Bounded Independence versus Symmetric Tests[*]

Ravi Boppana[†]   Johan Håstad[‡]   Chin Ho Lee[§]   Emanuele Viola[§]

April 4, 2019

Abstract

For a test $T \subseteq \{0,1\}^n$ define $k^*(T)$ to be the maximum k such that there exists a k-wise uniform distribution over $\{0,1\}^n$ whose support is a subset of T.

For $H_t = \{x \in \{0,1\}^n : |\sum_i x_i - n/2| \le t\}$ we prove $k^*(H_t) = \Theta(t^2/n + 1)$.

For $S_{m,c} = \{x \in \{0,1\}^n : \sum_i x_i \equiv c \pmod{m}\}$ we prove that $k^*(S_{m,c}) = \Theta(n/m^2)$. For some $k = O(n/m)$ we also show that any k-wise uniform distribution puts probability mass at most $1/m + 1/100$ over $S_{m,c}$. Finally, for any fixed odd m we show that there is an integer $k = (1 - \Omega(1))n$ such that any k-wise uniform distribution lands in T with probability exponentially close to $|S_{m,c}|/2^n$; and this result is false for any even m.

1 Introduction and our results

A distribution on $\{0,1\}^n$ is *k-wise uniform*, aka *k-wise independent*, if any k bits are uniform in $\{0,1\}^k$. The study of k-wise uniformity has been central to theoretical computer science since at least the seminal work [CW79] by Carter and Wegman. A specific direction has been to show that k-wise uniformity "looks random" to several classes of tests. These classes include combinatorial rectangles [EGL$^+$92, CRS00] (for an exposition see e.g. Lecture 1 in [Vio17]), bounded-depth circuits, aka AC^0, [Baz09, Raz09, Bra10, Tal17, HS16] (see e.g. Lectures 2-3 in [Vio17]), and halfspaces [RS10, DGJ$^+$10, GOWZ10, DKN10]. More recently a series of works consider smoothed versions of the first two classes, where the input is perturbed with noise, and gives improved bounds [AW89, GMR$^+$12, HLV18, LV17a]. These results have in turn found many applications. For example, the recent exciting constructions of 2-source extractors for polylogarithmic min-entropy [CZ16, BDT16] rely on the results that k-wise uniformity looks random to AC^0 [Bra10] and the majority function [DGJ$^+$10].

In this work we extend this direction by giving new bounds for two classes of tests, both symmetric. First we consider the class of mod m tests.

Definition 1. *For an input length n, and integers m and c, we define the set $S_{m,c} := \{x \in \{0,1\}^n : \sum_i x_i \equiv c \pmod{m}\}$.*

These tests have been extensively studied at least since circuit complexity theory "hit the wall" of circuits with mod m gates for composite m in the 80's. However, the effect of k-wise uniformity on mod m tests does not seem to have been known before this paper. We study for what values of k there exist a k-wise uniform distribution over $\{0,1\}^n$ *supported* on $S_{m,c}$. Note that such k-wise uniform distributions are obviously not pseudorandom against mod m tests. Our first main result gives tight bounds on the maximum value of such a k, establishing k is about n/m^2.

Theorem 2. *For all integers $m \geq 2$ and c, there exists an integer $k \geq n/(32m^2) - 1$ and a k-wise uniform distribution on $\{0,1\}^n$ that is supported on $S_{m,c}$.*

Theorem 3. *For all integers $m \geq 2$, c, and $k \geq 140n/m^2 + 4$, no k-wise uniform distribution on $\{0,1\}^n$ can be supported on $S_{m,c}$.*

Theorem 2 is trivial for $m = 2$, as the uniform distribution over $S_{2,0}$ is $(n-1)$-wise uniform. But already for $m = 3$ the result is not trivial.

Theorem 3 is equivalent to saying that when $k \geq 140n/m^2 + 4$ then every k-wise uniform distribution must land in $\overline{S_{m,c}}$ with non-zero probability.

For motivation, recall from above the line of works [Baz09, Raz09, Bra10, Tal17, HS16] showing that k-wise uniform distributions fool AC^0 circuits. Specifically, these works show $k = \text{poly} \log n$ suffices to fool AC^0 circuits on n bits of size poly(n) and depth $O(1)$. It is natural to ask whether the same distribution also fools AC^0 circuits with mod m gates, a "frontier" class for which we have exponential lower bounds [Raz87, Smo87] (when m is prime) but not good pseudorandom generators. A positive answer might have looked plausible, given that for example the parity function is hard even with mod 3 gates [Smo87]. But in fact Theorem 2 gives a strong negative answer, showing that $k = \Omega(n)$ is necessary even for a single mod 3 gate.

