# Pseudorandomness, symmetry, smoothing: II

Harm Derksen*
Northeastern University
ha.derksen@northeastern.edu

Peter Ivanov†
Northeastern University
ivanov.p@northeastern.edu

Chin Ho Lee‡
North Carolina State University
chinho.lee@ncsu.edu

Emanuele Viola§
Northeastern University
viola@ccs.neu.edu

July 16, 2024

**Abstract**

We prove several new results on the Hamming weight of bounded uniform and small-bias distributions.

We exhibit bounded-uniform distributions whose weight is anti-concentrated, matching existing concentration inequalities. This construction relies on a recent result in approximation theory due to Erdéyi (Acta Arithmetica 2016). In particular, we match the classical tail bounds, generalizing a result by Bun and Steinke (RANDOM 2015). Also, we improve on a construction by Benjamini, Gurel-Gurevich, and Peled (2012).

We give a generic transformation that converts any bounded uniform distribution to a small-bias distribution that almost preserves its weight distribution. Applying this transformation in conjunction with the above results and others, we construct small-bias distributions with various weight restrictions. In particular, we match the concentration that follows from that of bounded uniformity and the generic closeness of small-bias and bounded-uniform distributions, answering a question by Bun and Steinke (RANDOM 2015). Moreover, these distributions are supported on only a constant number of Hamming weights.

We further extend the anti-concentration constructions to small-bias distributions perturbed with noise, a class that has received much attention recently in derandomization. Our results imply (but are not implied by) a recent result of the authors (CCC 2024), and are based on different techniques. In particular, we prove that the standard Gaussian distribution is far from any mixture of Gaussians with bounded variance.

*Partially supported by NSF grant DMS 2147769.

†Supported by NSF grant CCF-2114116.

‡Work done in part at Harvard University, supported by Madhu Sudan's and Salil Vadhan's Simons Investigator Awards.

§Supported by NSF grant CCF-2114116.

# 1 Introduction and our results

A distribution $D$ over $\{-1,1\}^n$ is $(\varepsilon,k)$-*biased* if for every $S \subseteq [n]$ of size $0 < |S| \leq k$, we have $|\mathbb{E}[D^S]| \leq \varepsilon$, where $D^S := \prod_{i\in S} D_i$. If $\varepsilon = 0$ then any $k$ bits are uniform and $D$ is called $k$-*wise uniform* or simply $k$-*uniform*. If $k = n$ then $D$ is called $\varepsilon$-*biased*.

The study of these distributions permeates and precedes theoretical computer science. They were studied already in the 40's under the name of orthogonal arrays [RR47], and are closely related to universal hash functions [CW79], error-correcting codes (see e.g. [HH23]), and in their modern guise were introduced in the works [ABI86, CGH+85, NN90].

Exploited in countless works, one of the most useful properties of such distributions $D$ is that the distribution of their Hamming weight

$$1^\top D := \sum_{i=1}^{n} D_i$$

is approximately the (centered) binomial distribution

$$B := 1^\top U,$$

where $U$ is uniform in $\{-1,1\}^n$. (For simplicity, in this work we focus on the simplest setting where the distributions are supported on $\{-1,1\}^n$.)

Yet, perhaps surprisingly, available bounds were loose or only applied to specific settings of parameters. In this work, we prove several new results on the Hamming weight of bounded-uniform distributions, small-bias distributions, and an extension of the latter obtained by perturbing it with noise. We discuss each of these in turn.

## 1.1 Bounded uniformity

One of the key properties of bounded uniform distributions is their tail bounds. They have myriad applications in a wide range of areas in computer science, including hashing, load balancing, streaming algorithms, derandomization and cryptography. For example, the work [SSS95] titled "Chernoff–Hoeffding bounds for applications with limited independence" has 500+ references according to Google scholar.

These bounds can be obtained from applying Markov's inequality on an upper bound of the higher moments of the Hamming weight $1^\top D := \sum_{i=1}^n D_i$ of the distribution [SSS95, BR94]. See Section 3.4 in [DP09] for details of this proof strategy and others.

**Fact 1.** *Let $D$ be a $(2k)$-uniform distribution on $\{-1,1\}^n$. For every integer $t > 0$, we have*

$$\mathbb{P}\big[|1^\top D| \geq t\big] \leq \sqrt{2}\left(\frac{2kn}{et^2}\right)^k.$$

Note that one can replace $k$ with any $k' \leq k$ in the bound, as any $k$-uniform distribution is also $k'$-uniform. Thus, this tail bound becomes non-trivial whenever $t \geq c\sqrt{n}$. In this paper, as in [Vio23], every occurrence of "$c$" denotes a possibly different positive real number. The notation "$c_x$" for parameter(s) $x$ indicates that this number may depend on $x$ and only

on $x$. Replacing "$c$" with $O(1)$ everywhere is consistent with one common interpretation of the big-O notation.

When $t \le c\sqrt{n \log k}$, pseudorandomness results on $k$-uniformity against halfspaces [BGGP12, DGJ+10, DKN10] imply that every $k$-uniform distribution puts roughly the same mass on $\mathbb{P}[1^\top D \ge t]$ as $B$.

**Theorem 2** ([BGGP12, DGJ+10, DKN10])**.** *Let $D$ be a $k$-uniform distribution on $\{-1,1\}^n$. For every integer $t > 0$, we have $|\mathbb{P}[1^\top D \ge t] - \mathbb{P}[B \ge t]| \le c/\sqrt{k}$.*

Therefore, for thresholds in this regime, essentially the upper and lower bounds on the tail of the (centered) binomial distribution extend to $k$-uniform distributions.

We note that tight estimates on the moments of the sum of independent bounded random variables have been established [Sko22]. But these bounds do not imply the tightness of Fact 1. More generally, and perhaps surprisingly, lower bounds on the tail mass of $k$-uniform distributions remain scarce. This question was revisited by Bun and Steinke [BS15]. They show that for every $k \ge c \log n$, there exists a $k$-uniform distribution $D$ such that $\mathbb{P}[1^\top D \ge t] \ge c^{-k}$ for $t = c\sqrt{nk}$. However, their result only applies to $k \ge c \log n$ and ties $t$ and $k$, and so it does not apply in several regimes of interest.

**Our results.** In this work, we obtain matching lower bounds to Fact 1 and Theorem 2, generalizing or strengthening a number of previous works.

**Theorem 3.** *For every $k$ and $t$, there exists a $k$-uniform distribution $D$ on $\{-1,1\}^n$ such that*

$$\mathbb{P}[1^\top D \ge t] - \mathbb{P}[B \ge t] \ge c\sqrt{\tfrac{n}{k}}\,\mathbb{P}[B = t] \ge c2^{-t^2/n}/\sqrt{k}.$$

Note that $\mathbb{P}[B = t]$ on the right hand side cannot be replaced with $\mathbb{P}[B \ge t]$.

Theorem 3 matches Theorem 2 up to constant, in particular removing a logarithmic factor from a lower bound sketched in [BGGP12]. Moreover, it shows that Fact 1 is tight for $t \in [\sqrt{n}, c\sqrt{nk}]$.

For $t \ge c\sqrt{nk}$, we obtain the following lower bound, again matching Fact 1.

**Theorem 4.** *For every $k \le (n/9)^{1/3}$ and $t \ge \sqrt{nk}$, there exists a $k$-uniform distribution $D$ on $\{-1,1\}^n$ such that $\mathbb{P}[1^\top D \ge t] \ge \frac{1}{3k^{3/2}}(\frac{kn}{16t^2})^{k/2}$.*

Theorem 4 generalizes [BS15]. To illustrate this regime, note that in particular for $t = c\sqrt{n \log n}$ we show that the error is large: $\ge (c/\log n)^k$, whereas $\mathbb{P}[B \ge t] \le 1/n^c$ is exponentially smaller.

There remain some parameter regimes that are not covered by our or previous works. For example, does *every* $k$-uniform distribution put mass at least $1/k^c$ on strings of weight $c\sqrt{n \log k}$? In general, we raise the question of establishing the best possible error for any threshold $t$ between $k$-uniform distributions and binomial, interpolating between the error $c/\sqrt{k}$ for $t \le c\sqrt{n \log k}$ [BGGP12, DGJ+10, DKN10] and the exponentially small error for larger $t$ (Fact 1 and Theorem 4).

