pseudobinomial in the L_1 -sense, then it is $\epsilon \lg \frac{n+1}{\epsilon}$ -pseudobinomial in the entropy-sense.
- b) If μ is ϵ -pseudobinomial in entropy-sense then it is $\sqrt{(2 \ln 2)(n\delta + \epsilon)}$ -pseudobinomial in the L_1 -sense.

Min-weight entropy and average-weight entropy. In Section 4, we elaborate on the notion of min-weight entropy and we study the related notion of average-weight entropy.

Definition (Average-weight entropy). If μ be a probability distribution on $\{0, 1\}^n$, define the average-weight entropy of μ :

$$H_{avg}(\mu) \stackrel{\text{def}}{=} E_{u \sim U_n} H(\overline{\sigma_u \mu}) = H(|X + U| \mid U),$$

where $X \sim \mu$ and $U \sim U_n$ are independent.

There are distributions μ on $\{0, 1\}^n$ such that the weight distribution $\bar{\mu}$ of μ is the uniform distribution on $[0 : n]$, and hence $H(\bar{\mu}) = \lg(n+1) \approx 2H(\text{bin}_n)$. We note that each probability μ on $\{0, 1\}^n$ has a translation whose weight distribution has entropy at most $H(\text{bin}_n)$; we show that

$$H_{min}(\mu) \leq H_{avg}(\mu) \leq H(\text{bin}_n),$$

where the inequalities $H_{avg}(\mu) \leq H(\text{bin}_n)$ and $H_{min}(\mu) \leq H(\text{bin}_n)$ are strict unless μ is the uniform distribution U_n on $\{0, 1\}^n$. That is, U_n is the unique maximum min-weight entropy distribution and the unique maximum average-weight entropy distribution.

² If μ is ϵ -pseudobinomial in the L_1 -sense, then it $\epsilon/2$ -fools all binary threshold functions $\{t_{a,b}\}_{a,b}$. Conversely, if μ ϵ -fools all binary threshold functions $\{t_{a,b}\}_{a,b}$, then it is 2ϵ -pseudobinomial in the L_∞ -sense, and hence $2(n+1)\epsilon$ -pseudobinomial in the L_1 -sense.

Local pseudobinomiality. We note in Section 5 that small-bias does not imply pseudobinomiality in the L_1 -sense or the entropy-sense even if the bias is exponentially small, but it is enough to guarantee local pseudobinomiality on small subsets of indices.

Average case pseudobinomiality. In Section 6, we study average-case pseudobinomiality. We show that for spaces with small bias, the weight distribution of a random translation of the space is close to the binomial distribution. First we show that if μ is a δ -biased probability distribution,

$$E_{u \sim U_n} \|\overline{\sigma_u \mu} - \text{bin}_n\|_2^2 \leq \delta^2 \quad (7)$$

$$E_{u \sim U_n} \|\overline{\sigma_u \mu} - \text{bin}_n\|_1 \leq \delta \sqrt{n+1}. \quad (8)$$

Then we conclude a similar bound for the average-weight entropy:

$$0 \leq I(|X+U|; U) = H(\text{bin}_n) - H_{\text{avg}}(\mu) = O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta}), \quad (9)$$

where $X \sim \mu$ and $U \sim U_n$ are independent, and I is the mutual information function. Thus, small-bias implies the weaker notion of average-case pseudobinomiality.

The key behind the average L_2 -bound (7) is the following lemma which is inspired by an argument in the paper of Viola [27].

Lemma (Variance bound). If $f : \{0, 1\}^n \rightarrow \mathbb{C}$ and μ is a δ -biased probability distribution on $\{0, 1\}^n$, then

$$E_{u \sim U_n} |E_{\sigma_u \mu} f - E_{U_n} f|^2 \leq \delta^2 (E_{U_n} |f|^2 - |E_{U_n} f|^2).$$

The proof of the variance bound is based on Parseval's equality. We derive (7) by applying the variance bound to $f = I_w$ for all $w \in [0 : n]$, where $I_w : \{0, 1\}^n \rightarrow \{0, 1\}$ is the weight indicator function given by $I_w(x) = 1$ iff $|x| = w$. The average L_1 -bound (8) follows from (7) via Jensen's inequality. The average-weight entropy bound (9) follows from (8), Jensen's inequality, and the negative spectrum lemma.

Sum of spaces conjectures. The work of Viola [7, 19, 27] suggests exploring the derandomization capabilities of small-bias spaces. Reingold and Vadhan asked whether the sum of two independent $n^{-O(1)}$ -biased spaces fools log-space [21]. Since any distribution which fools log-space must be pseudobinomial, a natural question is whether the sum of two independent δ -biased spaces is $O((n\delta)^{\Theta(1)})$ -pseudobinomial. Using the above results, we show in Section 7 that the following are equivalent:

- I) **Pseudobinomiality of sum conjecture.** For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, the \mathbb{F}_2 -sum $X + Y$ is $O((n\delta)^{\Theta(1)})$ -pseudobinomial in the L_1 -sense.
- II) **Entropy of sum conjecture.** For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, the Shannon entropy of the weight of $X + Y$ satisfies $H(|X + Y|) \geq H(\text{bin}_n) - O((n\delta)^{\Theta(1)})$.
- III) **Entropy of sum conjecture: max version.** For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, $H(|X + Y|) \geq \max\{H(|X|), H(|Y|)\} - O((n\delta)^{\Theta(1)})$.
- IV) **Entropy of sum conjecture: min version.** For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, $H(|X + Y|) \geq \min\{H(|X|), H(|Y|)\} - O((n\delta)^{\Theta(1)})$.

If A and B are real valued random variables (taking values in a finite set for instance), a simple conditioning argument show that

$$H(A + B) \geq \max\{H(A), H(B)\}.$$

Unfortunately, the picture is more complex in the context of weight distributions and mod-2 sums; we show that for highly biased X or Y , the entropy $H(|X + Y|)$ can be smaller than $H(|X|)$ and $H(|Y|)$.

The equivalence between (I) and (II) follows from the equivalence between pseudobinominality in the L_1 -sense and the entropy-sense, the fact that small-bias is preserved by restrictions and is invariant under translations, and the local pseudobinominality of small-bias spaces. In (III) and (IV), we get rid of the binomial distribution using (9), which says that the average-weight entropy of a small-bias space is a good approximation of the entropy of the binomial distribution. Namely, to show that (II) follows from (IV), we argue using (9) that there is $u \in \{0, 1\}^n$ such that both $H(|X + u|)$ and $H(|Y + u|)$ are close to $H(\text{bin}_n)$, and hence by (IV) applied to $X + u$ and $Y + u$, $H(|X + Y|) = H(|(X + u) + (Y + u)|)$ is close to $H(\text{bin}_n)$. The fact that (III) follows from (II) is based on (6), which asserts that the entropy of the weight distribution of a probability distribution on $\{0, 1\}^n$ with small bias cannot significantly exceed $H(\text{bin}_n)$.

Finally, we note that, by conditioning on Y , the above conjectures follow from the following (possibly stronger) conjecture.

- V) **Lower sandwiching the weight-entropy function conjecture.** For any δ -biased random vector $X \in \{0, 1\}^n$, the weight-entropy function $h : \{0, 1\}^n \rightarrow \mathbb{R}$ given by $h(u) \stackrel{\text{def}}{=} H(|X + u|)$ has a lower sandwiching function $g \leq h$ such that $E_{U_n}(h - g) = O((n\delta)^{\Theta(1)})$ and $\delta \|\widehat{g}\|_1 = O((n\delta)^{\Theta(1)})$.

Dual characterization of pseudobinominality. In Section 8, we point out the dual characterization of the space of functions fooled by pseudobinominality in terms of tight sandwichability between sums of translations of symmetric functions $\{f_i\}_i$ on subsets of the variables such that the total L_∞ -norm $\sum_i \|f_i\|_\infty$ is small.

1.3 Preliminaries

1.3.1 Fourier transform preliminaries

The study of boolean functions using harmonic analysis methods dates back to the 70's (e.g., [17, 16, 18]). We summarize below some basic notions used in this paper. Identify the hypercube $\{0, 1\}^n$ with the group \mathbb{Z}_2^n . The *characters* of the abelian group \mathbb{Z}_2^n are $\{\chi_z\}_{z \in \mathbb{Z}_2^n}$, where $\chi_z : \{0, 1\}^n \rightarrow \{-1, 1\}$ is given by $\chi_z(x) = (-1)^{\sum_{i=1}^n x_i y_i}$. Consider the vector space $\mathcal{L}(\{0, 1\}^n)$ of complex³ valued functions on $\{0, 1\}^n$ endowed with the inner product $\langle \cdot, \cdot \rangle$ associated with the uniform distribution on $\{0, 1\}^n$:

$$\langle f, g \rangle = E_{U_n} f \bar{g} = \frac{1}{2^n} \sum_x f(x) \overline{g(x)},$$

³Except for Lemma 6.1, all the objects in this paper are over the reals.

where $\bar{\cdot}$ is the complex conjugation operator. The characters $\{\chi_z\}_z$ form an orthonormal basis of $\mathcal{L}(\{0, 1\}^n)$, i.e., for each $z, z' \in \{0, 1\}^n$,

$$\langle \chi_z, \chi_{z'} \rangle = \begin{cases} 1 & \text{if } z = z' \\ 0 & \text{if } z \neq z'. \end{cases}$$

If $f \in \mathcal{L}(\{0, 1\}^n)$, its Fourier transform $\widehat{f} \in \mathcal{L}(\{0, 1\}^n)$ is given by the coefficients of the unique expansion of f in terms of $\{\chi_z\}_z$:

$$f(x) = \sum_z \widehat{f}(z) \chi_z(x) \quad \text{and} \quad \widehat{f}(z) = \langle f, \chi_z \rangle = E_{U_n} f \chi_z.$$

We have $\widehat{\widehat{f}} = 2^n f$ and

$$\langle f, g \rangle = 2^n \langle \widehat{f}, \widehat{g} \rangle = \sum_z \widehat{f}(z) \widehat{g}(z). \quad (10)$$

A special case of (10) is *Parseval's equality*:

$$E_{U_n} |f|^2 = \sum_z |\widehat{f}(z)|^2 = \|\widehat{f}\|_2^2. \quad (11)$$

1.3.2 Information theory preliminaries

We summarize below basic information theoretic definitions and notations used in this paper (see [8] for a general reference). Recall from (4) and (5) the definitions of *relative entropy* D and *Shannon entropy* H . The relative entropy is not symmetric but it satisfies the nonnegativity property $D(\gamma_1 || \gamma_2) \geq 0$ with equality iff $\gamma_1 = \gamma_2$. The entropy is a function of the probability distribution, but in some cases it is convenient to argue on the random variables. If A is a random variable taking values in a finite set, its entropy $H(A) \stackrel{\text{def}}{=} H(\mu_A)$, where μ_A is the probability distribution of A . If B is another random variable taking values in a finite set, for each value of b of B , $H(A|B = b) \stackrel{\text{def}}{=} H(\mu_{A|B=b})$, where $\mu_{A|B=b}$ is the probability distribution of A given $B = b$. The *conditional entropy* of A given B is $H(A|B) \stackrel{\text{def}}{=} E_{b \sim \mu_B} H(A|B = b)$. The *mutual information* is $I(A; B) \stackrel{\text{def}}{=} H(A) - H(A|B)$. The *joint entropy* is $H(A, B) \stackrel{\text{def}}{=} H(\mu_{A,B})$, where $\mu_{A,B}$ is the joint probability distributions of A and B . The basic properties of mutual information are:

$$I(A; B) = I(B; A) = H(A, B) - H(A) - H(B) \geq 0.$$

2 Entropy of weight distributions and small-bias

In this section, we elaborate on the material introduced in Section 1.2.1. First we derive the entropy-difference converse bound from the negative spectrum lemma, then we give two specific applications of the negative spectrum lemma in Section 2.1. The proof of the negative spectrum lemma is in Section 9. In Section 2.2, we compare the relative entropy similarity measure to the L_1 -norm in the proximity of the binomial distribution.