Theorem 2 proves a conjecture in [LV17b] where this question is also raised. Their motivation was a study of the "mod 3" rank of k-wise uniform distributions, started in [MZ09], which is the dimension of the space spanned by the support of the distribution over $\mathbb{F}_3$. The work of Lee and Viola [LV17b, Claim 1.10] shows that achieving $100 \log n$-wise uniformity with dimension $\leq n^{0.49}$ would have applications to pseudorandomness. It also exhibits a distribution with dimension $n^{0.72}$ and pairwise uniformity. Theorem 2 yields a new piece of information in this study: there exist distributions with dimension $n-1$ and $\Omega(n)$-wise uniformity. However, we are still far from achieving the parameters required in [LV17b, Claim 1.10].

We then prove another theorem which is in the same spirit of Theorem 3 but gives different information.

Theorem 4. *Let m and c be two integers.*

(a) *For $k \geq 2n/m+2$, no k-wise uniform distribution on $\{0,1\}^n$ can be supported on $S_{m,c}$.*

(b) *If m is odd, then there is a $\gamma > 0$ depending only on m such that for any $(1-\gamma)n$-wise uniform distribution D on $\{0,1\}^n$, we have $|\Pr[D \in S_{m,c}] - |S_{m,c}|/2^n| \leq 2^{-\gamma n}$.*

(c) *There exists a universal constant C such that for every $\varepsilon > 0$, $n \geq Cm^2 \log m$, and any $C(n/m)(1/\varepsilon)^2$-wise uniform distribution D on $\{0,1\}^n$, $\Pr[D \in S_{m,c}] \leq |S_{m,c}|/2^n + \varepsilon$.*

First, Theorem 4 (a) shows that the largest possible value of k in Theorem 2 is $k \leq 2(n+1)/m+2$. Compared to Theorem 3, this result is asymptotically inferior, but gives better constants and has a simpler proof. Theorem 4 (b) shows that when m is odd, if k is larger than $(1-\gamma)n$ for a positive constant γ depending only on m then k-wise uniformity fools $S_{m,c}$ with exponentially small error. The proof of Theorem 4 (b) however does not carry to the setting of $k < n/2$, for any m. So we establish Theorem 4 (c), which gives a worse error bound but allows for k to become smaller for larger m, specifically, $k = O(n/m)$ for constant error. The error bound in Theorem 4 (c) and the density of $S_{m,c}$ are such that it only provides a meaningful upper bound on the probability that the k-wise uniform distribution lands in $S_{m,c}$, but not a lower bound. In fact, we conjecture that no lower bound is possible in the sense that there is a constant $C > 0$ such that for every m there is a Cn-wise uniform distribution supported on the complement of $S_{m,c}$.

We note that Theorem 4 (b) is false when m is even because the uniform distribution on $S_{2,0}$ has uniformity $k = n-1$ but puts about $2/m$ mass on $S_{m,0}$, a set which as we shall see later (cf. Remark 1) has density about $1/m$. The latter density bound, in combination with Theorem 4 (b) and Theorem 4 (c) implies that for some $k = \min\{O(n/m), (1-\Omega(1))n\}$, every k-wise uniform distribution puts probability mass at most $1/m + 1/100$ over $S_{m,c}$ for odd m and any integer c.

A remaining open problem is to understand for what values of k does k-wise uniformity ensure closeness to uniform modulo m. Note that to get close to uniform, or even to put probability mass on all possible moduli, one would want ε in Theorem 4 (c) to be roughly $1/m$. However, this gives $k \geq n$ in the theorem, which is trivial.

We then consider another class of tests which can be written as the intersection of two halfspaces.

Definition 5. *For an input length n, and an integer t, we define the set $H_t := \{x \in \{0,1\}^n : |\sum_i x_i - n/2| \leq t\}$.*

2

Again, we ask for what values of k there exists a k-wise uniform distribution over $\{0,1\}^n$ supported on H_t. We obtain tight bounds for k up to a constant factor, showing that the maximum value of such a k is $\Theta(t^2/n + 1)$.