## 1.2 Small-bias

We develop a paradigm to obtain small-bias distributions from $k$-uniform distributions while retaining some of their deviation properties. The paradigm has two steps. First, *symmetrize* the distribution. As $k$-uniform distributions are typically supported on nearly balanced strings, this step has the effect of making the bias small on tests of size not too large. Second, add noise, following [LV17]. This makes the bias small on large tests. Moreover, the small-bias distributions that we construct are supported on few Hamming weights.

**Lemma 5.** *Let $D$ be any $k$-uniform distribution. There exists a distribution $D_{\mathsf{bias}}$ supported on $ck^2$ Hamming weights that is simultaneously $k$-uniform and $(ck/n)^{k/4}$-biased such that for every interval $[a, b]$,*

$$\mathbb{P}\big[1^\top D_{\mathsf{bias}} \in [a-k, b+k]\big] \geq \mathbb{P}\big[1^\top D \in [a, b]\big].$$

We note that while the idea of adding noise is from [LV17], here we inject much less noise than in [LV17]. This is made possible by the symmetrization step, which lets us reduce the bias for moderate-size parities. On the other hand, symmetrization destroys linearity which is essential in [LV17].

Using this approach, we carry results on the Hamming weight of $k$-uniform distributions, including our new results for $k$-uniformity (Theorems 3 and 4), to small-bias distributions.

First, we obtain the following theorem stating that there are small-bias distributions supported on few Hamming weights that are nearly "balanced." This is obtained from the construction of $k$-uniform distributions in [BHLV19] (Lemma 22).

**Theorem 6.** *For every $k$, there exists a $(ck/n)^k$-biased distribution $D$ supported on at most $ck^2$ weights such that $|1^\top D| \leq 21\sqrt{kn}$ always.*

Now we state results corresponding to Theorem 3 and Theorem 4.

**Theorem 7.** *For every $k$ and $t$, there exists a $(ck/n)^k$-biased distribution $D$ supported on at most $ck^2$ weights such that*

$$\mathbb{P}[1^\top D \geq t-k] - \mathbb{P}[B \geq t] \geq c\sqrt{\tfrac{n}{k}}\,\mathbb{P}[B = t] \geq c2^{-t^2/n}/\sqrt{k}.$$

In typical settings, one can replace $t-k$ with $t$. However this in general depends on $t$ and $k$ so we leave the general statement for simplicity.

**Theorem 8.** *For every $k \leq (n/9)^{1/3}$ and $t \geq \sqrt{nk}$, there exists a $(ck/n)^{k/2}$-biased distribution $D$ supported on at most $ck^2$ weights such that*

$$\mathbb{P}[1^\top D \geq t] \geq \frac{1}{3k^{3/2}}\left(\frac{kn}{64t^2}\right)^{k/2}.$$

Several researchers, see for example [BS15], posed the question of understanding tail bounds for small-bias distributions. And yet, our understanding of their tail probabilities is notably lacking. In terms of upper bounds, existing tail bounds follow from Fact 1 via a connection to the closeness between small-bias and $k$-uniform distributions [AGM03, AAK+07, OZ18]. On

the opposite direction, Bazzi [Baz15] showed that an exponentially small-biased distribution does not fool the majority function with error $c/\sqrt{n}$. The recent work [DILV24] shows that $\mathbb{P}[1^\top D \geq \sqrt{nk}] \geq c^k$ for some $(ck/n)^k$-biased distribution $D$. However, this result applies to a specific threshold and does not scale as the threshold varies.

By contrast, Theorem 8 shows that the tail bounds for bounded uniformity cannot be improved even for small-bias distributions, in general parameter settings. This in particular answers a question in [BS15] and generalizes the recent work [DILV24].

### 1.3 Small-bias plus noise

In the past decade, researchers have considered bounded uniform and small-bias distributions which are perturbed by noise.

**Definition 9.** $N_\rho$ is the noise distribution on $\{-1,1\}^n$, where each bit is independently set to uniform with probability $1-\rho$ and 1 otherwise. We write $D \cdot N_\rho$ for the coordinate-wise product of $D$ and $N_\rho$, which corresponds to bit-wise xor over $\{0,1\}$. We also call this the *$\rho$-smoothed distribution*. Note that $x \cdot N_1 = x$ and $x \cdot N_0 = U$, for any $x$.

Note that smoothing does not increase the distance of any two distributions, with respect to any class of tests which is closed under shifts. So distinguishing smoothed distributions is at least as hard as distinguishing the corresponding (non-smooth) distributions. A main motivation for considering smoothed distributions comes from several paradigms for constructing pseudorandom generators (PRGs) that combine $(\varepsilon, k)$-biased distributions in different ways. These paradigms have been proposed in the last 15 years or so; for additional background, we refer the readers to the recent monograph [HH23].

A main result from [DILV24] is that smoothed $n^{-k}$-bias distributions do not fool thresholds with error less than $c^k$. This is then used to show that they do not fool, even with constant error, other models such as small-space algorithms or small-depth circuits.

**Our results.** Using the small-bias distributions constructed in this work, we obtain alternative proofs of the main result from [DILV24]. We note that these proofs provide different information. In [DILV24], the pseudorandom distributions put *more* mass than the binomial on the tail. In some of the proofs presented here, the mass will be *less* and the distribution is supported only on a *few* weights. This also provides a more complete picture of how these distributions can be designed.

In more detail, the small-bias distributions given in Theorems 6 and 8 put noticeable different masses on its tail than the binomial distribution. From this, one can argue that they remain so after perturbed with noise, and thus can be distinguished by thresholds.

**Theorem 10.** *For any $\rho \in (0,1]$ and $k \leq c\rho^2 n^{1/3}$, there is a $(ck/n)^k$-biased distribution $D$ on $\{-1,1\}^n$ and some threshold $t = c\sqrt{nk}/\rho$ such that*

$$\mathbb{P}[B \geq t] \geq \mathbb{P}[1^\top (D \cdot N_\rho) \geq t] + 2^{-ck/\rho^2}.$$

Moreover, we show that a threshold can distinguish smoothed small-bias from *any* $k$-uniform distribution, answering a question that was left open in [DILV24] in the affirmative.

**Theorem 11.** *For every $k \le (n/9)^{1/3}$ and $\rho \in [0,1)$, let $k' := c\log(1/\rho)k$. There exists a $(ck/n)^{k/2}$-biased distribution $D$ and a threshold $t$ such that for every $k'$-uniform distribution $D_{k'}$,*

$$\mathbb{P}[1^\top (D \cdot N_\rho) \ge t] \ge \mathbb{P}[1^\top D_{k'} \ge t] + \Big(\frac{c\rho^2}{\log(1/\rho)}\Big)^{k/2}.$$

Note that Theorem 11 shows that a smoothed small-bias distribution puts more mass on the tail than the uniform distribution (which is also $k'$-uniform), whereas Theorem 10 shows the opposite.

In addition, we show that any distribution supported on a few weights, after perturbed with noise, can be distinguished from the binomial distribution by a threshold. This uses a new bound on the total variation distance between any mixture of few Gaussian distributions with bounded variance the standard Gaussian distribution (Lemma 31).

**Theorem 12.** *Let $D$ be any distribution on $\{-1,1\}^n$ supported on $k$ weights. For any $\rho \in (0,1]$, there is some $t$ such that*

$$\big|\mathbb{P}[B \ge t] - \mathbb{P}[1^\top (D \cdot N_\rho) \ge t]\big| \ge 2^{-ck/\rho}.$$

Applying Theorem 12 to our small-bias distributions $D$ that are supported on $ck^2$ weights implies that their smoothed distributions $D \cdot N_\rho$ can be distinguished from uniform with advantange $2^{-ck^2/\rho}$. We remark that one can further improve the advantage to $2^{-ck/\rho}$, by applying Lemma 20 to "sparsify" any $k$-uniform distribution so that it is supported on $k+1$ weights, and then observing that $D \cdot N_{2\rho} \equiv (D \cdot N_\rho) \cdot N_\rho$ and $D \cdot N_\rho$ is small-biased.

# 2 Proof of Theorem 4

At a high level, the proof of Theorem 4 follows the same strategy in [BS15]. However, there are some noticeable differences. First, we decouple the threshold and the error parameters. Second, we do not pass the argument to Gaussian distributions. Finally and most importantly, their proof incurs a loss of a $1/\sqrt{n}$ factor in their lower bound on the tail mass (and thus requires $k \ge c\log n$), which is significant in certain regimes of parameters.