Definition 2.1 (Log-binomial function). Let $L : \{0, 1\}^n \rightarrow \mathbb{R}$ be given by:

$$L(x) = \lg \text{bin}_n(|x|).$$

Lemma 2.2. *Let μ be a probability distribution on $\{0, 1\}^n$, then*

$$H(\text{bin}_n) - H(\bar{\mu}) = (E_\mu L - E_{U_n} L) + D(\bar{\mu} \parallel \text{bin}_n)$$

Proof. Expand

$$D(\bar{\mu} \parallel \text{bin}_n) = \sum_w \bar{\mu}(w) \lg \frac{\bar{\mu}(w)}{\text{bin}_n(w)} = - \sum_w \bar{\mu}(w) \lg \text{bin}_n(w) - H(\bar{\mu}) = -E_\mu L - H(\bar{\mu}).$$

Thus, $H(\text{bin}_n) - H(\bar{\mu}) = E_\mu L + H(\text{bin}_n) + D(\bar{\mu} \parallel \text{bin}_n)$. The lemma then follows from noting that $H(\text{bin}_n) = -E_{U_n} L$. \square

The key behind the entropy-difference converse bound is the following lemma whose proof is in Section 9.

Lemma 2.3 (Negative spectrum lemma). $\hat{L}(z) \leq 0$ for each $z \in \{0, 1\}^n$. Moreover, $\hat{L}(z) = 0$ if $|z|$ odd, and $\hat{L}(z) < 0$ if $z \neq 0$ and $|z|$ even.

Corollary 2.4. $\|\hat{L}\|_1 = n$.

Proof. The key point is that $\hat{L} \leq 0$ by the negative spectrum lemma, hence

$$\|\hat{L}\|_1 = - \sum_z \hat{L}(z) = -L(0) = -\lg \text{bin}_n(0) = n.$$

\square

Corollary 2.5. *Let μ be a probability distribution on $\{0, 1\}^n$. If μ is δ -biased, then*

$$|E_\mu L - E_{U_n} L| \leq \delta n.$$

Proof. Consider the Fourier expansion of the log-binomial function L : $L(x) = \sum_z \hat{L}(z) \chi_z(z)$. Thus, $E_\mu L - E_{U_n} L = \sum_{z \neq 0} \hat{L}(z) E_\mu \chi_z(z)$ since $\hat{L}(0) = E_{U_n} L$. It follows that

$$|E_{U_n} L - E_\mu L| = \left| \sum_{z \neq 0} \hat{L}(z) E_\mu \chi_z \right| \leq \delta \|\hat{L}\|_1^{\neq 0},$$

where $\|\hat{L}\|_1^{\neq 0} = \sum_{z \neq 0} |\hat{L}(z)| \leq \|\hat{L}\|_1 = n$. \square

Corollary 2.6 (Entropy-difference converse bound). *Let μ be a probability distribution on $\{0, 1\}^n$. If μ is δ -biased, then*

$$D(\bar{\mu} \parallel \text{bin}_n) \leq n\delta + H(\text{bin}_n) - H(\bar{\mu}). \quad (12)$$

Hence

$$\|\bar{\mu} - \text{bin}_n\|_1^2 \leq (2 \ln 2)(n\delta + H(\text{bin}_n) - H(\bar{\mu})). \quad (13)$$

Proof. By Lemma 2.2 and Corollary 2.5:

$$D(\bar{\mu} \parallel \text{bin}_n) = E_{U_n} L - E_\mu L + H(\text{bin}_n) - H(\bar{\mu}) \leq n\delta + H(\text{bin}_n) - H(\bar{\mu}).$$

Bound (13) follows from Bound (12) and Pinsker's bound (Part (b) of Theorem 1.1). \square

We note below that one consequence of the above is that if μ has small bias, then unlike arbitrary probability distribution on $\{0, 1\}^n$, $H(\bar{\mu})$ cannot be significantly larger than $H(\text{bin}_n)$ if it has small bias.

Corollary 2.7 (Maximum entropy of the weight distributions of small-bias spaces). *Let μ be a probability distribution on $\{0, 1\}^n$. If μ is δ -biased, then $H(\bar{\mu}) \leq H(\text{bin}_n) + n\delta$.*

Proof. The bound follows from Corollary 2.6 since $D(\bar{\mu}||\text{bin}_n) \geq 0$. □

2.1 Specific applications of the negative spectrum lemma

2.1.1 Entropy of distributions with nonnegative Fourier spectrum

A consequence of negative spectrum lemma is that if the Fourier transform of a probability distribution μ on $\{0, 1\}^n$ is nonnegative, then $H(\text{bin}_n) - H(\bar{\mu}) \leq D(\bar{\mu}||\text{bin}_n)$. For instance, if μ is uniformly supported by a linear code, then its Fourier transform is nonnegative. Another example of a distribution with nonnegative Fourier transform is the convolution of a probability distribution on $\{0, 1\}^n$ with itself.

The bound is slightly better than the one resulting from combining Pinsker's bound and the entropy-difference bound: $H(\text{bin}_n) - H(\bar{\mu}) \leq \beta \lg \frac{(n+1)}{\beta}$, where $\beta = \sqrt{(2 \ln 2)D(\bar{\mu}||\text{bin}_n)}$ under the assumption that $D(\bar{\mu}||\text{bin}_n) \leq \frac{1}{8 \ln 2}$ ⁴.

Corollary 2.8. *Let μ be a probability distribution on $\{0, 1\}^n$ whose Fourier transform is nonnegative, i.e., $E_\mu \chi_z \geq 0$ for each $z \in \{0, 1\}^n$. Then $H(\text{bin}_n) - H(\bar{\mu}) \leq D(\bar{\mu}||\text{bin}_n)$.*

Proof. As in Corollary 2.5, consider the Fourier expansion of the log-binomial function L : $L(x) = \sum_z \hat{L}(z) \chi_z(z)$. Thus, $E_\mu L - E_{U_n} L = \sum_{z \neq 0} \hat{L}(z) E_\mu \chi_z(z) \leq 0$ since, by Lemma 2.3, $\hat{L} \leq 0$. Hence the claim follows from Lemma 2.2. □

Below are examples of distributions with the nonnegative Fourier transform property:

- Let $C \subset \mathbb{F}_2^n$ be an \mathbb{F}_2 -linear code, and let μ_C be the distributions resulting from choosing a uniformly random element of C . Thus,

$$E_{\mu_C} \chi_z = \begin{cases} 1 & \text{if } z \in C^\perp \\ 0 & \text{otherwise,} \end{cases}$$

where C^\perp is the dual of C . It follows that $H(\text{bin}_n) - H(\bar{\mu}_C) \leq D(\bar{\mu}_C||\text{bin}_n)$.

- Let μ be probability distribution on $\{0, 1\}^n$ and consider the convolution $\mu * \mu$ of μ with itself, i.e., $(\mu * \mu)(a) = \sum_x \mu(y) \mu(x+a)$. That is, choose $x \sim \mu$ and $y \sim \mu$ independently and output their \mathbb{F}_2 -sum $x+y$. We have $E_{\mu * \mu} \chi_z = E_{x,y \sim \mu} \chi_z(x+y) = (E_\mu \chi_z)^2 \geq 0$, for each $z \in \{0, 1\}^n$. It follows that $H(\text{bin}_n) - H(\bar{\mu} * \bar{\mu}) \leq D(\bar{\mu} * \bar{\mu}||\text{bin}_n)$.

⁴ By Pinsker's bound, $\epsilon \stackrel{\text{def}}{=} \|\text{bin}_n - \bar{\mu}\|_1 \leq \beta \stackrel{\text{def}}{=} \sqrt{(2 \ln 2)D(\bar{\mu}||\text{bin}_n)} \leq \frac{1}{2}$. Thus, by the entropy-difference bound $H(\text{bin}_n) - H(\bar{\mu}) \leq \epsilon \lg \frac{(n+1)}{\epsilon} \leq \beta \lg \frac{(n+1)}{\beta}$ since the function $x \lg \frac{n+1}{x}$ is increasing for $0 \leq x \leq \frac{1}{2}$, for all $n \geq 1$.

2.1.2 Entropy of even weight strings

It follows from the negative spectrum lemma that the entropy of the weight distribution of the even weight strings is strictly larger than that of odd weight strings if n is even.

Corollary 2.9. *Let $E_n \in \{0, 1\}^n$ be a uniformly random vector of even weight and $O_n \in \{0, 1\}^n$ a uniformly random vector of odd weight.*

- a) *If n is odd, then $H(|E_n|) = H(|O_n|)$.*
- b) *If n is even, then $H(|E_n|) > H(|O_n|)$.*

Proof. If n is odd, then $|E_n|$ and $n - |O_n|$ are identically distributed, hence $H(|E_n|) = H(n - |O_n|) = H(|O_n|)$. Assume in what follows that n is even.

We have

$$\begin{aligned}
 H(|O_n|) - H(|E_n|) &= - \sum_{w \text{ odd}} \frac{\binom{n}{w}}{2^{n-1}} \lg \frac{\binom{n}{w}}{2^{n-1}} + \sum_{w \text{ even}} \frac{\binom{n}{w}}{2^{n-1}} \lg \frac{\binom{n}{w}}{2^{n-1}} \\
 &= \sum_w (-1)^w \frac{\binom{n}{w}}{2^{n-1}} \lg \frac{\binom{n}{w}}{2^{n-1}} \\
 &= 2 \sum_w (-1)^w \frac{\binom{n}{w}}{2^n} \lg \frac{\binom{n}{w}}{2^n} \quad (\text{Since } \sum_w (-1)^w \binom{n}{w} = 0) \\
 &= 2\widehat{L}(\vec{1}),
 \end{aligned}$$

where $\vec{1} \in \{0, 1\}^n$ is the all ones vector. Hence (b) follows from the negative spectrum lemma for $z = \vec{1}$. \square

2.2 Relative entropy versus L1 in the proximity of the binomial

Pinsker's bound implies that relative entropy is in general stronger than L_1 . We note in Corollary 2.10 that for weight distributions $\bar{\mu}$ of probability distributions μ on $\{0, 1\}^n$, $D(\bar{\mu} || \text{bin}_n)$ is equivalent to $\|\bar{\mu} - \text{bin}_n\|_1$ in the $n^{-\Theta(1)}$ -error regime.

Unlike in the entropy-difference converse bound, small-bias is not essential here; the bound in Part (b) below, which assumes small bias, is only slightly better than Part (a). Parts (c) and (d) are simple variations of (a) and (b) without the assumption $\|\gamma_1 - \gamma_2\|_1 \leq \frac{1}{2}$. We will use Part (d) in Section 6.

Corollary 2.10. *Let μ be a probability distribution on $\{0, 1\}^n$ and $\epsilon = \|\bar{\mu} - \text{bin}_n\|_1$. Then:*

- a) $D(\bar{\mu} || \text{bin}_n) \leq n\epsilon + \epsilon \lg \frac{n+1}{\epsilon}$ if $\epsilon \leq \frac{1}{2}$.
- b) $D(\bar{\mu} || \text{bin}_n) \leq n\delta + \epsilon \lg \frac{n+1}{\epsilon}$ if $\epsilon \leq \frac{1}{2}$ and μ is δ -biased
- c) $D(\bar{\mu} || \text{bin}_n) \leq n\epsilon + 3\epsilon \lg \frac{n+1}{\epsilon}$ if $n \geq 7$
- d) $D(\bar{\mu} || \text{bin}_n) \leq n\delta + 3\epsilon \lg \frac{n+1}{\epsilon}$ if μ is δ -biased and if $n \geq 7$.