Theorem 6. *For every integer t, there exists an integer $k \geq t^2/(50n)$ and a k-wise uniform distribution over $\{0,1\}^n$ that is supported on H_t.*

Theorem 7. *For all integers t and $k \geq 36t^2/n + 3$, no k-wise uniform distribution over $\{0,1\}^n$ can be supported on H_t.*

In fact, as we will see in the proof, the distribution in Theorem 6 is actually supported on $H_t \cap S_{m,0}$.

One motivation for these results is to understand for which tests the smoothed version of the test obtained by perturbing coordinates with random noise is fooled by k-wise uniformity. As mentioned earlier, this understanding underlies recent, state-of-the-art pseudorandom generators [AW89, GMR+12, HLV18, LV17a]. See also [LV17b]. Using Theorem 6 we prove that $\Omega(\log n)$-wise uniformity is necessary to fool read-once DNF on n bits, even with constant noise. Note that $O(\log n)$-uniformity is sufficient, even without noise [EGL+92].

Theorem 8. *There exists a read-once DNF $d : \{0,1\}^n \to \{0,1\}$, a constant α, and an $\alpha \log n$-wise uniform distribution D such that $|\Pr[d(U) = 1] - \Pr[d(D + N_\alpha) = 1]| \geq \Omega(1)$, where U is uniform over $\{0,1\}^n$, N_α is the distribution over $\{0,1\}^n$ whose bits are independent and are set to uniform with probability α and 0 otherwise, and '$+$' is bit-wise XOR.*

Proof. Let d be the Tribes DNF with width $w = \log n - \log \ln n + o_n(1)$, see e.g. [O'D14]. We have $|\Pr[d(U) = 1] - 1/2| = o(1)$. Partition the n bits into n/w blocks of size w. The distribution D has i.i.d. blocks. The projection of each block is an αw-wise uniform distribution with Hamming weight $\leq 2w/3$. The probability that $d(D + N_\alpha) = 1$ is at most the probability that there exists a block where the noise vector N_α has Hamming weight $\geq w/3$. This probability is at most

$$(n/w)2^w(\alpha/2)^{w/3} \leq 1/3,$$

for a sufficiently small α. $\qquad\square$

1.1 Techniques

We give two related approaches to proving Theorem 2. At a high level, both approaches are similar to the work of Alon, Goldreich, and Mansour [AGM03], which shows that one can apply a small perturbation to the probability masses of every almost k-wise uniform distribution on $\{0,1\}^n$ to make it k-wise uniform, showing that every ε-almost k-wise uniform distribution on $\{0,1\}^n$ is $n^{O(k)}\varepsilon$-close to a k-wise uniform distribution, in statistical distance. However, in their setting there is no constraint on the support. This makes our proof significantly more technical.

Our first approach uses the following equivalent definition for a distribution on $\{0,1\}^n$ to be k-wise uniform: the distribution is unbiased under any parity test on at most k bits. To construct our distribution, we first start with the uniform distribution over the set $S_{m,c}$,

and show that the bias under each of these parity tests is small enough, so that they can be made zero by a small perturbation of the probability masses of the distribution. Our goal is then to show that the change in the probability on each weight is no more than the probability we start with, so that it remains non-negative after the perturbation. In the conference version of this paper [BHLV16], we use this approach to prove a slightly weaker version of Theorem 2. We refer the interested readers to [BHLV16] for details.

We now give an overview of the second approach, which is developed in this paper. Instead of looking at the biases of parity tests, we consider another equivalent characterization of k-wise uniform distributions that are *symmetric*. To simplify the calculations, we will switch to $\{-1, 1\}$ and consider distributions supported on $S'_{m,c} := \{y \in \{-1, 1\}^n : \sum_i y_i \equiv c \pmod{m}\}$. One can then translate results for $\{-1, 1\}^n$ back to $\{0, 1\}^n$. (See Fact 12.) A symmetric distribution is k-wise uniform on $\{-1, 1\}^n$ if and only if the first k moments of the sum of its n bits match the corresponding moments of the uniform bits. Similar to the first approach, we start with the uniform distribution on $S'_{m,c}$, and show that the first k moments of the sum of the bits are close to the ones of the uniform bits. Then, we perturb the probabilities on $k + 1$ of the sums $\sum_i y_i$ of the distribution to match these moments exactly. Once again, our goal is to show that the amount of correction is small enough so that the adjusted probabilities remain non-negative. Note that in this approach we work with distributions over the integers instead of $\{0, 1\}^n$.