To prove Theorem 4, we use tools in approximation theory. In particular, to remove the $1/\sqrt{n}$ loss in [BS15], instead of applying a Markov–Bernstein type inequality for $L_\infty$-norms, we rely on the following inequality by Erdéyi,

**Lemma 13** (Theorem 2.1 ($q = 1$ case) in [Erd16])**.** *For $m \in \mathbb{N}$ and $L > 0$, let $Q \in \mathbb{C}[x]$ be a degree-$d$ univariate polynomial (possibly with complex coefficients) such that*

$$|Q(0)| > \frac{1}{L}\Big(\sum_{j=1}^{m} |Q(j)|\Big).$$

*Then $d \ge 7\sqrt{m/L}$.*

We also need the following inequality due to Ehlich, Zeller, Coppersmith, Rivlin, and Cheney.

**Lemma 14** (Lemma 20 in [BS15]). *Let $p$ be a univariate degree-$d$ polynomial such that $|p(i)| \le 1$ on $i \in \{0, \dots, m\}$, where $3d^2 \le m$. Then $|p(x)| \le 3/2$ for every $x \in [0, m]$.*

We will also use the following extremal property of Chebyshev polynomials $T_k$ (cf. [SV13, Propositions 2.4 and 2.5]).

**Fact 15.** *Let $p$ be a univariate polynomial of degree $k$ such that $|p(t)| \le 1$ on $[-1, 1]$. For every $s \ge 1$, $|p(s)| \le T_k(s) \le (2|s|)^k$, where $T_k$ is the Chebyshev polynomial of degree $k$.*

**Fact 16** (cf. Lemma 23 in [BHLV19]). *If $a$ is an integer such that $|a| \le n$ and $a \equiv n \bmod 2$, then $\mathbb{P}[B = a] \ge 2^{-a^2/n} \frac{1}{2\sqrt{n}}$.*

**Proof of Theorem 4.** By strong duality (cf. [BS15]), we have

$$\max_D \mathbb{P}\big[\mathbb{1}(1^\top D \ge t)\big] = \min_p \mathbb{E}\big[p(U)\big],$$

where the maximum is over all $k$-uniform distributions $D$, and the minimum is over all degree-$k$ (upper sandwiching) polynomials $p\colon \{-1, 1\}^n \to \mathbb{R}$ such that $p(x) \ge \mathbb{1}(1^\top x \ge t)$.

Let $p$ be a degree-$k$ polynomial attaining $\delta := \min_p \mathbb{E}[p(U)]$. Define the univariate polynomial $q$ to be the symmetrization of $p$, that is, $q(\sum_{i=1}^n x_i) := p(x)$.

We use Lemmas 13 and 14 to bound $q(t)$ on $t \in [-\sqrt{kn}, \sqrt{kn}]$. We first state the lemma and defer its proof to the end of this section.

**Lemma 17.** *Suppose $3k^2 \le \sqrt{kn}$. Then we have $|q(t)| \le 3\delta \cdot k^{3/2} \cdot 2^k$ for every $t \in [-\sqrt{kn}, \sqrt{kn}]$.*

Note that the upper bound is independent on $n$, whereas the bound in [BS15, Theorem 9] has a polynomial dependence on $n$.

We now continue the proof assuming Lemma 17. Let $s := \sqrt{kn}$. Observe that $q(t) \ge \mathbb{1}(t \ge t) = 1$. Let $\widetilde{q}_s(\theta) = q(\theta s)$. By Lemma 17, we have $\max_{\theta \in [-1,1]} |\widetilde{q}(\theta)| \le 3\delta \cdot k^{3/2} \cdot 2^k$. By Fact 15, for $t \ge s$, we have

$$\begin{aligned} 1 \le q(t) &= \widetilde{q}_s\Big(\frac{t}{s}\Big) \\ &\le \Big(\frac{2t}{s}\Big)^k \cdot \max_{\theta \in [-1,1]} |\widetilde{q}_s(\theta)| \\ &\le \Big(\frac{2t}{\sqrt{kn}}\Big)^k \cdot 3\delta \cdot k^{3/2} \cdot 2^k. \end{aligned}$$

Rearranging gives $\delta \ge \frac{1}{3k^{3/2}} (\frac{kn}{16t^2})^{k/2}$, proving Theorem 4. □

To prove Lemma 17, we use Lemma 13 to bound $q(t)$ on the integer points between $-\sqrt{kn}$ and $\sqrt{kn}$, and then extend the bound to the whole interval using Lemma 14.

**Claim 18.** $0 \le q(t) \le 2\delta \cdot k^{3/2} \cdot 2^k$ *for every* $t \in \{-\sqrt{kn}, \dots, \sqrt{kn}\}$,

*Proof of Claim 18.* Assume $n$ is even so that $B$ is supported on even integers. Fix an even integer $t_0$ such that $|t_0| \le \sqrt{kn}$ and without loss of generality assume $t_0 > 0$. As $\mathbb{P}[B = t] \ge \mathbb{P}[B = \sqrt{nk}]$ for every even $t$ with $|t| \le \sqrt{nk}$ and $q$ is nonnegative, we have

$$\begin{aligned}\mathbb{P}\Big[B = \sqrt{kn}\Big] \sum_{j=1}^{\sqrt{nk}/2} q(t_0 - 2j) &\le \sum_{j=1}^{\sqrt{nk}/2} \mathbb{P}\big[B = t_0 - 2j\big] q(t_0 - 2j) \\ &\le \mathbb{E}[q(B)] = \delta.\end{aligned}$$

Rearranging and using Fact 16 gives

$$\sum_{j=1}^{\sqrt{nk}/2} q(t_0 - 2j) \le \frac{\delta}{\mathbb{P}[B = \sqrt{kn}]} \le 2\delta \cdot \sqrt{n} \cdot 2^k.$$

Consider the polynomial $Q(t) := q(t_0 - 2t)$ of degree $k$. Let $m = \sqrt{nk}/2$, and $L = m/k^2 = n^{1/2}k^{-3/2}$. As $k < 7k = 7\sqrt{m/L}$, by the contrapositive of Lemma 13, we have

$$\begin{aligned}|q(t_0)| = |Q(0)| &\le \frac{1}{L}\sum_{j=1}^{m}|Q(j)| && \text{(Lemma 13)} \\ &= \frac{1}{L}\sum_{j=1}^{\sqrt{kn}/2} q(t_0 - 2j) \\ &\le 2\delta \cdot \frac{\sqrt{n} \cdot 2^k}{L} = 2\delta \cdot k^{3/2} \cdot 2^k. && \square\end{aligned}$$

*Proof of Lemma 17.* Let $Q$ be the degree-$k$ polynomial $Q(j) := q(\sqrt{kn} - 2j)$. By Claim 18, we have $|Q(j)| \le M$ on $j \in \{0, \dots, \sqrt{kn}\}$, where $M := 2\delta \cdot k^{3/2} \cdot 2^k$. As $3k^2 \le \sqrt{kn}$ for $k \le (n/9)^{1/3}$, by Lemma 14, we have $q(t) \le 3M/2$ for every $t \in [0, \sqrt{kn}]$. $\square$

# 3 Proof of Theorem 3

We rely on the result from [BHLV19, Theorem 2] that for every $a$, $m$, and $k \le n/(8m^2)$, there is a $k$-uniform distribution $D$ supported on $\{x \in \{-1, 1\}^n : 1^\top x \equiv a \pmod m\}$. Moreover, implicit in the proof they show that the probability mass on every point $s$ in the support of $D$ is at least $(m/4) \cdot \mathbb{P}[B = s]$.

Let $m := \sqrt{n/(8k)}$. We can pick an integer $a$ such that $t$ belongs to the support of some $k$-uniform $D$, from which we conclude that

$$\varepsilon := \mathbb{P}[1^\top D = t] - \mathbb{P}[B = t] \ge (m/4 - 1)\,\mathbb{P}[B = t].$$

Now, we have either

$$\begin{aligned}\mathbb{P}[1^\top D \ge t] - \mathbb{P}[B \ge t] &\ge \varepsilon/2 \text{ or} \\ \mathbb{P}[1^\top D \le t] - \mathbb{P}[B \le t] &\ge \varepsilon/2,\end{aligned}$$

as otherwise, summing both inequalities give $(1 + \mathbb{P}[1^\top D = t]) - (1 + \mathbb{P}[B = t]) < \varepsilon$, a contradiction. If we are in the second case, we can consider $\overline{D}$, the complement of $D$, which is also $k$-uniform, and we have $\mathbb{P}[1^\top D \le t] = \mathbb{P}[1^\top(\overline{D}) \ge t]$.