Proof. Part (b) follows from combining Corollary 2.6 with the entropy-difference bound (Part (b) of Theorem 1.1). The bound in Part (d) follows from Corollary 2.6 and Lemma 2.11 below which is a simple variation of the entropy-difference bound that does not assume $\|\gamma_1 - \gamma_2\|_1 \leq \frac{1}{2}$. To establish (a) and (c), note that by Lemma 2.2, we have

$$D(\bar{\mu} \parallel \text{bin}_n) = E_{U_n} L - E_{\mu} L + H(\text{bin}_n) - H(\bar{\mu}).$$

Moreover, $E_{U_n} L - E_{\mu} L = \sum_w (\bar{\mu}(w) - \text{bin}_n(w)) \lg \text{bin}_n(w)$. Thus, $|E_{U_n} L - E_{\mu} L| \leq \|\bar{\mu} - \text{bin}_n\|_1 \times \|L\|_{\infty} = \epsilon n$. It follows that $D(\bar{\mu} \parallel \text{bin}_n) \leq \epsilon n + |H(\text{bin}_n) - H(\bar{\mu})|$. Thus, (a) and (c) follow from the entropy-difference bound and its variation below. \square

Lemma 2.11. *Let γ_1 and γ_2 be two probability distributions on a finite set \mathcal{X} . If $|\mathcal{X}| \geq 8$, then $|H(\gamma_1) - H(\gamma_2)| \leq 3\epsilon \lg \frac{|\mathcal{X}|}{\epsilon}$, where $\epsilon = \|\gamma_1 - \gamma_2\|_1$.*

Proof. Let $m = |\mathcal{X}|$. By the entropy-difference bound, $|H(\gamma_1) - H(\gamma_2)| \leq \epsilon \lg \frac{m}{\epsilon}$ if $\epsilon \leq \frac{1}{2}$. Note that $\|\gamma_1 - \gamma_2\|_1$ is at most 2 and $|H(\gamma_1) - H(\gamma_2)|$ is at most $\lg m$. Thus, it is enough to verify that $3\epsilon \lg \frac{m}{\epsilon} \geq \lg m$ for all $\frac{1}{2} \leq \epsilon \leq 2$. We have $3\epsilon \lg \frac{m}{\epsilon} \geq \frac{3}{2} \lg \frac{m}{2} = \frac{3}{2} \lg m - \frac{3}{2} \geq \lg m$ for $m \geq 8$. \square

Remark 2.12. *Recall that the relative entropy is not symmetric. It is worth mentioning that unlike $D(\bar{\mu} \parallel \text{bin}_n)$, the similarity measure $D(\text{bin}_n \parallel \bar{\mu})$ is problematic in our context since if $\bar{\mu}(w) = 0$ for some $w \in [0 : n]$, then $D(\text{bin}_n \parallel \bar{\mu}) = \infty$. This issue can be fixed by small perturbations of μ to guarantee that for some $\beta > 0$, $\bar{\mu}(w) > \beta$ for all $w \in [0 : n]$. By arguing as Corollary 2.10, it follows easily from the entropy-difference bound that under this assumption, $D(\text{bin}_n \parallel \bar{\mu}) \leq \epsilon \lg \frac{1}{\beta} + \epsilon \lg \frac{n+1}{\epsilon}$, if $\epsilon = \|\bar{\mu} - \text{bin}_n\|_1 \leq \frac{1}{2}$. Accordingly, for $\beta = 2^{-\Omega(n^{\Theta(1)})}$, we get from Pinsker's bound that $D(\text{bin}_n \parallel \bar{\mu})$ is equivalent to $\|\bar{\mu} - \text{bin}_n\|_1$ in the $n^{-\Theta(1)}$ -error regime.*

3 Pseudobinomial spaces

In this section we elaborate on the notion of pseudobinomial probability distributions introduced in Section 1.2.2. We conclude from the entropy-difference bound and the entropy-difference converse bound that for spaces with $n^{-\Theta(1)}$ -small bias, pseudobinomiality in the L_1 -sense is equivalent to pseudobinomiality in the entropy-sense in the $n^{-\Theta(1)}$ -error regime.

First we elaborate on the notations introduced in Section 1.2.2. If μ is a probability distribution on $\{0, 1\}^n$ and $u \in \{0, 1\}^n$ is a translation vector, define the *translation* $\sigma_u \mu$ of μ by u to be the probability distribution on $\{0, 1\}^n$ given by $(\sigma_u \mu)(x) = \mu(x + u)$. If $I \subset [n]$ is nonempty, the *restriction* μ^I of μ on I is the probability distribution on $\{0, 1\}^I$ given by $\mu^I(y) = \mu(x : x_I = y)$, for each $y \in \{0, 1\}^I$. We will also use the previously defined notations on $\{0, 1\}^n$ for probability distributions on $\{0, 1\}^I$. For instance, $\text{bin}_{|I|}$ is the binomial distribution on $[0 : |I|]$, and if $u \in \{0, 1\}^I$, then $\overline{\sigma_u \mu^I}$ is the weight distribution of the translation $\sigma_u \mu^I$ of the restriction μ^I of μ .

Definition 3.1 (Pseudobinomiality in the L_p -sense). *Let $\epsilon > 0$ and consider the L_p -norm, where $p = 1, 2$ or ∞ . A probability distribution μ on $\{0, 1\}^n$ is called ϵ -pseudobinomial in the L_p -sense if for each nonempty set of indices $I \subset [n]$ and each translation vector $u \in \{0, 1\}^I$, we have $\|\overline{\sigma_u \mu^I} - \text{bin}_{|I|}\|_p \leq \epsilon$. That is, the weight distribution of each translation of a restriction of μ is ϵ -close to the binomial distribution in the L_p -sense.*

A subset $S \subset \{0, 1\}^n$ is called ϵ -pseudobinomial in the L_p -sense if the probability distribution μ_S on $\{0, 1\}^n$ resulting from choosing a uniformly random element of S is ϵ -pseudobinomial in the L_p -sense.

Pseudobinomiality in the L_1 -sense, L_2 -sense, and L_∞ -sense are equivalent in the $\epsilon = n^{-\Theta(1)}$ regime. Namely, with the L_p pseudobinomiality error of μ defined as $\epsilon_p(\mu) \stackrel{\text{def}}{=} \min_{I,u} \|\overline{\sigma_u \mu^I} - \text{bin}_{|I|}\|_p$, we have

$$\epsilon_\infty(\mu) \leq \epsilon_2(\mu) \leq \epsilon_1(\mu) \leq \sqrt{(n+1)\epsilon_2(\mu)} \leq (n+1)\epsilon_\infty(\mu).$$

We focus mostly in what follows on the L_1 -sense. Recall the total variation equation (3):

$$\|\overline{\sigma_u \mu^I} - \text{bin}_{|I|}\|_1 = 2 \max_{A \subset I} |\overline{\sigma_u \mu^I}(A) - \text{bin}_{|I|}(A)|.$$

Thus, μ is ϵ -pseudobinomial in L_1 -sense iff $2|\overline{\sigma_u \mu^I}(A) - \text{bin}_{|I|}(A)| \leq \epsilon$, for all nonempty $I \subset [n]$, all $A \subset I$ and $u \in \{0, 1\}^I$.

Lemma 3.2. *If a probability distribution μ on $\{0, 1\}^n$ is ϵ -pseudobinomial in L_1 -sense, then it is ϵ -biased.*

Proof. Consider any nonzero $z \in \{0, 1\}^n$, let I be the support of z , and let \mathcal{E}_I the set of even numbers in $[0 : |I|]$. Thus, $|E_\mu \chi_z| = 2|\overline{\mu^I}(\mathcal{E}_I) - \text{bin}_{|I|}(\mathcal{E}_I)|$. The lemma follows from the total variation equation. \square

Compared to ϵ -bias, ϵ -pseudobinomiality in the L_1 -sense extends the requirement $2|\overline{\sigma_u \mu^I}(A) - \text{bin}_{|I|}(A)| \leq \epsilon$ from $A = \mathcal{E}_I$ to all the subsets $A \subset [0 : |I|]$ (note that $|\overline{\sigma_u \mu^I}(\mathcal{E}_I) - \text{bin}_{|I|}(\mathcal{E}_I)| = |\overline{\mu^I}(\mathcal{E}_I) - \text{bin}_{|I|}(\mathcal{E}_I)|$, for all $u \in \{0, 1\}^I$, since $\mathcal{E}_I + u$ is either \mathcal{E}_I or its complement). In this sense, ϵ -pseudobinomiality in the L_1 -sense is a natural extension of the ϵ -bias property, which is also invariant under translations and preserved by restrictions.

Problem 3.3. *Given n and ϵ , explicitly construct a $\text{poly}(n, \frac{1}{\epsilon})$ -size subset $S \subset \{0, 1\}^n$ such that S is ϵ -pseudobinomial in the L_1 -sense.*

For convenience, we repeat below the following two definitions introduced in Section 1.2.2.

Definition 3.4 (Minimum weight entropy). *If μ is a probability distribution on $\{0, 1\}^n$, define the min-weight entropy of μ :*

$$H_{\min}(\mu) = \min_{u \in \{0, 1\}^n} H(\overline{\sigma_u \mu}),$$

i.e., $H_{\min}(\mu)$ is the minimum Shannon entropy of the weight distribution of a translation of μ .

Definition 3.5 (Pseudobinomiality in the entropy-sense). *A probability distribution μ on $\{0, 1\}^n$ is called ϵ -pseudobinomial in the entropy-sense if for each nonempty index subset $I \subset [n]$, we have $H_{\min}(\mu^I) \geq H(\text{bin}_{|I|}) - \epsilon$.*

Using the entropy-difference bound and the entropy-difference converse bound, we obtain the following equivalence.

Corollary 3.6 (Pseudobinomiality: L_1 and entropy equivalence). *let μ be a δ -biased probability distribution μ on $\{0, 1\}^n$ and $\epsilon > 0$. Then:*

- a) *If $\epsilon \leq 1/2$ and μ is ϵ -pseudobinomial in the L_1 -sense, then it $\epsilon \lg \frac{n+1}{\epsilon}$ -pseudobinomial in the entropy-sense.*

b) If μ is ϵ -pseudobinomial in entropy-sense then it is $\sqrt{(2 \ln 2)(n\delta + \epsilon)}$ -pseudobinomial in the L_1 -sense.

Proof. Part (a) follows from the Difference in entropy Bound (Lemma 1.1). Part (b) follows from the Difference in entropy Converse Bound (Theorem 2.6) and the fact that δ -bias is preserved by restrictions and invariant under translations. \square

Question 3.7. *Does small-bias follow from pseudobinomiality in entropy? We know from Lemma 3.2 that ϵ -pseudobinomiality in the L_1 -sense implies ϵ -bias. Does ϵ -pseudobinomiality in the entropy-sense imply δ -bias, where δ is small (e.g., $\delta = (\epsilon n)^c$, for some absolute constant $c > 0$)?*

4 Min-weight entropy, average-weight entropy, and the binomial entropy

In this section, we elaborate on the notion of min-weight entropy and we study the related notion of average-weight entropy. There are distributions μ on $\{0, 1\}^n$ such that the weight distribution $\bar{\mu}$ of μ is the uniform distribution on $[0 : n]$, and hence $H(\bar{\mu}) = \lg(n + 1) \approx 2H(\text{bin}_n)$. We note below that each probability μ on $\{0, 1\}^n$ has a translation whose weight distribution has entropy at most $H(\text{bin}_n)$, and hence $H_{\min}(\mu) \leq H(\text{bin}_n)$. To do so, we need the notion of average-weight entropy which we obtain by replacing the min with an average and we can interpret in terms of conditional entropy.

Definition 4.1 (Average-weight entropy). *If μ is a probability distribution on $\{0, 1\}^n$, define the average-weight entropy of μ :*

$$H_{\text{avg}}(\mu) = E_{u \sim U_n} H(\overline{\sigma_u \mu}),$$

i.e., $H_{\text{avg}}(\mu)$ is the average Shannon entropy of the weight distribution of a random translation of μ . In terms of conditional entropy, we have the following equivalent definition. Let $X \in \{0, 1\}^n$ a random vector generated according to μ , and let $U \in \{0, 1\}^n$ be a uniformly distributed random vector independent from X . Then

$$H_{\text{avg}}(\mu) = H(|X + U| \mid U).$$

By definition, $H_{\min}(\mu) \leq H_{\text{avg}}(\mu)$. On the other hand, $H(\text{bin}_n) - H_{\text{avg}}(\mu)$ is the mutual information between the weight $|X + U|$ of $X + U$ and U :

$$H(\text{bin}_n) - H_{\text{avg}}(\mu) = H(|X + U|) - H(|X + U| \mid U) = I(|X + U|; U) \geq 0.$$

It follows that $H_{\min}(\mu) \leq H_{\text{avg}}(\mu) \leq H(\text{bin}_n)$. We argue below that the inequality $H_{\text{avg}}(\mu) \leq H(\text{bin}_n)$ is strict unless μ is the uniform distribution. To do so, we need the following lemma which we will use also in Section 6.