While the two approaches seem similar to each other, the latter allows us to perform a more refined analysis on the tests we consider in this paper, and obtain the tight lower bounds for both modular and threshold tests.

Organization. We begin with Theorem 4 in Section 2 because it is simpler. We use the second approach to prove the tight lower bounds for $S_{m,c}$ (Theorem 2) and H_t (Theorem 6) in Sections 3 and 4, respectively. The proof of Theorem 2 involves a somewhat technical lemma which we defer to Section 5. Finally, we prove our tight upper bounds for $S_{m,c}$ (Theorem 3) and H_t (Theorem 7) in Sections 6 and 7, respectively.

2   Proof of Theorem 4

In this section we prove Theorem 4. We start with the following theorem which will give Theorem 4 (a) as a corollary.

Theorem 9. *Let $I \subseteq \{0, 1, \ldots, n\}$ be a subset of size $|I| \le n/2$. There does not exist a $2|I|$-wise uniform distribution on $\{0, 1\}^n$ that is supported on $S := \{x \in \{0, 1\}^n : \sum_i x_i \in I\}$.*

Proof. Suppose there exists such a distribution D. Define the n-variate nonzero real polynomial p by

$$p(x) := \prod_{i \in I} \left(-i + \sum_{j=1}^n x_j \right).$$

Note that $p(x) = 0$ when $x \in S$. And so $\mathbb{E}[p^2(D)] = 0$ in particular. However, since p^2 has degree at most $2|I|$, we have $\mathbb{E}[p^2(D)] = \mathbb{E}[p^2(U)] > 0$, where U is the uniform distribution over $\{0, 1\}^n$, a contradiction. $\qquad \square$

4

Proof of Theorem 4 (a). When I corresponds to the mod m test $S_{m,c}$, $|I| \leq n/m + 1$. $\square$

We now move to Theorem 4 (b). First we prove a lemma giving a useful estimate of

$$\sum_{x \in S_{m,c}} (-1)^{\sum_{i=1}^{k} x_i}.$$

Similar bounds have been established elsewhere, cf. e.g. Theorem 2.9 in [VW08], but we do not know of a reference with an explicit dependence on m, which will be used in the next section. Theorem 4 (b) follows from bounding above the tail of the Fourier coefficients of the indicator function of $S_{m,c}$.

Lemma 10. *For any $1 \leq k \leq n - 1$ and any $0 \leq c \leq m - 1$, we have*

$$\left| \sum_{x \in S_{m,c}} (-1)^{\sum_{i=1}^{k} x_i} \right| \leq 2^n \left(\cos \frac{\pi}{2m} \right)^n,$$

while for $k = 0$, we have

$$\left| |S_{m,c}| - 2^n/m \right| \leq 2^n \left(\cos \frac{\pi}{2m} \right)^n.$$

For odd m the first bound also holds for $k = n$.

Proof. Consider an expansion of

$$p(y) = (1 - y)^k (1 + y)^{n-k}$$

into 2^n terms indexed by $x \in \{0, 1\}^n$ where $x_i = 0$ indicates that we take the term 1 from the ith factor. It is easy to see that the coefficient of y^d is $\sum_{|x|=d}(-1)^{\sum_{i=1}^{k} x_i}$, where $|x|$ denotes the number of occurrences of 1 in x. Denote $\zeta := e^{2\pi i/m}$ as an mth root of unity. Recall the identity

$$\frac{1}{m} \sum_{j=0}^{m-1} \zeta^{j(d-c)} = \begin{cases} 1 & \text{if } d \equiv c \pmod{m} \\ 0 & \text{otherwise.} \end{cases}$$

Thus the sum we want to bound is equal to

$$\frac{1}{m} \sum_{j=0}^{m-1} \zeta^{-jc} p(\zeta^j).$$

Note that $p(\zeta^0) = p(1) = 0$ for $k \neq 0$ while for $k = 0$, $p(\zeta^0) = 2^n$. For the other terms we have the following bound.