# 4 From bounded-uniformity to small-bias

In this section, we prove Lemma 5, which gives a generic way to transform any $k$-uniform distribution into a $(ck/n)^{k/4}$-bias distribution while preserving some of the weight properties of the distribution.

We start with a lemma that lets us "sparsify" the weight distribution of any $k$-uniform distribution. This uses Carathéodory's theorem from convex geometry, stated next, which has a simple proof (see e.g. Theorem 2.3 in Chapter 1 in [Bar02]).

**Lemma 19** (Carathéodory's theorem)**.** *Every point in the convex hull of a set $S \subseteq \mathbb{R}^k$ can be represented as a convex combination of $k+1$ points from $S$.*

We use Lemma 19 to sparsify any *symmetric* $k$-uniform distribution so that it is supported on $k+1$ weights. Here we use the fact that a symmetric distribution on $\{-1,1\}^n$ is $k$-uniform if and only if the first $k$ moments of $1^\top D$ match the ones of $B$.

**Lemma 20.** *Let $D$ be any symmetric $k$-uniform distribution such that $1^\top D$ is supported on a set $S \subseteq \{-1,1\}^n$. Then there is a symmetric $k$-uniform distribution $D'$ such that $1^\top D'$ is supported on a subset $S' \subseteq S$ of size at most $k+1$.*

*Proof.* For each $i$, define $p_i := \mathbb{P}[1^\top D = i]$ and $v^i := (i, i^2, \ldots, i^k) \in \mathbb{R}^k$. Let $b \in \mathbb{R}^k$ be the vector $(\mathbb{E}[B], \mathbb{E}[B^2], \ldots, \mathbb{E}[B^k])$. As $D$ is $k$-uniform, we have $b = \sum_{i \in S} p_i v^i$, and so $b$ lies in the convex hull of the $v_i$'s. Thus, there is a set $S'$ of $k+1$ indices such that $b = \sum_{i \in S'} q_i v_i$.

We define the symmetric distribution $D'$ by $\mathbb{P}[1^\top D = i] := q_i \mathbb{1}(i \in S')$ It is clear that $\sum_i \mathbb{P}[1^\top D' = i] v^i = b$, and thus $D'$ is $k$-uniform. □

We will also use the following bound on the bias of the uniform distribution on a "Hamming slice." This follows from an upper bound on Krawtchouk polynomials that was established in [DILV24].

**Lemma 21** (Claim 31 and Corollary 16 in [DILV24])**.** *Let $W_t$ be the uniform distribution on $\{x \in \{-1,1\}^n : \sum_{i=1}^n x_i = t\}$. For every subset $S \subseteq [n]$ of size $\ell$, we have*

$$\left|\mathbb{E}\left[W_t^{[n]\setminus S}\right]\right| = \left|\mathbb{E}\left[W_t^S\right]\right| \leq \left(\frac{\ell}{n} + \frac{t^2}{n^2}\right)^{\frac{\ell}{2}}.$$

**Proof of Lemma 5.** Let $D$ be any $k$-uniform distribution on $\{-1,1\}^n$. We obtain our small-bias distribution $D_{\mathsf{bias}}$ in 3 steps as follows:

First, we symmetrize $D$ to obtain $D_{\mathsf{sym}}$. Then, we use Lemma 20 to sparsify $D_{\mathsf{sym}}$ to obtain $D'$ so that it is supported on $k+1$ weights. Finally, we add $k/2$ bits of noise, obtained by repeating the following process independently $k/2$ times: pick a uniform random coordinate of $D'$ and set it to uniform.

We now claim that $D_{\mathsf{bias}}$ is $(ck/n)^{\frac{k}{4}}$-biased. Let $S \subseteq [n]$ be any non-empty subset. Note that the noise applied in the last step can only reduce the bias of $D'$. We consider three cases:

1. $|S| \in [1,k]$: observe that $D'$ remains $k$-uniform. So $\mathbb{E}[D'^S] = 0$.

2. $|S| \in [k+1, n-k-1]$: let $t = (kn^3)^{\frac{1}{4}}$. By Lemma 21 and Fact 1, we have

$$\begin{aligned}\left|\mathbb{E}[D'^S]\right| &\le \left|\mathbb{E}\left[D'^S \mid |1^\top D'| \le t\right]\right| + \mathbb{P}\left[|1^\top D'| > t\right] \\ &\le \left(\frac{|S|}{n} + \frac{t^2}{n^2}\right)^{|S|/2} + \sqrt{2}\left(\frac{kn}{et^2}\right)^{k/2} \\ &\le 2 \cdot (2k/n)^{\frac{k}{4}}.\end{aligned}$$

3. $|S| \in [n-k, n]$: If a bit in $S$ is set to uniform by the noise, which happens with probability $(1-k/n)$, then the bias is 0. So

$$\left|\mathbb{E}[D^S_{\mathsf{bias}}]\right| \le (1-(1-k/n))^{k/2} = (k/n)^{k/2}.$$

Since changing a bit $x_i \in \{-1, 1\}$ of $x$ can only change $1^\top x$ by at most 2, the lemma follows. □

### 4.1 Proofs of Theorems 6, 7 and 8

Theorems 7 and 8 directly follow from applying Lemma 5 to Theorems 3 and 4 respectively. Theorem 6 follows from applying the lemma to following result, which exhibits bounded uniform distributions that are supported only on nearly-balanced strings.

**Lemma 22** ([BHLV19])**.** *For any integer $k$, there is a distribution $D$ supported on $\{x \in \{-1,1\}^n : |1^\top x| \le 10\sqrt{kn}\}$ which is $(2k)$-uniform.*

We note that one can instead apply Lemma 5 to the standard randomness-efficient construction of $k$-uniform distributions via BCH codes, and then use results from algebraic geometry to bound the Hamming weight, specifically Theorem 18 in [MS77]. However, the support is slightly less concentrated to the center than Lemma 22.

**Claim 23.** *There exists a linear distribution $D$ supported on $\{x \in \{-1,1\}^n : |1^\top x| \le ck\sqrt{n}\}$ which is $k$-uniform.*

## 5 Distinguishing small-bias plus noise

In this section we prove Theorems 10, 11 and 12. We will apply several results on sums of independent random variables. We first state these results before proceeding to the proofs.

Let $\mathcal{N}(0,1)$ denote the standard normal distribution, which has mean 0 and variance 1.

**Lemma 24** (Theorem 11.2 in [Das08])**.** *Let $Y_1, \ldots, Y_n$ be $n$ independent random variables with $\mathbb{E}[Y_i] = 0$, $\mathbf{Var}[Y_i] = \sigma_i^2$, $\mathbb{E}[|Y_i|^3] < \infty$. Let $Y := \sum_{i=1}^n Y_i$. For every $\theta \in \mathbb{R}$,*

$$\left|\mathbb{P}\left[\frac{Y}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} \ge \theta\right] - \mathbb{P}\left[\mathcal{N}(0,1) \ge \theta\right]\right| \le \frac{\sum_{i=1}^n \mathbb{E}[|Y_i|^3]}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}}.$$

For fixed $\rho_i$'s and $\sigma_i$'s, the additive error given by Lemma 24 is roughly $1/\sqrt{n}$. So, for Theorem 12 to hold with $k \geq c \log n$, we would need a more refined approximation. For this reason, we will be using the following Cramér's estimate of sums of independent random variables, which gives a multiplicative rather than additive approximation in terms of $\mathcal{N}(0,1)$.