Lemma 4.2. *Let μ be a probability distribution on $\{0, 1\}^n$, then*

$$H(\text{bin}_n) - E_{u \sim U_n} H(\overline{\sigma_u \mu}) = E_{u \sim U_n} D(\overline{\sigma_u \mu} \parallel \text{bin}_n). \quad (14)$$

Proof. By Lemma 2.2, for each $u \in \{0, 1\}^n$, we have $H(\text{bin}_n) - H(\overline{\sigma_u \mu}) = e(u) + D(\overline{\sigma_u \mu} \| \text{bin}_n)$, where $e(u) = E_{\sigma_u \mu} L - E_{U_n} L$. The Lemma follows from the fact that the average $E_{U_n} e = 0$:

$$E_{u \sim U_n} e(u) = E_{u \sim U_n} E_{x \sim \mu} L(x + u) - E_{U_n} L = E_{x \sim \mu} E_{u \sim U_n} L(x + u) - E_{U_n} L = 0.$$

□

Lemma 4.3. *If $H_{\text{avg}}(\mu) = H(\text{bin}_n)$, then μ is the uniform distribution U_n on $\{0, 1\}^n$.*

Proof. If $H_{\text{avg}}(\mu) = H(\text{bin}_n)$, then the LHS of (14) is zero. Since $D(\overline{\sigma_u \mu} \| \text{bin}_n) \geq 0$, we must have $D(\overline{\sigma_u \mu} \| \text{bin}_n) = 0$, i.e., $\overline{\sigma_u \mu} = \text{bin}_n$, for each u . It follows that for each symmetric function $s : \{0, 1\}^n \rightarrow \mathbb{R}$ (i.e., $s(x)$ depends only on $|x|$) and for each translation vector $u \in \{0, 1\}^n$, we have $E_{\mu} \sigma_u s = E_{U_n} \sigma_u s$, where $\sigma_u s$ is the translation of s by u given by $(\sigma_u s)(x) = s(x + u)$. In particular this is true for the symmetric function $s : \{0, 1\}^n \rightarrow \{0, 1\}$ given by $s(x) = 1$ iff $x = 0$. Since the \mathbb{R} -span of $\{\sigma_u s\}_u$ is the set of all functions $\{0, 1\}^n \rightarrow \mathbb{R}$, we get that $\mu = U_n$. □

In summary, we have the following.

Corollary 4.4 (Min-weight entropy, avg-weight-entropy, and the binomial entropy). *Let μ be a probability distribution on $\{0, 1\}^n$. Then:*

- a) $H_{\min}(\mu) \leq H_{\text{avg}}(\mu) \leq H(\text{bin}_n)$, where the inequalities $H_{\text{avg}}(\mu) \leq H(\text{bin}_n)$ and $H_{\min}(\mu) \leq H(\text{bin}_n)$ are strict unless μ is the uniform distribution U_n on $\{0, 1\}^n$. That is, U_n is the unique maximum min-weight entropy distribution and the unique maximum average-weight entropy distribution.
- b) If $X \in \{0, 1\}^n$ is a random vector generated according to μ , and $U \in \{0, 1\}^n$ is uniformly distributed random vector independent from X . Then

$$0 \leq I(|X + U|; U) = H(\text{bin}_n) - H_{\text{avg}}(\mu) = E_{u \sim U_n} D(\overline{\sigma_u \mu} \| \text{bin}_n).$$

We will argue in Section 6 that if μ has small bias, then $E_{u \sim U_n} D(\overline{\sigma_u \mu} \| \text{bin}_n)$ is small. Namely, if μ is δ -biased, then $E_{u \sim U_n} D(\overline{\sigma_u \mu} \| \text{bin}_n) = O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta})$. It follows that small bias is enough to guarantee that the average-weight entropy is close to that of the binomial (see Corollary 6.4).

A final remark, is that it is possible that $H_{\min}(\mu) = H_{\text{avg}}(\mu)$ for $\mu \neq U_n$. For instance, if the support of μ contains only one vector, or two vectors whose hamming distance is odd.

5 Local pseudobinomiality

We note in this section that small-bias does not imply pseudobinomiality in the L_1 -sense or the entropy-sense even if the bias is exponentially small, but it is enough to guarantee local pseudobinomiality on small subsets of indices.

Let $S_3 = \{x \in \{0, 1\}^n : |x| \equiv 0 \pmod{3}\}$, and consider the distribution μ_{S_3} resulting from choosing a uniformly random element of S . Then μ_{S_3} is $2^{-\Omega(n)}$ -biased (see [10]), but $\|\overline{\mu_{S_3}} - \text{bin}_n\|_{\infty} = \Theta(\frac{1}{\sqrt{n}})$ and $\|\overline{\mu_{S_3}} - \text{bin}_n\|_1 = \Theta(1)$ (this follows easily from de Moivre-Laplace normal approximation of the binomial (see Theorem 9.4)). Hence, by the entropy-difference converse bound (Corollary 2.6), $H(\text{bin}_n) - H(\overline{\mu_{S_3}}) \geq \frac{1}{2 \ln 2} \|\overline{\mu_{S_3}} - \text{bin}_n\|_1^2 - n\delta = \Theta(1)$.

Similarly k -wise independence does not imply small-pseudobinomiality unless $k = n$. For instance, let S_2 be the set even weight vectors, then μ_{S_2} is $(n - 1)$ -wise independent, but $\|\overline{\mu_{S_2}} - \text{bin}_n\|_\infty = \Theta(\frac{1}{\sqrt{n}})$ and $\|\overline{\mu_{S_2}} - \text{bin}_n\|_1 = \Theta(1)$.

However, we can guarantee proximity to the binomial distribution for small subsets of indices $I \subset [n]$. In particular, if $|I| = O(\log n)$, then sufficiently small $\delta = n^{-\Theta(1)}$ is enough.

Lemma 5.1 (Local pseudobinomiality). *Let $k \geq 1$ and μ be a δ -biased probability distribution μ on $\{0, 1\}^n$. Then for each nonempty set of indices $I \subset [n]$ of size $|I| \leq k$, and each translation vector $u \in \{0, 1\}^I$, we have $\|\overline{\sigma_u \mu^I} - \text{bin}_{|I|}\|_1 \leq 2\delta 2^{k/2}$.*

The proof of the lemma follows from a standard argument known as Vazirani XOR Lemma [26] (see also [22]), which we include below for completeness.

Proof. By the total variation expression (3),

$$\|\overline{\sigma_u \mu^I} - \text{bin}_{|I|}\|_1 = 2 \max_{A \subset [0:|I|]} |\overline{\sigma_u \mu^I}(A) - \text{bin}_{|I|}(A)|.$$

Let $f : \{0, 1\}^I \rightarrow \{0, 1\}$ be given by $f(x) = 1$ iff $|x + u| \in A$. Thus,

$$|\overline{\sigma_u \mu^I}(A) - \text{bin}_{|I|}(A)| = |E_\mu f - E_{U_n} f| \leq \delta \|f\|_1 \leq \delta \sqrt{2^{|I|}} \|f\|_2 \leq \delta 2^{k/2},$$

where the last equality follows from Parseval's equality $\|f\|_2^2 = E_{U_n} f^2 = E_{U_n} f \leq 1$ since f is 0/1 valued. \square

6 Average case pseudobinomiality

In this section, we show that if μ is a δ -biased probability distribution, then $E_{u \sim U_n} \|\overline{\sigma_u \mu} - \text{bin}_n\|_1 \leq \delta \sqrt{n + 1}$, i.e., the average L_1 -distance between the binomial distribution and the weight distribution of the translation of μ by a random vector in $\{0, 1\}^n$ is most $\delta \sqrt{n + 1}$. Thus, if the bias is small, almost all translation of μ have weight distributions close to the binomial distribution. In this sense, small-bias implies average case pseudobinomiality. We conclude a similar bound for average-weight entropy: if μ be a δ -biased, then $H_{\text{avg}}(\mu) = H(\text{bin}_n) - O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta})$.

The following Lemma is inspired by the paper of Viola [27] (the argument used to establish Lemma 3 in [27]) and it follows from Parseval's equality.

Lemma 6.1 (Variance bound). *If $f : \{0, 1\}^n \rightarrow \mathbb{C}$ and μ is a δ -biased probability distribution on $\{0, 1\}^n$, then*

$$E_{u \sim U_n} |E_{\sigma_u \mu} f - E_{U_n} f|^2 \leq \delta^2 (E_{U_n} |f|^2 - |E_{U_n} f|^2).$$

Proof. Define $\Delta : \{0, 1\}^n \rightarrow \mathbb{R}$ by $\Delta(u) = E_{\sigma_u \mu} f - E_{U_n} f$. Consider the Fourier expansion of f : $f = \sum_z \hat{f}(z) \chi_z$. Thus,

$$\Delta(u) = E_{y \sim \mu} \sum_z \hat{f}(z) \chi_z(y + u) - E_{U_n} f = \sum_z \chi_z(u) \hat{f}(z) E_\mu \chi_z - E_{U_n} f = \sum_{z \neq 0} \chi_z(u) \hat{f}(z) E_\mu \chi_z$$

since $\chi_0 = 1$ and $\hat{f}(0) = E_{U_n} f$. Hence $\hat{\Delta}(0) = 0$ and $\hat{\Delta}(z) = \hat{f}(z) E_\mu \chi_z$ for each $z \neq 0$. It follows from Parseval's Equality (11) that

$$E_{u \sim U_n} |\hat{\Delta}(z)|^2 = \sum_{z \neq 0} |\hat{f}(z)|^2 (E_\mu \chi_z)^2 \leq \delta^2 \sum_{z \neq 0} |\hat{f}(z)|^2 = \delta^2 (E_{U_n} |f|^2 - |E_{U_n} f|^2)$$

since $\widehat{f}(0) = E_{U_n} f$ and $E_{U_n} |f|^2 = \sum_z |\widehat{f}(z)|^2$, by Parseval's equality. \square

Corollary 6.2 (Average case pseudobinomiality). *Let μ be a δ -biased probability distribution on $\{0, 1\}^n$, then*

- a) $E_{u \sim U_n} \|\overline{\sigma_u \mu} - \text{bin}_n\|_2^2 \leq \delta^2$
- b) $E_{u \sim U_n} \|\overline{\sigma_u \mu} - \text{bin}_n\|_1 \leq \delta \sqrt{n+1}$
- c) $E_{u \sim U_n} D(\overline{\sigma_u \mu} \| \text{bin}_n) = O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta})$.

Proof. If $w \in [0 : n]$, define the indicator function $I_w : \{0, 1\}^n \rightarrow \{0, 1\}$ by $I_w(x) = 1$ iff $|x| = w$. Thus, $\text{bin}_n(w) = E_{U_n} I_w$ and $\overline{\sigma_u \mu}(w) = E_{\sigma_u \mu} I_w$. Hence $\|\overline{\sigma_u \mu} - \text{bin}_n\|_2^2 = \sum_w |E_{\sigma_u \mu} I_w - E_{U_n} I_w|^2$, and accordingly,

$$E_{u \sim U_n} \|\overline{\sigma_u \mu} - \text{bin}_n\|_2^2 = \sum_w E_{u \sim U_n} |E_{\sigma_u \mu} I_w - E_{U_n} I_w|^2.$$

Fix $w \in [0 : n]$. Applying Lemma 6.1 with $f = I_w$, we get

$$E_{u \sim U_n} |E_{\sigma_u \mu} I_w - E_{U_n} I_w|^2 \leq \delta^2 (E_{U_n} |I_w|^2 - |E_{U_n} I_w|^2) \leq \delta^2 E_{U_n} I_w$$

since $E_{U_n} |I_w|^2 = E_{U_n} I_w$ as I_w is 0/1 valued. It follows that $E_{u \sim U_n} \|\overline{\sigma_u \mu} - \text{bin}_n\|_2^2 \leq \delta^2 \sum_w E_{U_n} I_w = \delta^2$ since $\sum_w I_w = 1$.

Part (b) follows from Part (a) via Jensen's inequality applied to $g(u, w) = \overline{\sigma_u \mu}(w) - \text{bin}_n(w)$ (namely, $(E_{u, w} |g(u, w)|)^2 \leq E_{u, w} |g(u, w)|^2$).