**Lemma 25** (Chapter VIII, Equation (2.41) in [Pet75])**.** *There exists a constant $c > 0$ such that the following holds. Let $Y_1, \ldots, Y_n$ be $n$ independent random variables with $\mathbb{E}[Y_i] = 0$, and $\mathbb{E}[Y_i^2] = \sigma_i^2$ for each $i \in [n]$. Let $Y := \sum_{i=1}^n Y_i$. For $0 \leq \theta \leq cn^{1/6}$, there exists an $\varepsilon \in [0, \frac{(\theta+1)}{c\sqrt{n}}]$ such that*

$$\mathbb{P}\left[\frac{Y}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} \geq \theta\right] = \mathbb{P}\Big[\mathcal{N}(0,1) \geq \theta\Big] \cdot \exp\left(\frac{\sum_{i=1}^n \mathbb{E}[Y_i^3]}{6\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} \cdot \theta^3\right)(1+\varepsilon).$$

To relate Lemma 25 to Lemma 24, note that when $\theta$ is small, $\exp(\frac{\sum_{i=1}^n \mathbb{E}[Y_i^3]}{6(\sum_{i=1}^n \sigma_i^2)^{3/2}} \cdot \theta^3)$ is roughly $1 + \frac{\sum_{i=1}^n \mathbb{E}[Y_i^3]}{(\sum_{i=1}^n \sigma_i^2)^{3/2}}$.

Specializing Lemma 25 to our applications, we obtain the following lemma. We first need a simple claim.

**Claim 26.** *For every $x \in \{-1,1\}^n$ and $i \in [n]$, let $Y_i$ be the mean zero variables $Y_i := (x \cdot N_\rho)_i - \rho x_i$. Then $\mathbb{E}[Y_i^2] = 1 - \rho^2$ and $\mathbb{E}[Y_i^3] \leq -2\rho(1-\rho^2)x_i \in [-1, 0]$.*

*Proof.* It suffices to compute the first 3 moments of $(x \cdot N_\rho)_i$. We have $\mathbb{E}[(x \cdot N_\rho)_i] = \rho x_i$, $\mathbb{E}[(x \cdot N_\rho)_i^2] = 1$, and $\mathbb{E}[(x \cdot N_\rho)_i^3] = \rho x_i$. □

**Lemma 27.** *For every $x \in \{-1,1\}^n$, $\rho \in [0,1)$, and $\theta \in [0, cn^{1/6}]$, we have*

1. $\mathbb{P}[B \geq \sqrt{n} \cdot \theta] \geq \mathbb{P}[\mathcal{N}(0,1) \geq \theta]$, *and*
2. $\mathbb{P}\left[\frac{1^\top (x \cdot N_\rho) - \rho \cdot 1^\top x}{\sqrt{n}(1-\rho^2)^{1/2}} \geq \theta\right] \leq 2\,\mathbb{P}\Big[\mathcal{N}(0,1) \geq \theta\Big]$.

*Proof.* We apply Lemma 25 to obtain both inequalities. For the first inequality, observe that $\mathbb{E}[Y_i^3] = \mathbb{E}[B_i^3] = 0$. For the second inequality, we apply Lemma 25 using Claim 26, which gives $\mathbb{E}[Y_i^3] \leq 0$. Note that $e^y \leq 1$ for any $y \leq 0$, and $c(1+\theta)/\sqrt{n} \leq 1$. □

We will use the following approximation to compare the tails of the standard normal distribution.

**Lemma 28** (Lemma 22.2 in [Kle20])**.** *For any $\theta > 0$,*

$$\frac{1}{\theta + \frac{1}{\theta}} \leq \mathbb{P}\Big[\mathcal{N}(0,1) \geq \theta\Big] \cdot \frac{\sqrt{2\pi}}{e^{-\theta^2/2}} \leq \frac{1}{\theta}.$$

Finally, we state the following tail bounds for sums of independent bounded random variables with small variances.

**Claim 29** (Bernstein's inequality)**.** *Let $Y_1, \ldots, Y_n$ be independent mean-zero random variables. Suppose $|Y_i| \leq M$ for every $i \in [n]$. Then*

$$\mathbb{P}\Big[\sum_{i=1}^n Y_i \geq t\Big] \leq \exp\left(-\frac{t^2/2}{\sum_{i=1}^n \mathbb{E}[Y_i^2] + Mt/3}\right).$$

Again, specializing Claim 29 to our applications, we obtain the following.

**Claim 30.** $\mathbb{P}[|1^\top(x \cdot N_\rho) - \rho \cdot 1^\top x| \geq s] \leq 2\exp(-\frac{cs^2}{(1-\rho^2)n+s})$ *for every* $x \in \{-1,1\}^n$ *and* $s > 0$.

*Proof.* We will only prove that $\mathbb{P}[1^\top(x \cdot N_\rho) - \rho \cdot 1^\top x \geq s] \leq \exp(-\frac{cs^2}{(1-\rho^2)n+s})$. The other direction is analogous and the conclusion follows from a simple union bound. Let us consider the mean zero variables $Y_i := (x \cdot N_\rho)_i - \rho x_i$. We have $|Y_i| \leq 1 + \rho \leq 2$. Applying Claim 29 with Claim 26, we have

$$\mathbb{P}\big[1^\top(x \cdot N_\rho) - \rho \cdot 1^\top x \geq s\big] = \mathbb{P}\Big[\sum_{i=1}^n Y_i \geq s\Big] \leq \exp\Big(-\frac{cs^2}{(1-\rho^2)n+s}\Big). \qquad \square$$

## 5.1 Proof of Theorem 10

Let $D$ be the $(ck/n)^k$-biased distribution in Theorem 8 such that $|1^\top D| \leq 21\sqrt{kn}$. Let $t := \beta\sqrt{kn}/\rho$ for a sufficiently large constant $\beta > 0$, and $\theta = t/\sqrt{n} = \beta\sqrt{k}/\rho$. Using $(1-\rho^2)^{-1/2} \geq 1 + \rho^2/2$, we have

$$\frac{t - 21\rho\sqrt{kn}}{\sqrt{n}(1-\rho^2)^{1/2}} = \sqrt{k} \cdot \frac{\frac{\beta}{\rho} - 21\rho}{(1-\rho^2)^{1/2}} \geq \beta\frac{\sqrt{k}}{\rho} \cdot \Big(1 - \frac{21\rho^2}{\beta}\Big) \cdot \Big(1 + \frac{\rho^2}{2}\Big) \geq (1 + \rho^2/4) \cdot \theta. \tag{1}$$

By assumption, $k \leq c\rho^2 n^{1/3}$ and so $\theta \leq cn^{1/6}$ for which Lemma 27 applies. As $|1^\top D| \leq 21\sqrt{kn}$, using the second inequality in Lemma 27 and (1), we have

$$\begin{aligned}\mathbb{P}\big[1^\top(D \cdot N_\rho) \geq t\big] &\leq \mathbb{P}\Big[1^\top(D \cdot N_\rho) - \rho \cdot 1^\top D \geq t - 21\rho\sqrt{kn}\Big] \\ &\leq 2\,\mathbb{P}\big[\mathcal{N}(0,1) \geq (1 + \rho^2/4) \cdot \theta\big]. \end{aligned} \tag{2}$$

On the other hand, from the first inequality of Lemma 27, we have

$$\mathbb{P}\big[B \geq t\big] \geq \mathbb{P}\big[\mathcal{N}(0,1) \geq \theta\big]. \tag{3}$$

By Lemma 28,

$$\begin{aligned}\mathbb{P}[\mathcal{N}(0,1) \geq \theta] &\geq \frac{1}{2\theta}\frac{1}{\sqrt{2\pi}}e^{-\theta^2/2} \\ &\geq \frac{c\sqrt{k}}{\rho}e^{-\theta^2/2} + \frac{2}{\theta}\frac{1}{\sqrt{2\pi}}e^{-(1+\rho^2/4)\theta^2/2} \\ &\geq e^{-ck/\rho^2} + 2\,\mathbb{P}\big[\mathcal{N}(0,1) \geq (1 + \rho^2/4) \cdot \theta\big]. \end{aligned} \tag{4}$$

Putting (2), (3) and (4) together completes the proof. $\square$

## 5.2 Proof of Theorem 11

Let $k' = C\log(1/\rho)k$ and $t = \sqrt{k'n}$. For every $k'$-uniform distribution $D_{k'}$, by Fact 1, we have

$$\mathbb{P}[1^\top D_{k'} \geq t] \leq \sqrt{2}\Big(\frac{k'n}{et^2}\Big)^{k'/2} \leq \rho^{Ck/2}.$$

We will construct a small-distribution $D$ that puts more mass on the tail of an even larger threshold than $t$. Specifically, let $t' = 2t/\rho$. Applying Theorem 8 to $k$ and $t'$, we obtain a $(ck/n)^{k/2}$-biased distribution $D$ such that

$$\mathbb{P}[1^\top D \geq t'] \geq \frac{1}{3k^{3/2}}\Big(\frac{\rho^2}{256C\log(1/\rho)}\Big)^{k/2}.$$

We now show that conditioned on $1^\top D \geq t'$, the smoothed distribution $D \cdot N_\rho$ still puts at least half the mass beyond $t$. Since $\rho t' - t \geq t$, by Claim 30,

$$\mathbb{P}\big[|1^\top(x \cdot N_\rho) - \rho \cdot 1^\top x| \geq t\big] \leq 2e^{-\frac{ct^2}{(1-\rho^2)n+t}} \leq 1/2.$$

Therefore

$$\mathbb{P}[1^\top(D \cdot N_\rho) \geq t] \geq \frac{1}{6k^{3/2}}\Big(\frac{\rho^2}{256C\log(1/\rho)}\Big)^{k/2} \geq \Big(\frac{c\rho^2}{\log(1/\rho)}\Big)^{k/2}. \quad \square$$

## 5.3 Proof of Theorem 12

We first state the main technical result we need, which may be of independent interest. We defer its proof to the next section.

**Lemma 31.** *Let $M$ be a mixture of $k$ Gaussian distributions each with variance $\sigma^2 = 1 - \rho^2$. Then there exists an interval $I$ such that*

$$\big|\mathbb{P}[\mathcal{N}(0,1) \in I] - \mathbb{P}[M \in I]\big| \geq 2^{-ck/\rho}.$$

*In particular, up to a factor 2 the same bound applies with the interval replaced with some threshold.*

*Proof of Theorem 12 assuming Lemma 31.* We first claim that $D \cdot N_\rho$ is $(cn^{-1/2})$-close to a mixture of $k$ Gaussian distributions with variance $1 - \rho^2$ in Kolmogorov (aka. CDF) distance. This follows from applying Lemma 24 and Claim 26 to any fixed weight $w$ of $D$. Specifically, we have that for every $x \in \{-1,1\}^n$ and $\theta$,

$$\big|\mathbb{P}\big[1^\top(x \cdot N_\rho) \geq \theta\sqrt{n}\big] - \mathbb{P}\big[\mathcal{N}(\mu, 1-\rho^2) \geq \theta\big]\big| \leq c/\sqrt{n},$$

for some $\mu$ that depends only on $1^\top x$ and $\rho$. On the other hand, again by Lemma 24, for every $\theta$, we have

$$\big|\mathbb{P}\big[B \geq \theta\sqrt{n}\big] - \mathbb{P}\big[\mathcal{N}(0,1) \geq \theta\big]\big| \leq 1/\sqrt{n}.$$

Combining the above with Lemma 31, we conclude that there exists some $\theta$ such that

$$\big|\mathbb{P}\big[B \geq \theta\sqrt{n}\big] - \mathbb{P}\big[1^\top(D \cdot N_\rho) \geq \theta\sqrt{n}\big]\big| \geq 2^{-ck/\rho} - c/\sqrt{n}. \quad \square$$

# 6 Proof of Lemma 31

We wish to show that a linear sum of $k$ exponential functions with variance $< 1$ cannot approximate the standard normal well. We can factor the two expressions so the mixture becomes a linear sum of $k$ exponential functions, which can be written as $\sum_{i\in[k]} a_i e^{b_i}$, while the standard normal can be written as $e^{\alpha x^2}$. This factoring crucially uses the fact the variances in the mixture are identical.

We then argue the distance must be large for some point. To prove this, we show the entries in the inverse of a Vandermonde like matrix which corresponds to the $e^{\alpha x^2}$ are not too large. This step is the bulk of the proof. On the other hand, the Vandermonde matrix corresponding to the sum of exponentials is singular. After some matrix norm manipulations this allows us to achieve the desired result.

**Lemma 32.** *Suppose that $f(x)$ is the PDF of a Gaussian distribution with variance 1, and $g(x)$ is the PDF of a mixture of $k$ Gaussian distributions with variance $1-\rho^2 = \sigma^2 < 1$. Then*

$$\|f - g\|_\infty \geq e^{-ck/\rho}.$$

We let $\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ denote the probability density function of $\mathcal{N}(0,1)$.

*Proof of Lemma 31 assuming Lemma 32.* Let $g(x)$ denote the PDF of $M$ and set $h(x) = \phi(x) - g(x)$. By Lemma 32, there exists some $a \in \mathbb{R}$ with

$$|h(a)| \geq e^{-ck/\rho}.$$

We have $|\phi'(x)| \leq (2\pi e)^{-1/2}$ and $|g'(x)| \leq (2\pi e)^{-1/2}/\sigma^2$ for all $x$, so $|h'(x)| \leq (2\pi e)^{-1/2}(1 + \frac{1}{\sigma^2}) < \frac{1}{2\sigma^2}$. We claim there exists an interval $I \subseteq \mathbb{R}$ with

$$|\,\mathbb{P}[\mathcal{N}(0,1) \in I] - P[M \in I]| \geq 2\sigma^2 \left(e^{-ck/\rho}\right)^2 = e^{-ck/\rho}.$$

To see this, assume that $h(0) \geq e^{-ck/\rho}$. Then $h(x) \geq e^{-ck/\rho} - \frac{1}{2\sigma^2}|x|$ for any $x$. So we set the interval $I = [-2\sigma^2 e^{-ck/\rho}, 2\sigma^2 e^{-ck/\rho}]$, and then $\int_I h(x)dx \geq 2\sigma^2(e^{-ck/\rho})^2$. Finally, note the assumption that $a = 0$ can be made without loss of generality. □

## 6.1 Proof of Lemma 32

The main technical result we need is the following.

**Lemma 33.** *Let $\alpha, D > 0$ be fixed. Let*

$$\Delta(k) := \inf_g \|e^{\alpha x^2} - g(x)\|_\infty$$

*where the infimum is over $g$ that are a linear combination of $k$ exponential functions, and the norm $\|\cdot\|_\infty$ is the supremum over the interval $[-D\sqrt{k}, D\sqrt{k}]$. Then we have*

$$\Delta(k) \geq \exp\left(\frac{-ck}{D^2\alpha}\right).$$

*Proof of Lemma 32 assuming Lemma 33.* Without loss of generality we may assume that $f(x)$ is the PDF of the standard normal distribution with mean 0 and variance 1. Define $\overline{f}(x) = e^{x^2/2\sigma^2} f(x) = e^{\alpha x^2}$ with $\alpha = \frac{1}{2\sigma^2} - \frac{1}{2}$, and $\overline{g}(x) = e^{x^2/2\sigma^2} g(x)$. Now $g(x)$ is a linear combination of $k$ exponential functions. If we choose some $D > 0$ then Lemma 33 gives us

$$\Delta(k) \geq \exp\Big(\frac{-ck}{D^2\alpha}\Big),$$

where $\Delta(k)$ is the supremum of $|\overline{f} - \overline{g}|$ over the interval $[-D\sqrt{k}, D\sqrt{k}]$. It follows that

$$\|f - g\|_\infty = \|e^{-x^2/2\sigma^2}(\overline{f} - \overline{g})\|_\infty \geq \exp\Big(\frac{-D^2 k}{2\sigma^2}\Big)\Delta(k) = \exp\Big(-k\Big(\frac{D^2}{2\sigma^2} + \frac{c}{D^2\alpha}\Big)\Big),$$

Then if we set $D^2 = c\sqrt{\sigma^2/\alpha}$ we have

$$\|f - g\|_\infty \geq \exp\Big(\frac{-ck}{\sqrt{\sigma^2\alpha}}\Big) = \exp\Big(\frac{-ck}{\sqrt{1-\sigma^2}}\Big). \qquad \square$$

## 6.2 Proof of Lemma 33

**Definition 34.** For a function $f : \mathbb{Z} \to \mathbb{R}$ define the $(k+1) \times (k+1)$ matrix $M_k(f)$ by $M_k(f)_{i,j} = f(i + j - k - 2)$. For example,

$$M_3(f) = \begin{pmatrix} f(-3) & f(-2) & f(-1) & f(0) \\ f(-2) & f(-1) & f(0) & f(1) \\ f(-1) & f(0) & f(1) & f(2) \\ f(0) & f(1) & f(2) & f(3) \end{pmatrix}$$

**Fact 35.** *If $f(x) = \sum_{i=1}^k a_i e^{b_i x}$, then $\det M_k(f) = 0$.*

*Proof.* We claim the sequence $\ldots, f(-1), f(0), f(1), f(2), \ldots$ satisfies a linear recurrence of order $k$, which implies the columns of $M_k(f)$ are linearly dependent.