We derive Part (c) from Part (b) and Part (d) of Corollary 2.10. For each $u \in \{0, 1\}^n$, let $\epsilon(u) = \|\overline{\sigma_u \mu} - \text{bin}_n\|_1$. By Part (d) of Corollary 2.10, we have $D(\overline{\sigma_u \mu} \| \text{bin}_n) \leq n\delta + 3\epsilon(u) \lg \frac{n+1}{\epsilon(u)}$ if $n \geq 7$. Since the function $x \lg \frac{n+1}{x}$ is concave, $E_{u \sim U_n} \epsilon(u) \lg \frac{n+1}{\epsilon(u)} \leq E_{u \sim U_n} \epsilon(u) \lg \frac{n+1}{E_{u \sim U_n} \epsilon(u)}$. By Part (b), we have $E_{u \sim U_n} \epsilon(u) \leq \delta \sqrt{n+1}$. The function $x \lg \frac{n+1}{x}$ is increasing for all $0 \leq x \leq (n+1)/e$, thus if $\delta \sqrt{n+1} \leq (n+1)/e$, we have

$$E_{u \sim U_n} D(\overline{\sigma_u \mu} \| \text{bin}_n) \leq n\delta + 3\delta \sqrt{n+1} \lg \frac{\sqrt{n+1}}{\delta} = O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta}).$$

Finally, note that the condition $\delta \sqrt{n+1} \leq (n+1)/e$ can be ignored since the bias $\delta \leq 1$ and $\sqrt{n+1} < (n+1)/e$ for all $n \geq 7$. \square

Remark 6.3. *The bound in Part (c) uses Part (d) of Corollary 2.10, which relies on the negative spectrum lemma. If instead of Part (d) of Corollary 2.10, we use the bound in Part (c) which does not depend on the negative spectrum lemma, we get the slightly weaker bound $E_{u \sim U_n} D(\overline{\sigma_u \mu} \| \text{bin}_n) = O(n^{3/2}\delta + \sqrt{n}\delta \lg \frac{1}{\delta})$.*

It follows from (c) and from Part (b) of Corollary 4.4 that small bias is enough to guarantee that the average-weight entropy is close to that of the binomial:

Corollary 6.4 (Average-weight entropy of small-bias spaces). *Let μ be a δ -biased probability distribution on $\{0, 1\}^n$. Then*

$$0 \leq I(|X + U|; U) = H(\text{bin}_n) - H_{\text{avg}}(\mu) = O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta}),$$

where $X \in \{0, 1\}^n$ is generated according to μ , and $U \in \{0, 1\}^n$ is a uniformly distributed random vector independent from X .

Remark 6.5. Note that since $I(|X + U|; U) = I(U; |X + U|) = H(U) - H(U | |X + U|)$ and $H(U) = n$, we have $H(U | |X + U|) = n - O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta})$.

Lemma 6.1 fails if instead of a δ -biased probability distribution we have a probability distribution uniformly supported by an \mathbb{F}_2 -linear code $Q \subsetneq \mathbb{F}_2^n$. By a more involved argument, a statement similar to Part (b) of Corollary 6.2 can be derived for linear codes with large bilateral minimum distance. The *bilateral minimum* distance of an \mathbb{F}_2 -linear code $Q \subset \mathbb{F}_2^n$ is the maximum d such that all nonzero codewords have weights between d and $n - d$.

Theorem 6.6 ([5, 6] **Global average case pseudobinomiality for linear codes**). ⁵ Let $Q \subsetneq \{0, 1\}^n$ be an \mathbb{F}_2 -linear code whose dual has bilateral minimum distance at least $2t + 1$, where $t \geq 1$ is an integer.

If $t \geq 1$,

$$E_{u \sim U_n} \|\overline{\mu_{Q+u}} - \text{bin}_n\|_\infty \leq \min \left\{ \left(e \ln \frac{n}{2t} \right)^t \left(\frac{2t}{n} \right)^{\frac{t}{2}}, \sqrt{2} e^{-\frac{t}{10}} \right\}.$$

If $t \geq 3$,

$$E_{u \sim U_n} \|\overline{\mu_{Q+u}} - \text{bin}_n\|_1 \leq \min \left\{ (2t + 1) \left(e \ln \frac{n}{2t} \right)^t \left(\frac{2t}{n} \right)^{\frac{t}{2}-1}, \sqrt{2}(n + 1)e^{-\frac{t}{10}} \right\}.$$

Thus, for $t = \Theta(1)$,

$$E_{u \sim U_n} \|\overline{\mu_{Q+u}} - \text{bin}_n\|_1 = O \left(\frac{(\ln n)^t}{n^{\frac{t}{2}-1}} \right).$$

It follows that, for $t \geq 3$, almost all cosets of Q have weight distributions close to the binomial distribution in the L_1 -sense. An extended version of dual BCH codes gives explicit codes of size $2(n + 1)^t$ and bilateral minimum distance at least $2t + 1$ [5, 6]. We can interpret Theorem 6.6 as a statement about linear codes with large dual bilateral minimum distance being *globally* pseudobinomial on the average in the L_1 -sense (they cannot be locally pseudobinomial due the defining linear constraints which makes them highly biased).

7 Sum of spaces conjectures

Viola [7, 19, 27] proved that the sum of independent small-bias spaces fools constant degree polynomials. Naturally, this suggests the question of whether other simple functions can be derandomized by sums of independent small-bias spaces. Reingold and Vadhan asked whether the sum of two independent $n^{-O(1)}$ -biased spaces fools log-space [21]. Meka and Zuckerman [21] ruled out the possibility that the sum of two independent spaces with constant-bias gives a hitting set for log-space. Since any distribution which ϵ -fools log-space must be ϵ -pseudobinomial in L_∞ -sense (and hence $(n + 1)\epsilon$ -pseudobinomial in the L_1 -sense), a natural question is whether the sum of two independent δ -biased spaces is $O((\delta n)^{\Theta(1)})$ -pseudobinomial in the L_1 -sense.

If $X, Y \in \{0, 1\}^n$ are independent random vectors distributed according to μ_X and μ_Y , consider the sum $X + Y$ over \mathbb{F}_2 . The probability distribution of $X + Y$ is the convolution $\mu_X * \mu_Y$: $(\mu_X * \mu_Y)(z) = \sum_x \mu_X(x) \mu_Y(x + z)$.

⁵Theorem 6.6 is not used in the proofs in this paper except for the proof of Lemma 7.7, which is stated to compare with highly biased spaces.

Conjecture 7.1 (Pseudobinomiality of sum). *For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, the sum $X + Y$ is $O((n\delta)^{\Theta(1)})$ -pseudobinomial in the L_1 -sense.*

That is, there exist constants $a, b, c, n_0 > 0$, such that for each $\delta > 0$, for each integer $n > n_0$, and all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, the sum $X + Y$ is $a\delta^b n^c$ -pseudobinomial in the L_1 -sense.

We argue that Conjecture 7.1 is equivalent to each of the following three conjectures.

Conjecture 7.2 (Entropy of sum). *For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, the Shannon entropy of the weight of $X + Y$ satisfies $H(|X + Y|) \geq H(\text{bin}_n) - O((n\delta)^{\Theta(1)})$.*

That is, there exist constants $a, b, c, n_0 > 0$, such that for each $\delta > 0$, for each integer $n > n_0$, and all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, $H(|X + Y|) \geq H(\text{bin}_n) - a\delta^b n^c$.

Conjecture 7.3 (Entropy of sum without the binomial: max version). *For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, $H(|X + Y|) \geq \max\{H(|X|), H(|Y|)\} - O((n\delta)^{\Theta(1)})$.*

Conjecture 7.4 (Entropy of sum without the binomial: min version). *For all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, $H(|X + Y|) \geq \min\{H(|X|), H(|Y|)\} - O((n\delta)^{\Theta(1)})$.*

Lemma 7.5. *Conjecture 7.1 is equivalent to Conjecture 7.2.*

Lemma 7.6. *Conjectures 7.2, 7.3, and 7.4 are equivalent.*

The proof of Lemma 7.5 is in Section 7.1. The equivalence follows from the equivalence between pseudobinomiality in the L_1 -sense and the entropy-sense, and the fact that small-bias is preserved by restrictions and is invariant under translations (we also need the fact that small-bias implies local pseudobinomiality on small sets of induces (Lemma 5.1) to handle the n_0 parameters in the statement of Conjecture 7.1).

The proof of Lemma 7.6 is in Section 7.2. In Conjectures 7.3 and 7.4, we get rid of the binomial distribution using Corollary 6.4 which says that average-weight entropy of a small-bias space is a good approximation of the entropy of the binomial distribution. To show that Conjecture 7.2 follows from Conjecture 7.4, we argue using Corollary 6.4 that there is $u \in \{0, 1\}^n$ such that both $H(|X + u|)$ and $H(|Y + u|)$ are close to $H(\text{bin}_n)$, and hence by Conjecture 7.4 applied to $X + u$ and $Y + u$, $H(|X + Y|) = H(|(X + u) + (Y + u)|)$ is close to $H(\text{bin}_n)$. Obviously, Conjecture 7.4 follows from Conjecture 7.3. The fact that Conjecture 7.3 follows from Conjecture 7.2 is based on Corollary 2.7, which asserts that the entropy of the weight distribution of a probability distribution on $\{0, 1\}^n$ with small bias cannot significantly exceed $H(\text{bin}_n)$.

If A and B are real valued random variables (taking values in a finite set for instance), a simple conditioning argument show that $H(A+B) \geq H(A+B|B) = H(A)$ and similarly $H(A+B) \geq H(B)$. Thus,

$$H(A + B) \geq \max\{H(A), H(B)\}.$$

Unfortunately, the picture in our context is more complex. For highly biased X and Y , the entropy $H(|X + Y|)$ can be significantly smaller than $H(|X|)$ and $H(|Y|)$.

Lemma 7.7. *For infinitely many values of n , there exists a coset $S \subset \mathbb{F}_2^n$ of an \mathbb{F}_2 -linear code such that $H(|X|) = H(|Y|) = \Theta(\log n)$ but $H(|X + Y|) = \Theta(1)$, where X and Y are random vectors chosen independently and uniformly from S .*

The proof uses Theorem 6.6 and it is in Section 7.3. Lemma 7.7 shows that we need at least one variable to have small bias. The following Lemma shows that we need small bias in both.

Lemma 7.8. *There exist a $2^{-\Omega(n)}$ -biased random vector $X \in \{0, 1\}^n$ and a deterministic vector $Y \in \{0, 1\}^n$ such that $H(|X + Y|) = H(|X|) - \Omega(1)$.*

Proof. We know from Section 5 that there exists a random vector $Z \in \{0, 1\}^n$ such that Z is δ -biased but $H(|Z|) = H(\text{bin}_n) - \Omega(1)$, where $\delta = 2^{-\Omega(n)}$. By Corollary 6.4, $H_{\text{avg}}(Z) = H(\text{bin}_n) - O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta})$. Thus, there exist $u \in \{0, 1\}^n$ such that $H(|Z + u|) = H(\text{bin}_n) - 2^{-\Omega(n)}$. Let $X = Z + u$ and $Y = u$, hence $X + Y = Z$. \square

Another way get to rid of the $H(\text{bin}_n)$ term in Conjecture 7.2 is via averaging followed by conditioning. Consider replacing $H(\text{bin}_n)$ in Conjecture 7.2 by the average entropy $H_{\text{avg}}(X) = H(|X + U| | U)$ of X . By Corollary 6.4, we have $H(\text{bin}_n) = H_{\text{avg}}(X) + O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta})$. Thus, Conjecture 7.2 is equivalent to showing that

$$H(|X + Y|) = H(|X + U| | U) - O((n\delta)^{\Theta(1)}).$$

Since conditioning reduces entropy

$$H(|X + Y|) = H(|X + Y| | Y) + I(|X + Y|; Y) \geq H(|X + Y| | Y),$$

the conjecture would follow if (we don't have a proof of the other direction) we can show that

$$H(|X + Y| | Y) = H(|X + U| | U) - O((n\delta)^{\Theta(1)}),$$

for all independent δ -biased X and Y . Compared to the argument of Viola [27] for low-degree polynomials, the loss incurred by conditioning is at a high level similar to the loss resulting from Jensen's inequality in Viola's argument (since Jensen's inequality is also behind the fact that conditioning reduces entropy).