To prove the claim, we show the existence of $c_1, \ldots, c_k$ such that

$$f(x) = c_1 f(x-1) + \cdots + c_k f(x-k).$$

Solving for $c_1, \ldots, c_k$, we obtain $k$ linear constraints

$$1 = \frac{c_1}{e^{b_1}} + \cdots + \frac{c_k}{e^{kb_1}}, \ldots, 1 = \frac{c_1}{e^{b_k}} + \cdots + \frac{c_k}{e^{kb_k}}.$$

There exists a solution to this system. $\square$

**Fact 36.** *If $|x| < 1$ then*

$$\prod_{i=1}^\infty (1 - x^i) \geq \exp\Big(\frac{-c}{1-x}\Big).$$

*Proof.* Using absolute convergence, and the inequality $(1 - x^j) \geq j x^j (1 - x)$, we get

$$\sum_{i=1}^\infty \log(1-x^i) = -\sum_{i=1}^\infty \sum_{j=1}^\infty \frac{x^{ij}}{j} = -\sum_{j=1}^\infty \sum_{i=1}^\infty \frac{x^{ij}}{j} = -\sum_{j=1}^\infty \frac{x^j}{j(1-x^j)} \geq -\sum_{j=1}^\infty \frac{1}{j^2(1-x)} = -\frac{\pi^2}{6(1-x)}.$$

Then apply the exponential function to both sides. $\square$

**Lemma 37.** *Suppose that $q > 1$ and $k$ is a positive integer. Let $A := M_k(q^{x^2})^{-1}$. Then*

$$|A_{i,j}| \leq \frac{\binom{k}{i-1}\binom{k}{j-1}}{\prod_{i=1}^{k}(1-q^{-2i})}.$$

*Proof of Lemma 33.* Let $f(x) = e^{\alpha x^2}$ and $g(x) = \sum_{i=1} a_i e^{b_i x}$. Define $\widetilde{f}(x) = f(Dx/\sqrt{k}) = q^{x^2}$, where $q = e^{D^2\alpha/k}$ and $\widetilde{g}(x) = g(Dx/\sqrt{k})$. By Fact 35, $M_k(\widetilde{g})$ is singular. Let $A$ be the inverse of $M_k(\widetilde{f})$. The matrix

$$AM_k(\widetilde{g}) = AM_k(\widetilde{f}) - AM_k(\widetilde{f} - \widetilde{g}) = I - AM_k(\widetilde{f} - \widetilde{g})$$

is singular. It follows that

$$\|A\|_\sigma \|M_k(\widetilde{f} - \widetilde{g})\|_\sigma \geq \|AM_k(\widetilde{f} - \widetilde{g})\|_\sigma \geq 1$$

where $\|A\|_\sigma$ is the spectral norm of $A$. The matrix $A$ is positive definite symmetric and the sum of the singular values is the sum of the eigenvalues which is equal to the trace of $A$. By Lemma 37 we get

$$\|A\|_\sigma \leq \mathrm{trace}(A) \leq \frac{\sum_{i=0}^{k}\binom{k}{i}^2}{\prod_{i=1}^{k}(1-q^{-2i})} = \frac{4^k}{\prod_{i=1}^{k}(1-q^{-2i})}.$$

On the other hand,

$$\|M_k(\widetilde{f} - \widetilde{g})\|_\infty \leq (k+1)\|f - g\|_\infty.$$

Combining everything and using Fact 36, we get

$$\|f - g\|_\infty \geq \frac{\prod_{i=1}^{k}(1-q^{-2i})}{4^k(k+1)} \geq \exp\left(\frac{-c}{1-q^{-2}} - \log(4)k - \log(k+1)\right) = \exp\left(\frac{-ck}{D^2\alpha}\right). \quad \square$$

## 6.3 Proof of Lemma 37

First we define Vandermonde matrices.

**Definition 38.**

$$\mathrm{Vand}(x_0, x_1, x_2, \ldots, x_k) = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_0 & x_1 & x_2 & \cdots & x_k \\ x_0^2 & x_1^2 & x_2^2 & \cdots & x_k^2 \\ \vdots & \vdots & \vdots & & \vdots \\ x_0^k & x_1^k & x_2^k & \cdots & x_k^k \end{pmatrix}.$$

*Proof.* We can transform $M_k(q^{x^2})$ into $\mathsf{Vand}(1, q^2, q^4, \ldots, q^{2k})$ by multiplying the rows and columns of $M_k(q^{x^2})$ with powers of $q$. Thus by Proposition 39, stated at the end, $(-1)^{i+j}A_{i,j}\prod_{b=1}^{k}(q^{2b}-$

1) is a sum of $\binom{k}{i-1}\binom{k}{j-1}$ powers of $q$. This implies $(-1)^{i+j}A_{i,j}\prod_{b=1}^{k}(1-q^{-2b})$ is also a sum of $\binom{k}{i-1}\binom{k}{j-1}$ powers of $q$.

Next we claim no positive powers of $q$ appear in the aforementioned sum. We define

$$B_k(q) := \begin{pmatrix} q^{(-k)^2/2} & & & \\ & q^{(2-k)^2/2} & & \\ & & \ddots & \\ & & & q^{k^2/2} \end{pmatrix}$$

so that we can write

$$M_k(q^{x^2}) = B_k(q)C_k(q)B_k(q)$$

where $C_k(q)$ is a matrix with 1 on the diagonal and negative powers of $q$ outside of the diagonal. In particular, $C_k(q)$ converges to the identity matrix as $q \to \infty$. So

$$A = A(q) = M_k(q^{x^2})^{-1} = B_k(q)^{-1}C_k(q)^{-1}B_k(q)^{-1}$$

converges as $q \to \infty$ because both $B_k(q)^{-1}$ and $C_k(q)^{-1}$ converge. This shows that $A_{i,j}\prod_{b=1}^{k}(1-q^{-2b})$ cannot have positive powers of $q$ in its corresponding sum. Since

$$|A_{i,j}|\prod_{b=1}^{k}(1-q^{-2b})$$

is a sum of $\binom{k}{i-1}\binom{k}{j-1}$ non-positive powers of $q$ and $q > 1$ we get

$$|A_{i,j}|\prod_{b=1}^{k}(1-q^{-2b}) \leq \binom{k}{i-1}\binom{k}{j-1}.$$ □

**Proposition 39.** *Let* $V = \mathrm{Vand}(1, q, q^2, \ldots, q^k)$. *Then*

$$(-1)^{i+j}(V^{-1})_{i,j}\prod_{b=1}^{k}(q^b-1)$$

*is a sum of* $\binom{k}{i-1}\binom{k}{j-1}$ *powers of* $q$.

*Proof.* Note that $\det(V) = \prod_{0\leq a<b\leq k}(q^b - q^a)$. Let $\widetilde{V}_i$ be the matrix $V$ with the $i$-th column removed, and $\widetilde{V}_{j,i}$ be the matrix $V$ with the $j$-th row and $i$-th column removed. By the formula of $V^{-1}$ from Cramer's rule we get

$$(V^{-1})_{i,j} = \frac{(-1)^{i+j}\det(\widetilde{V}_{j,i})}{\det(V)}.$$

Note that $\widetilde{V}_{k+1,i} = \mathrm{Vand}(1, q, \cdots, q^{i-2}, q^i, \ldots, q^{k+1})$, so

$$\det(\widetilde{V}_{k+1,i}) = \prod_{\substack{0\leq a<b\leq k \\ a,b\neq i-1}}(q^b - q^a).$$

So we have

$$\frac{\prod_{b=1}^{k}(q^b-1)\det(\widetilde{V}_{k+1,i})}{\det(V)} = \frac{\prod_{b=1}^{k}(q^b-1)}{\prod_{j=i}^{k}(q^j-q^{i-1})\prod_{j=0}^{i-2}(q^{i-1}-q^j)}$$

which is up to a power of $q$ factor equal to

$$\frac{\prod_{j=1}^{k}(q^j-1)}{\prod_{j=1}^{k-i+1}(q^j-1)\prod_{j=1}^{i-1}(q^j-1)} = \binom{k}{i-1}_q,$$

where the right-hand side is a Gaussian $q$-binomial coefficient which is a sum of $\binom{k}{i-1}$ powers of $q$. To see this, consider the generating function $\prod_{j=0}^{k-1}(1+q^jt) = \sum_{j=0}^{k} q^{j(j-1)/2}\binom{k}{j}_q t^j$. This implies that $\binom{k}{i-1}_q$ is a sum of $\binom{k}{i-1}$ powers of $q$.