By looking at the LP dual, we get the following.

Definition 7.9 (Weight-entropy function). *If $X \in \{0, 1\}^n$ is a random vector, define the weight-entropy function $h_{\mu_X} : \{0, 1\}^n \rightarrow \mathbb{R}$ by*

$$h_{\mu_X}(u) \stackrel{\text{def}}{=} H(|X + u|) = H(\sigma_u \mu_X),$$

where μ_X is the probability distribution of X .

Conjecture 7.10 (Lower sandwiching the weight-entropy function). *For each δ -biased random vector $X \in \{0, 1\}^n$, there exists $g : \{0, 1\}^n \rightarrow \mathbb{R}$ such that:*

- $g \leq h_{\mu_X}$
- $E_{U_n}(h_{\mu_X} - g) = O((n\delta)^{\Theta(1)})$
- $\delta \|\hat{g}\|_1 = O((n\delta)^{\Theta(1)})$.

Lemma 7.11. *Conjecture 7.10 implies Conjecture 7.2.*

The proof of Lemma 7.11 is in Section 7.4.

Remark 7.12. We conclude this section with some observations:

- 1) If Conjecture 7.4 holds with error $\beta(n, \delta)$ (i.e., $H(|X+Y|) \geq \min\{H(|X|), H(|Y|)\} - \beta(n, \delta)$), careful tracking of the bounds in the direct part of the proof of Lemma 7.5 shows that: for all independent δ -biased random vectors $X, Y \in \{0, 1\}^n$, $X + Y$ is $(\xi + \beta(n, \delta))$ -pseudobinomial in the L_1 -sense, where $\xi = O\left(\sqrt{n\delta} + n^{1/4}\sqrt{\delta \lg(1/\delta)}\right)$.
- 2) The correctness of the above conjectures leads to a PRG which, in addition to ϵ -pseudobinomiality, has the property of retaining ϵ -pseudobinomiality under the action of invertible linear transformations. Namely, for any invertible \mathbb{F}_2 -linear transformation $T : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$, μ^T is ϵ -pseudobinomial, where $\mu^T(x) = \mu(Tx)$.
- 3) Consider the more general scenario of $k \geq 2$ independent δ -biased random variables $X_1, \dots, X_k \in \{0, 1\}^n$. For $\delta = n^{-\Theta(1)}$, the complexity of the underlying PRG is $2^{O(k \log n)}$. The proofs of Lemmas 7.1 and 7.2 can be easily adapted to this setup without affecting the asymptotic error term $O((\delta n)^{\Theta(1)})$. Hence the equivalence of Conjectures 7.1, 7.2, 7.3, and 7.4 holds. In principle, the conjectures become weaker as k increases.
- 4) A more general framework for the above questions is the following. Let G be finite abelian Cayley graph of diameter n . Define weight function $W : G \rightarrow \mathbb{R}$, where $W(g)$ is the distance of g from zero. A probability distribution on G is δ -biased if $|E_\mu \chi| \leq \delta$ for each character χ of G [2]. Let X and Y be independent random variables taking values in G . Consider the sum $X+Y$ over G . The more general problem is about studying the Shannon entropy of $W(X+Y)$ compared to that of $W(U)$, where U be a uniformly random element of G . Note that for the hypercube, the diameter $n = \lg |G|$, i.e., it is very small compared to $|G|$; the setup of Cayley graphs with large diameters such as the circle graph C_n ⁶ does not capture the problem.

7.1 Proof of Lemma 7.5

Assume Conjecture 7.2 (entropy of sum) holds for $a, b, c, n_0 > 0$. Let $X, Y \in \{0, 1\}^n$ be independent δ -biased random vectors, where $n > n_0$. Let μ_X be the probability distribution of X and μ_Y that of Y . First, note that since X and Y are each δ -biased, $X + Y$ is δ^2 -biased (because $E_{X,Y} \chi_z(X + Y) = E_X \chi_z(X) E_Y \chi_z(Y)$, for each $z \in \{0, 1\}^n$). Let $I \subset [n]$ be nonempty and $u \in \{0, 1\}^I$. If $|I| \leq n_0$, by the local pseudobinomiality (Lemma 5.1) $\|\sigma_u(\mu_X * \mu_Y)^I - \text{bin}_{|I|}\|_1 \leq 2\delta^2 2^{n_0/2}$. Assume in what follows that $|I| > n_0$. Since the δ -bias property is invariant under translations and is preserved by restrictions, $X|_I$ and $Y|_I + u$ are δ -biased. By Conjecture 7.2, $H(|X + Y + u|) \geq H(\text{bin}_{|I|}) - a\delta^b |I|^c \geq H(\text{bin}_{|I|}) - a\delta^b n^c$. It follows from Corollary 2.6, that $\|\sigma_u(\mu_X * \mu_Y)^I - \text{bin}_{|I|}\|_1 \leq \sqrt{(2 \ln 2)(n\delta^2 + a\delta^b n^c)}$. Therefore, $X + Y$ is $\max\{\sqrt{(2 \ln 2)(n\delta^2 + a\delta^b n^c)}, 2\delta^2 2^{n_0/2}\}$ -pseudobinomial in the L_1 -sense, for all $n > n_0$.

Assume Conjecture 7.1 (pseudobinomiality of sum) holds for $a, b, c, n_0 > 0$, and let $X, Y \in \{0, 1\}^n$ be independent and δ -biased, where $n > n_0$. Thus, $X + Y$ is $a\delta^b n^c$ -pseudobinomial in the L_1 -sense. In particular, $\|\overline{\mu_X * \mu_Y} - \text{bin}_n\|_1 \leq a\delta^b n^c$. It follows from the entropy-difference bound (Part (b) of Lemma 1.1) that if $a\delta^b n^c \leq 1/2$, then

$$H(\text{bin}_n) - H(|X + Y|) \leq a\delta^b n^c \lg \frac{n+1}{a\delta^b n^c} \leq a\delta^b n^c \sqrt{\frac{n+1}{a\delta^b n^c}} = 2\sqrt{a}\delta^{b/2} n^{c/2} \sqrt{n+1},$$

⁶It is not hard to see that for C_n , the distributions of $W(X)$ and $W(U)$ are δ -close in the L_∞ -sense for any δ -biased X . That is, only random one variable is need.

since $\lg x \leq 2\sqrt{x}$, for all $x > 0$. If $a\delta^b n^c > 1/2$ (i.e., the bound of Conjecture 7.1 is not good for the given values of δ and n), then trivially $H(\text{bin}_n) - H(|X + Y|) \leq H(\text{bin}_n) \leq \lg(n + 1) \leq 2a\delta^b n^c \lg(n + 1)$. It follows that in all cases,

$$H(\text{bin}_n) - H(|X + Y|) \leq \max\{2\sqrt{a}\delta^{b/2}n^{c/2}\sqrt{n+1}, 2a\delta^b n^c \lg(n+1)\},$$

for all $n > n_0$.

7.2 Proof of Lemma 7.6

Clearly, Conjecture 7.4 (min version) follows from Conjecture 7.3 (max version). To show that Conjecture 7.2 follows from Conjecture 7.4, let $X, Y \in \{0, 1\}^n$ be independent δ -biased random vectors. We will argue that $H(|X + Y|) = H(\text{bin}_n) - O((n\delta)^{\Theta(1)})$ assuming Conjecture 7.4. Since X is δ -biased, by Corollary 6.4, $H(\text{bin}_n) - E_{u \sim U_n} H(|X + u|) = \epsilon$, where $\epsilon = O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta})$. Thus, the fraction of elements $u \in \{0, 1\}^n$ such that $H(\text{bin}_n) - H(|X + u|) > 3\epsilon$ is at most $\frac{1}{3}$. Similarly, since Y is δ -biased, the fraction of elements $u \in \{0, 1\}^n$ such that $H(\text{bin}_n) - H(|Y + u|) > 3\epsilon$ is at most $\frac{1}{3}$. Hence, there exists $u \in \{0, 1\}^n$ such that $H(\text{bin}_n) - H(|X + u|) \leq 3\epsilon$ and $H(\text{bin}_n) - H(|Y + u|) \leq 3\epsilon$. Fix such a u , let $X' = X + u$, and let $Y' = Y + u$. We have $H(|X'|) \geq H(\text{bin}_n) - 3\epsilon$ and $H(|Y'|) \geq H(\text{bin}_n) - 3\epsilon$. It follows from Conjecture 7.4 that $H(|X' + Y'|) \geq H(\text{bin}_n) - 3\epsilon - O((n\delta)^{\Theta(1)})$. Since $X' + Y' = X + u + Y + u = X + Y$, we get $H(|X + Y|) \geq H(\text{bin}_n) - 3\epsilon - O((n\delta)^{\Theta(1)}) = H(\text{bin}_n) - O((n\delta)^{\Theta(1)})$.

To derive Conjecture 7.3 from Conjecture 7.2, we use Corollary 2.7. By Corollary 2.7, $H(|X|) \leq H(\text{bin}_n) + n\delta$ and $H(|Y|) \leq H(\text{bin}_n) + n\delta$, i.e., $H(\text{bin}_n) \geq \max\{H(|X|), H(|Y|)\} - n\delta$. If Conjecture 7.2 holds, then

$$H(|X + Y|) \geq H(\text{bin}_n) - O((n\delta)^{\Theta(1)}) \geq \max\{H(|X|), H(|Y|)\} - n\delta - O((n\delta)^{\Theta(1)}).$$

7.3 Proof of Lemma 7.7

The proof uses Theorem 6.6. Let $n = 2^r - 1$, where $r \geq 2$ is an integer, and let D be the $(2^r - 1, 2^r - 1 - r, 3)$ -Hamming code, thus D^\perp is the $(2^r - 1, r, 2^{r-1})$ -Hadamard code. Let $Q = D^\perp \cup (D^\perp + \vec{1})$ be the extended Hadamard code of size $2^{r+1} = 2(n + 1)$, where $\vec{1}$ is the all ones vector. The dual Q^\perp of Q is the set of even weight codewords of D . Since $\vec{1} \in D$, each codeword in D other than 0 and $\vec{1}$ has weight between 3 and $n - 3$. Since n is odd, $\vec{1} \notin Q^\perp$. Thus Q^\perp has bilateral minimum distance at least 3 (actually, at least 4). Each nonzero codeword of the Hadamard code D^\perp has weight $\frac{n+1}{2}$, hence the possible weights of the codewords of Q are 0, $\frac{n-1}{2}$, $\frac{n+1}{2}$, and n . Thus, $H(\overline{\mu_Q}) = \Theta(1)$ (actually, $H(\overline{\mu_Q}) = 1 + o(1)$). It follows from Theorem 6.6 that $E_{u \sim U_n} \|\overline{\mu_{Q+u}} - \text{bin}_n\|_\infty = O\left(\frac{\ln n}{\sqrt{n}}\right)$. Thus, there exists $u \in \{0, 1\}^n$ such that $\|\overline{\mu_{Q+u}} - \text{bin}_n\|_\infty = O\left(\frac{\ln n}{\sqrt{n}}\right)$. Since $\|\text{bin}_n\|_\infty = O\left(\frac{1}{\sqrt{n}}\right)$, we get $\|\overline{\mu_{Q+u}}\|_\infty = O\left(\frac{\ln n}{\sqrt{n}}\right)$. It follows that $H(\overline{\mu_{Q+u}}) \geq \lg \frac{1}{\|\overline{\mu_{Q+u}}\|_\infty} = \Omega(\log n)$. To sum up, $H(\overline{\mu_{Q+u}}) = \Theta(\log n)$ but $H(\overline{\mu_{Q+u}} * \overline{\mu_{Q+u}}) = \Theta(1)$ since $\mu_{Q+u} * \mu_{Q+u} = \mu_Q$ because $(Q + u) + (Q + u) = Q$.