Next, by Claim 40 we have that

$$\frac{\det(\widetilde{V}_{j,i})}{\det(\widetilde{V}_{k+1,i})} = e_{k+1-j}(1,q,\ldots,q^{i-2},q^i,\ldots,q^k)$$

is a sum of $\binom{k}{j-1}$ powers of $q$. We conclude that

$$(-1)^{i+j}(V^{-1})_{i,j}\prod_{b=1}^{k}(q^b-1) = \frac{\det(\widetilde{V}_{j,i})}{\det(\widetilde{V}_{k+1,i})}\cdot\frac{\prod_{b=1}^{k}(q^b-1)\det(\widetilde{V}_{k+1,i})}{\det(V)}.$$

is a sum of $\binom{k}{i-1}\binom{k}{j-1}$ powers of $q$. □

**Claim 40.** *Let* $X = \mathrm{Vand}(x_0,\ldots,x_k)$. *Then*

$$\frac{\det(\widetilde{X}_{j,i})}{\det(\widetilde{X}_{k+1,i})} = e_{k+1-j}(x_0,\ldots,x_{i-2},x_i,\ldots,x_k).$$

*Proof.* Without loss of generality assume $i = 1$. Let $X' := \widetilde{X}_{k+1,1}$, and note $X' = \mathrm{Vand}(x_1,\ldots,x_k)$. We first sketch out the polynomial argument which proves that $\det(X') = \prod_{1\le a<b\le k}(x_b-x_a)$ (see [Wik]).

For $b\neq a$, $(x_b-x_a)$ is a factor of $\det(X')$, since if we replace $x_b$ with $x_a$ in $X'$ then the determinant becomes 0. Thus, $\det(X') = p_1\prod_{1\le a<b\le k}(x_b-x_a)$ for some polynomial $p_1$.

Now by the Leibniz formula for the determinant, since all entries of the $j$th row have degree $j-1$, $\det(X')$ is a homogeneous polynomial of degree $1+\cdots+k-1 = k(k-1)/2$. This implies that $p_1$ is a constant.

Finally, $p_1 = 1$ since the product of the diagonal entries of $X'$ is $x_2x_3^2\ldots x_k^{k-1}$, which is the monomial obtained by taking the first entry of each term in $\prod_{1\le a<b\le k}(x_b-x_a)$.

Next we have that $\det(\widetilde{X}_{j,i}) = p_2\prod_{1\le a<b\le k}(x_b-x_a)$ for some homogeneous polynomial $p_2$ of degree $k-(j-1)$. This follows by repeating the start of the previous argument.

Moreover we claim that $p_2$ is symmetric and squarefree. The first claim follows since if we swap $x_b$ with $x_a$ in $\widetilde{X}_{j,i}$ that will only change the sign of the determinant, and this occurs in $\prod_{1\le a<b\le k}(x_b-x_a)$. The second claim follows since there are terms in $\prod_{1\le a<b\le k}(x_b-x_a)$ where the degree of an individual variable is $k-1$. Thus if $p_2$ is not square free, the degree of these variables becomes $>k$, which is a contradiction.

This implies that $p_2 = e_{k+1-j}(x_1,x_2,\ldots,x_k)$. □

**Acknowledgements.** We thank Rocco Servedio for pointing us to [Pet75].

# References

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *ACM Symp. on the Theory of Computing (STOC)*, pages 496–505, 2007. 1.2

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized algorithm for the maximal independent set problem. *Journal of Algorithms*, 7:567–583, 1986. 1

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost $k$-wise independence versus $k$-wise independence. *Inform. Process. Lett.*, 88(3):107–110, 2003. 1.2

[Bar02] Alexander Barvinok. *A course in convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002. 4

[Baz15] Louay Bazzi. Weight distribution of cosets of small codes with good dual properties. *IEEE Trans. Inform. Theory*, 61(12):6493–6504, 2015. 1.2

[BGGP12] Itai Benjamini, Ori Gurel-Gurevich, and Ron Peled. On k-wise independent distributions and boolean functions, 2012. 1.1, 2, 1.1, 1.1

[BHLV19] Ravi Boppana, Johan Håstad, Chin Ho Lee, and Emanuele Viola. Bounded independence versus symmetric tests. *ACM Trans. Comput. Theory*, 11(4):Art. 21, 27, 2019. 1.2, 16, 3, 22

[BR94] Mihir Bellare and John Rompel. Randomness-efficient oblivious sampling. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 276–287. IEEE, 1994. 1.1

[BS15] Mark Bun and Thomas Steinke. Weighted polynomial approximations: limits for learning and pseudorandomness. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 40 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 625–644. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2015. 1.1, 1.1, 1.2, 1.2, 2, 2, 14, 2, 2

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions (preliminary version). In *26th Symposium on Foundations of Computer Science*, pages 396–407, Portland, Oregon, 21–23 October 1985. IEEE. 1

[CW79] J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *J. of Computer and System Sciences*, 18(2):143–154, 1979. 1

[Das08] Anirban DasGupta. *Asymptotic theory of statistics and probability*. Springer Texts in Statistics. Springer, New York, 2008. 24

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM J. on Computing*, 39(8):3441–3462, 2010. 1.1, 2, 1.1

[DILV24] Harm Derksen, Peter Ivanov, Chin Ho Lee, and Emanuele Viola. Pseudorandomness, symmetry, smoothing: I. In *Conf. on Computational Complexity (CCC)*, 2024. 1.2, 1.2, 1.3, 1.3, 1.3, 1.3, 4, 21

[DKN10] Ilias Diakonikolas, Daniel M. Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010*, pages 11–20. IEEE Computer Soc., Los Alamitos, CA, 2010. 1.1, 2, 1.1

[DP09] Devdatt Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009. 1.1

[Erd16] Tamás Erdélyi. Coppersmith-Rivlin type inequalities and the order of vanishing of polynomials at 1. *Acta Arith.*, 172(3):271–284, 2016. 13

[HH23] Pooya Hatami and William Hoza. Theory of unconditional pseudorandom generators. *Electron. Colloquium Comput. Complex.*, TR23-019, 2023. 1, 1.3

[Kle20] Achim Klenke. *Probability theory—a comprehensive course*. Universitext. Springer, Cham, [2020] ©2020. Third edition [of 2372119]. 28

[LV17] Chin Ho Lee and Emanuele Viola. Some limitations of the sum of small-bias distributions. *Theory of Computing*, 13, 2017. 1.2, 1.2, 1.2, 1.2

[MS77] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. II.* North-Holland Mathematical Library, Vol. 16. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977. 4.1

[NN90] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. In *22nd ACM Symp. on the Theory of Computing (STOC)*, pages 213–223. ACM, 1990. 1

[OZ18] Ryan O'Donnell and Yu Zhao. On Closeness to $k$-Wise Uniformity. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 54:1–54:19, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 1.2

[Pet75] V. V. Petrov. *Sums of independent random variables*. Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], Band 82. Springer-Verlag, New York-Heidelberg, 1975. Translated from the Russian by A. A. Brown. 25, 6.3

[RR47] C. Radhakrishna Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *Suppl. J. Roy. Statist. Soc.*, 9:128–139, 1947. 1

[Sko22] Maciej Skorski. Tight Chernoff-like bounds under limited independence. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 245 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 15, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2022. 1.1

[SSS95] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff-Hoeffding bounds for applications with limited independence. *SIAM J. Discrete Math.*, 8(2):223–250, 1995. 1.1, 1.1

[SV13] Sushant Sachdeva and Nisheeth K. Vishnoi. Faster algorithms via approximation theory. *Found. Trends Theor. Comput. Sci.*, 9(2):125–213, 2013. 2

[Vio23] Emanuele Viola. Mathematics of the impossible: The uncharted complexity of computation. 2023. 1.1

[Wik] Wikipedia contributors. Vandermonde matrix. https://en.wikipedia.org/wiki/Vandermonde_matrix. accessed July 9, 2024. 6.3