7.4 Proof of Lemma 7.11

Let $X, Y \in \{0, 1\}^n$ be independent δ -biased random vectors. We have $E_{U_n} h_{\mu_X} = H(|X + U| | U) = H_{\text{avg}}(X)$ and $E_{\mu_Y} h_{\mu_X} = H(|X + Y| | Y)$. Let $g : \{0, 1\}^n \rightarrow \mathbb{R}$ such that $g \leq h_{\mu_X}$, hence

$$E_{\mu_Y} h_{\mu_X} \geq E_{\mu_Y} g = E_{U_n} h_{\mu_X} - E_{U_n} (h_{\mu_X} - g) - (E_{U_n} g - E_{\mu_Y} g).$$

Since conditioning reduces entropy and $|E_{U_n}g - E_{\mu_Y}g| \leq \delta\|\widehat{g}\|_1$, we get

$$\begin{aligned} H(|X + Y|) &\geq H(|X + Y| | Y) \\ &\geq H_{avg}(X) - E_{U_n}(h_{\mu_X} - g) - \delta\|\widehat{g}\|_1 \\ &= H(\text{bin}_n) - O(n\delta + \sqrt{n}\delta \lg \frac{1}{\delta}) - E_{U_n}(h_{\mu_X} - g) - \delta\|\widehat{g}\|_1 \quad (\text{by Corollary 6.4}). \end{aligned}$$

8 Functions fooled by pseudobinomiality

Although a more critical problem at this stage is that of constructing small pseudobinomial spaces, it is worth mentioning the LP duality characterization of the space of functions fooled by pseudobinomiality. The class of functions fooled by the k -wise independence properly are characterized by tight sandwichability between low degree polynomials [3, 4]. For δ -biased distributions, we get small L_1 -norm in the Fourier domain instead of low degree polynomials [3, 4].

For pseudobinomiality, we get sums of translations of symmetric functions $\{f_i\}_i$ on subsets of the variables such that the total L_∞ -norm $\sum_i \|f_i\|_\infty$ is small.

Lemma 8.1 (Sandwiching). *Let $f : \{0, 1\}^n \rightarrow \mathbb{R}$ and $\epsilon, \alpha > 0$. Then the following are equivalent:*

- I) $|E_\mu f - E_{U_n} f| \leq \alpha$, for each ϵ -pseudobinomial probability distribution μ on $\{0, 1\}^n$
- II) There exist functions $f^l : \{0, 1\}^n \rightarrow \mathbb{R}$ (lower sandwiching function) and $f^h : \{0, 1\}^n \rightarrow \mathbb{R}$ (upper sandwiching function) which can be expressed as

$$f^l(x) = c^l + \sum_i f_i^l(x_{I_i^l} + u_i^l) \quad \text{and} \quad f^h(x) = c^h + \sum_i f_i^h(x_{I_i^h} + u_i^h),$$

where $c^l, c^h \in \mathbb{R}$, and for each i , $I_i^l, I_i^h \neq \emptyset \subset [n]$, $u_i^l \in \{0, 1\}^{I_i^l}$, $u_i^h \in \{0, 1\}^{I_i^h}$, $f_i^l : \{0, 1\}^{I_i^l} \rightarrow \mathbb{R}$ is a symmetric function (i.e., $f_i^l(y)$ depends only on the weight $|y|$ of y), and $f_i^h : \{0, 1\}^{I_i^h} \rightarrow \mathbb{R}$ is a symmetric function such that:

- a) (sandwiching) $f^l \leq f \leq f^h$
- b) $E_{U_n}(f - f^l) + \epsilon \sum_i \|f_i^l\|_\infty \leq \alpha$
- c) $E_{U_n}(f^h - f) + \epsilon \sum_i \|f_i^h\|_\infty \leq \alpha$.

Proof of the direct part. The key point is that if $g : [0 : n] \rightarrow \mathbb{R}$ and μ is a probability distribution on $\{0, 1\}^n$ such that $\|\overline{\mu} - \text{bin}_n\|_1 \leq \epsilon$, then $|E_\mu g - E_{\text{bin}_n} g| \leq \epsilon\|g\|_\infty$. Hence, following the reasoning in [3, 4], we get that if $f \leq f^h$, then

$$E_\mu f - E_{U_n} f \leq E_{U_n}(f^h - f) + (E_\mu f^h - E_{U_n} f^h) \leq E_{U_n}(f^h - f) + \epsilon \sum_i \|f_i^h\|_\infty.$$

Similarly, if $f^l \leq f$, we get $E_{U_n} f - E_\mu f \leq E_{U_n}(f - f^l) + \epsilon \sum_i \|f_i^l\|_\infty$. Thus, (II) implies (I).

The other direction follows from LP duality. Note that in addition to $\mu \geq 0$, $\sum_x \mu(x) = 1$, we have the following primal L_1 -constraints on μ : $\|\overline{\sigma_u \mu^I} - \text{bin}_{|I|}\|_1 \leq \epsilon$, for all nonempty $I \subset [n]$, and all $u \in \{0, 1\}^I$. Each of those L_1 -constraints can be represented in terms of $2^{|I|+1}$ linear constraints: $|E_\mu \beta_{I,u,a} - E_{U_n} \beta_{I,u,a}| \leq \epsilon$, for each $a : [0 : |I|] \rightarrow \{-1, 1\}$, where $\beta_{I,u,a} : \{0, 1\}^n \rightarrow \{-1, 1\}$ is given by $\beta_{I,u,a}(x) = a(|x_I + u|)$. \square

Note that, asymptotically, (II) is equivalent to $E_{U_n}(f^h - f^l) = O(\alpha)$, $N^l = O(\frac{\alpha}{\epsilon})$, and $N^h = O(\frac{\alpha}{\epsilon})$, where $N^l = \sum_i \|f_i^l\|_\infty$ and $N^h = \sum_i \|f_i^h\|_\infty$. Thus, the pseudobinomiality property fools f iff N^l and N^h are small (e.g., polynomial in n) and $E_{U_n}(f^h - f^l)$ is small (e.g., $O(n^{-\Theta(1)})$). Compared to the small-bias case, we have translations of symmetric functions on subsets of the variables instead of the characters $\{\chi_z\}_z$.

A natural resulting question is to study the boolean functions which can be approximated in the above sense. The simplest related question is probably: which functions $\{0, 1\}^n \rightarrow \{0, 1\}$ can be expressed as $\sum_i f_i(|x_{I_i} + u_i|)$ with $\{(I_i, u_i, f_i)\}_i$ as above, and $\sum_i \|f_i\|_\infty$ polynomial in n ?

9 Proof of the negative spectrum lemma

The proof is based on analyzing the binomial coefficients. A natural question is whether there is a less technical proof. For convenience, we repeat the statement of the lemma here.

Lemma 2.3 *Let $L : \{0, 1\}^n \rightarrow \mathbb{R}$ be the log-binomial function given by:*

$$L(x) = \lg \binom{n}{|x|}.$$

Then $\widehat{L}(z) \leq 0$ for each $z \in \{0, 1\}^n$. Moreover, $\widehat{L}(z) = 0$ if $|z|$ odd, and $\widehat{L}(z) < 0$ if $z \neq 0$ and $|z|$ even.

If $|z|$ is odd, then $\chi_z(x) = -\chi_z(x + \vec{1})$ and $\binom{n}{|x|} = \binom{n}{n - |x|} = \binom{n}{|x + \vec{1}|}$, hence $\chi_z(x) \binom{n}{|x|} + \chi_z(x + \vec{1}) \binom{n}{|x + \vec{1}|} = 0$ for each $x \in \{0, 1\}^n$. It follows that

$$\widehat{L}(z) = \frac{1}{2^n} \sum_x \chi_z(x) \lg \binom{n}{|x|} = 0.$$

If $z = 0$, then $\widehat{L}(0) = E_{U_n} L \leq 0$ since $L \leq 0$.

In what follows assume that $|z|$ even and $z \neq 0$. Let $S = \text{support}(z)$ and $m = |S|$, hence m is even and $m \geq 2$. We have

$$\begin{aligned} \widehat{L}(z) &= \frac{1}{2^n} \sum_x \chi_z(x) \lg \binom{n}{|x|} \\ &= \frac{1}{2^n} \sum_{x'' \in \{0, 1\}^{S^c}} \sum_{x' \in \{0, 1\}^S} (-1)^{|x'|} \lg \frac{\binom{|S| + |S^c|}{|x'| + |x''|}}{2^{|S| + |S^c|}} \\ &= \frac{1}{2^n} \sum_{x'' \in \{0, 1\}^{S^c}} \sum_{x' \in \{0, 1\}^S} (-1)^{|x'|} \lg \binom{|S| + |S^c|}{|x'| + |x''|} \quad (\text{since } \sum_{x' \in \{0, 1\}^S} (-1)^{|x'|} = 0) \\ &= \frac{1}{2^n} \sum_{x'' \in \{0, 1\}^{S^c}} \sum_{w=0}^{|S|} (-1)^w \binom{|S|}{w} \lg \binom{|S| + |S^c|}{w + |x''|} \\ &= \frac{1}{2^n} \sum_{x'' \in \{0, 1\}^{S^c}} \beta_m(|x''|, |S^c| - |x''|), \end{aligned}$$

where $\beta_m(a, b)$ is defined below.

Definition 9.1. If $m, a, b \geq 0$ are integers, define

$$\beta_m(a, b) \stackrel{\text{def}}{=} \sum_{w=0}^m (-1)^w \binom{m}{w} \lg \binom{m+a+b}{w+a}.$$

Lemma 2.3 then follows from the following lemma.

Lemma 9.2. If $m \geq 2$ is even and $a, b \geq 0$, then $\beta_m(a, b) < 0$.

This in turn follows from the following lemma.

Lemma 9.3. Let $m \geq 2$ be even, then

- a) $\beta_m(a, b) < \beta_m(a + a', b + b')$ for each $a, b, a', b' \geq 0$ such that not both a' and b' are zero
- b) $\lim_{c \rightarrow \infty} \beta_m(c, c) = 0$.

To derive Part (b) of Lemma 9.3, we need the following.

Theorem 9.4. (de Moivre-Laplace normal approximation of the binomial [11], page 184)
If $w \in [0 : n]$ is a function of n such that $|w - n/2| = o(n^{2/3})$, then

$$\text{bin}_n(w) = \sqrt{\frac{2}{\pi n}} e^{-2 \frac{(w-n/2)^2}{n}} (1 \pm o(1)).$$

Proof of Part (b) of Lemma 9.3. Let $c \geq 0$. We have

$$\begin{aligned} \beta_m(c, c) &= \sum_{w=0}^m (-1)^w \binom{m}{w} \lg \binom{m+2c}{w+c} \\ &= \sum_{w=0}^m (-1)^w \binom{m}{w} \lg \frac{\binom{m+2c}{w+c} \sqrt{m+2c}}{2^{m+2c}} \quad (\text{since } \sum_w (-1)^w \binom{m}{w} = 0). \end{aligned}$$

Thus,

$$\lim_{c \rightarrow \infty} \beta_m(c, c) = \sum_{w=0}^m (-1)^w \binom{m}{w} \lg \lim_{c \rightarrow \infty} A_c,$$

where

$$A_c = \frac{\binom{m+2c}{w+c} \sqrt{m+2c}}{2^{m+2c}} = \text{bin}_{m+2c}(w+c) \sqrt{m+2c}.$$

By de Moivre-Laplace normal approximation of the binomial, $\lim_{c \rightarrow \infty} A_c = \sqrt{\frac{2}{\pi}}$, hence

$$\lim_{c \rightarrow \infty} \beta_m(c, c) = \left(\lg \sqrt{\frac{2}{\pi}} \right) \sum_{w=0}^m (-1)^w \binom{m}{w} = 0.$$

□

To derive Part (a) of Lemma 9.3, expand

$$\begin{aligned} \lg \binom{m+a+b}{w+a} &= \lg(m+a+b)! - \lg(w+a)! - \lg(m-w+b)! \\ &= \lg(m+a+b)! - \lg w! - \lg(m-w)! - \sum_{i=1}^a \lg(w+i) - \sum_{i=1}^b \lg(m-w+i). \end{aligned}$$

Thus,

$$\begin{aligned} &\lg \binom{m+a+b}{w+a} - \lg \binom{m+a+a'+b+b'}{w+a+a'} \\ &= \lg \frac{(m+a+b)!}{(m+a+a'+b+b')!} + \sum_{i=a+1}^{a+a'} \lg(w+i) + \sum_{i=b+1}^{b+b'} \lg(m-w+i). \end{aligned}$$

It follows that

$$\beta_m(a, b) - \beta_m(a+a', b+b') = \sum_{i=a+1}^{a+a'} \gamma_m(i) + \sum_{i=b+1}^{b+b'} \gamma_m(i), \quad (15)$$

where $\gamma_m(i)$ is defined below.

Definition 9.5. *If $m, i \geq 0$ are integers, let*

$$\gamma_m(i) \stackrel{\text{def}}{=} \sum_w (-1)^w \binom{m}{w} \lg(w+i).$$

Note that we used in the derivation of (15) the fact that $\sum_w (-1)^w \binom{m}{w} = 0$, and the fact that

$$\sum_w (-1)^w \binom{m}{w} \lg(m-w+i) = \sum_w (-1)^w \binom{m}{w} \lg(w+i),$$

which holds because m is even.

Part (a) of Lemma 9.3 follows from (15) and the following lemma.

Lemma 9.6. *If $m \geq 2$ is even and $i \geq 0$, then $\gamma_m(i) < 0$.*

The proof of Lemma 9.6 is below and it uses the following lemmas.

Lemma 9.7. (Taylor series of \lg) *For all $q \geq 0$ and $0 \leq z \leq 1$, we have*

$$\lg(z+q) = \lg\left(q + \frac{1}{2}\right) + \sum_{k=1}^{\infty} \alpha_k(q) \left(z - \frac{1}{2}\right)^k,$$

where $\alpha_k(q) < 0$ for each even $k \geq 2$

Lemma 9.8. *If $m \geq 2$ and $k \geq 0$ are integers such that m is even, then*

$$\sum_w (-1)^w \binom{m}{w} (w - m/2)^k \begin{cases} = 0 & \text{if } k \text{ is odd} \\ = 0 & \text{if } k \text{ is even and } k < m \\ > 0 & \text{if } k \text{ is even and } k \geq m. \end{cases}$$

9.1 Proof of Lemma 9.6

We have

$$\begin{aligned}
\gamma_m(i) &= \sum_{w=0}^m (-1)^w \binom{m}{w} \lg(w+i) \\
&= \sum_{w=0}^m (-1)^w \binom{m}{w} \lg\left(\frac{w}{m} + \frac{i}{m}\right) \quad (\text{since } \sum_w (-1)^w \binom{m}{w} = 0) \\
&= \sum_{w=0}^m (-1)^w \binom{m}{w} \left(\lg\left(\frac{i}{m} + \frac{1}{2}\right) + \sum_{k=1}^{\infty} \alpha_k \left(\frac{i}{m}\right) \left(\frac{w}{m} - \frac{1}{2}\right)^k \right) \quad (\text{by Lemma 9.7}) \\
&= \sum_{k=1}^{\infty} \alpha_k \left(\frac{i}{m}\right) \frac{1}{m^k} \sum_{w=0}^m (-1)^w \binom{m}{w} \left(w - \frac{m}{2}\right)^k \quad (\text{since } \sum_w (-1)^w \binom{m}{w} = 0) \\
&= \sum_{\substack{k \geq m \\ \text{even}}} \alpha_k \left(\frac{i}{m}\right) \frac{1}{m^k} \sum_{w=0}^m (-1)^w \binom{m}{w} \left(w - \frac{m}{2}\right)^k \quad (\text{by Lemma 9.8}) \\
&< 0 \quad (\text{by Lemmas 9.7 and 9.8}).
\end{aligned}$$

Note that Lemma 9.7 is applicable since $k \geq m \geq 2$.

9.2 Proof of Lemma 9.7

Using the Taylor series of the natural logarithm

$$\ln(x+1) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k \quad \text{for } |x| \leq 1,$$

we get

$$\begin{aligned}
\lg(z+q) &= \lg\left(q + \frac{1}{2}\right) + \lg\left(\frac{z - \frac{1}{2}}{q + \frac{1}{2}} + 1\right) \\
&= \lg\left(q + \frac{1}{2}\right) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k \left(q + \frac{1}{2}\right)^k \ln 2} \left(z - \frac{1}{2}\right)^k,
\end{aligned}$$

if $\left|\frac{z - \frac{1}{2}}{q + \frac{1}{2}}\right| \leq 1$, i.e., if $-q \leq z \leq 1 + q$. Thus, the expansion holds for all $q > -\frac{1}{2}$ and $-q \leq z \leq 1 + q$, and in particular for $q \geq 0$ and $0 \leq z \leq 1$.

9.3 Proof of Lemma 9.8

The proof uses a standard argument based on the method of generating functions (e.g., [11], Chapter 11). If $k = 0$, the summation is clearly zero. Assume in what follows that $k \geq 1$. Let $f_k : \{0, 1\}^m \rightarrow \mathbb{R}$ be given by $f_k(x) = (-1)^{|x|} (|x| - m/2)^k$. Thus, with respect to the uniform distribution U_m on $\{0, 1\}^m$, we have

$$E_{U_m} f_k = \frac{1}{2^m} \sum_w (-1)^w \binom{m}{w} (w - m/2)^k.$$

We study below the sign of $E_{U_m} f_k$. Let s be a real variable and consider the moment generating function $g_s : \{0, 1\}^m \rightarrow \mathbb{R}$: $g_s(x) = (-1)^{|x|} e^{s(|x|-m/2)}$. Thus, $f_k(x) = \frac{\partial^k}{\partial s^k} g_s(x)|_{s=0}$, and hence $E_{U_m} f_k = \frac{\partial^k}{\partial s^k} E_{U_m} g_s|_{s=0}$. We have

$$\begin{aligned} E_{U_m} g_s &= \sum_w \frac{\binom{m}{w}}{2^m} (-1)^w e^{s(w-m/2)} = \frac{1}{2^m} \sum_w \binom{m}{w} (-e^{s/2})^w (e^{-s/2})^{m-w} \\ &= \left(-\sinh\left(\frac{s}{2}\right) \right)^m = \left(\sinh\left(\frac{s}{2}\right) \right)^m, \end{aligned}$$

since m is even. Therefore, $E_{U_m} f_k = \frac{\partial^k}{\partial s^k} \left(\sinh\left(\frac{s}{2}\right) \right)^m |_{s=0}$, for all $k \geq 1$. We have for all $z \in \mathbb{R}$,

$$\sinh(z) = \sum_{t \text{ odd} \geq 1} \frac{z^t}{t!}.$$

Thus,

$$\left(\sinh\left(\frac{s}{2}\right) \right)^m = \sum_{t \geq m \text{ even}} \binom{m}{t} \sum_{t_1, \dots, t_m \text{ odd} \geq 1: \sum_i t_i = t} \frac{1}{t_1! \dots t_m!}.$$

Note that t must be even since t_1, \dots, t_m are odd and m is even. Thus, $\frac{\partial^k}{\partial s^k} \left(\sinh\left(\frac{s}{2}\right) \right)^m |_{s=0} = 0$ if $k < m$, or if k is odd and $k \geq m$. If $k \geq m$ and k is even, we have

$$\frac{\partial^k}{\partial s^k} \left(\sinh\left(\frac{s}{2}\right) \right)^m |_{s=0} = \frac{k!}{2^k} \sum_{t_1, \dots, t_m \text{ odd} \geq 1: \sum_i t_i = k} \frac{1}{t_1! \dots t_m!} > 0.$$

Acknowledgments

The author would like to thank Ibrahim Abou-Faycal for helpful discussions and the anonymous reviewers for their comments.

References

- [1] NOGA ALON, ODED GOLDREICH, JOHAN HØASTAD, AND RENÉ PERALTA: Simple constructions of almost k -wise independent random variables. *Random Structures and Algorithms*, 3:289–304, 1992.
- [2] YOSSI AZAR, RAJEEV MOTWANI, AND JOSEPH NAOR: Approximating probability distributions using small sample spaces. *Combinatorica*, 18(2):151–171, 1998.
- [3] LOUAY BAZZI: *Minimum distance of error correcting codes versus encoding complexity, symmetry, and pseudorandomness*. Ph.D. thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 2003.
- [4] LOUAY BAZZI: Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput*, 38(6):2220–2272, 2009.
- [5] LOUAY BAZZI: Weight distribution of cosets of small codes with good dual properties. *CoRR*, abs/1408.5681, 2014.

- [6] LOUAY BAZZI: Weight distribution of cosets of small codes with good dual properties. In *Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT 2015), Hong Kong*, June 2015.
- [7] ANDREJ BOGDANOV AND EMANUELE VIOLA: Pseudorandom bits for polynomials. In *FOCS: IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 41–51, 2007.
- [8] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory*. Wiley, 1991.
- [9] ANINDYA DE: Beyond the central limit theorem: asymptotic expansions and pseudorandomness for combinatorial sums. *Electronic Colloquium on Computational Complexity (ECCC)*, TR14-125, 2014.
- [10] ANINDYA DE, OMID ETESAMI, LUCA TREVISAN, AND MADHUR TULSIANI: Improved pseudorandom generators for depth 2 circuits. In *APPROX-RANDOM*, pp. 504–517, 2010.
- [11] WILLIAM FELLER: *An introduction to probability theory and its applications. Volume I*. John Wiley & Sons, 1968.
- [12] PARIKSHIT GOPALAN, DANIEL M. KANE, AND RAGHU MEKA: Pseudorandomness for concentration bounds and signed majorities. *CoRR*, abs/1411.4584, 2014.
- [13] PARIKSHIT GOPALAN, RAGHU MEKA, OMER REINGOLD, AND DAVID ZUCKERMAN: Pseudorandom generators for combinatorial shapes. *SIAM J. Comput.*, 42(3):1051–1076, 2013.
- [14] RUSSELL IMPAGLIAZZO AND AVI WIGDERSON: $P = BPP$ if E requires exponential circuits: Derandomizing the XOR lemma. In *STOC: ACM Symposium on Theory of Computing (STOC)*, pp. 220–229, 1997.
- [15] PHILIPPE JACQUET AND WOJCIECH SZPANKOWSKI: Entropy computations via analytic de-poissonization. *IEEE Transactions on Information Theory*, 45(4):1072–1081, 1999.
- [16] JEFF KAHN, GIL KALAI, AND NATHAN LINIAL: The influence of variables on boolean functions. In *FOCS: IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 68–80, 1988.
- [17] ROBERT J. LECHNER: Harmonic analysis of switching functions. In *Recent Developments in Switching Theory*, ed. by Amar Mukhopadhyay. Academic Press, New York, 1971.
- [18] NATHAN LINIAL, YISHAY MANSOUR, AND NOAM NISAN: Constant depth circuits, fourier transform, and learnability. *JACM: Journal of the ACM*, 40(3):607–620, 1993.
- [19] SHACHAR LOVETT: Unconditional pseudorandom generators for low degree polynomials. *Theory of Computing*, 5(1):69–82, 2009.
- [20] SHACHAR LOVETT, OMER REINGOLD, LUCA TREVISAN, AND SALIL P. VADHAN: Pseudorandom bit generators that fool modular sums. In *APPROX-RANDOM*, pp. 615–630, 2009.
- [21] RAGHU MEKA AND DAVID ZUCKERMAN: Small-bias spaces for group products. In *APPROX-RANDOM*, pp. 658–672, 2009.

- [22] JOSEPH NAOR AND MONI NAOR: Small-bias probability spaces: Efficient constructions and applications. *SICOMP: SIAM Journal on Computing*, 22(4):838–856, 1993.
- [23] NOAM NISAN: Pseudorandom generators for space-bounded computation. *COMBINATORICA*, 12(4):449–461, 1992.
- [24] NOAM NISAN AND AVI WIGDERSON: Hardness vs. randomness. In *FOCS: IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 2–11, 1988.
- [25] YUVAL RABANI AND AMIR SHPILKA: Explicit construction of a small epsilon-net for linear threshold functions. *SIAM J. Comput.*, 39(8):3501–3520, 2010.
- [26] UMESH V. VAZIRANI: *Randomness, adversaries and computation*. Ph.D. thesis, University of California, Berkeley, Dec., 1986.
- [27] EMANUELE VIOLA: The sum of D small-bias generators fools polynomials of degree D. *Computational Complexity*, 18(2):209–217, 2009.