Claim 11. *For $1 \leq j \leq m - 1$, $|p(\zeta^j)| \leq 2^n \left(\cos \frac{\pi}{2m} \right)^k \left(\cos \frac{\pi}{m} \right)^{n-k}$.*

Proof. As $|1 + e^{i\theta}| = 2|\cos(\theta/2)|$ and $|1 - e^{i\theta}| = 2|\sin(\theta/2)|$ we have

$$|p(\zeta^j)| = |1 - \zeta^j|^k |1 + \zeta^j|^{n-k}$$

$$= 2^n \left| \sin \frac{j\pi}{m} \right|^k \left| \cos \frac{j\pi}{m} \right|^{n-k}$$

$$\leq 2^n \left(\cos \frac{\pi}{2m} \right)^k \left(\cos \frac{\pi}{m} \right)^{n-k},$$

5

where the last inequality holds for odd m because (1) $\sin \frac{j\pi}{m}$ is largest when $j = \frac{m-1}{2}$ or $j = \frac{m+1}{2}$, (2) $\sin(\frac{\pi}{2} - x) = \cos x$, and (3) $\cos \frac{j\pi}{m}$ is largest when $j = 1$ or $j = m - 1$. For even m the term with $j = m/2$ is 0, as in this case we are assuming that $k < n$, and the bounds for odd m are valid for the other terms. $\qquad \square$

Therefore, for $k \neq 0$ we have

$$\left| \sum_{x \in S_{m,c}} (-1)^{\sum_{i=1}^{k} x_i} \right| = \frac{m-1}{m} \cdot 2^n \left(\cos \frac{\pi}{2m} \right)^k \left(\cos \frac{\pi}{m} \right)^{n-k} \leq 2^n \left(\cos \frac{\pi}{2m} \right)^k \left(\cos \frac{\pi}{m} \right)^{n-k},$$

and we complete the proof using the fact that $\cos(\pi/m) \leq \cos(\pi/2m)$. For $k = 0$ we also need to include the term $p(1) = 2^n$ which divided by m gives the term $2^n/m$. $\qquad \square$

Remark 1. *Clearly the lemma for $k = 0$ simply is the well-known fact that the cardinality of $S_{m,c}$ is very close to $2^n/m$. Equivalently, if x is uniform in $\{0,1\}^n$ then the probability that $\sum_i x_i \equiv c \pmod{m}$ is very close to $1/m$.*

Proof of Theorem 4 (b). We use Fourier analysis (see e.g. [O'D14] for an introduction). Let $f \colon \{0,1\}^n \to \{0,1\}$ be the characteristic function of $S_{m,c}$. We write f in its Fourier expansion $f(x) = \sum_{\beta \in \{0,1\}^n} \hat{f}_\beta (-1)^{\sum_{i=1}^{n} \beta_i x_i}$. We first bound above the nonzero Fourier coefficients of f. Since f is symmetric, by Lemma 10, we have for any β with $|\beta| = k > 0$,

$$|\hat{f}_\beta| = \left| 2^{-n} \sum_{x \in S_{m,c}} (-1)^{\sum_{i=1}^{k} x_i} \right| \leq \left(\cos \frac{\pi}{2m} \right)^n \leq 2^{-\alpha n},$$

where $\alpha = -\log_2 \cos(\pi/2m) = \Theta(1/m^2)$ depends only on m. If D is k-wise uniform, we have $\mathbb{E}_{x \sim D}[(-1)^{\sum \beta_i x_i}] = 0$ whenever $0 < |\beta| \leq d$. Therefore,

$$|\mathbb{E}[f(D)] - \mathbb{E}[f(U)]| \leq \sum_{|\beta| > k} |\hat{f}_\beta| \cdot \left| \mathbb{E}_{x \sim D}\left[(-1)^{\sum \beta_i x_i} \right] \right| \leq \sum_{|\beta| > k} |\hat{f}_\beta| \leq 2^{-\alpha n} \sum_{t=k+1}^{n} \binom{n}{t} = 2^{-\alpha n} \sum_{t=0}^{n-k-1} \binom{n}{t}.$$

For $k \geq (1 - \delta)n$, we have an upper bound of $2^{n(H(\delta) - \alpha)}$. Pick δ small enough so that $H(\delta) \leq \alpha/2$. The result follows by setting $\gamma := \min\{\alpha/2, \delta\}$. $\qquad \square$

Note that the above proof fails when m is even as we cannot handle the term with $|\beta| = n$. Finally, we prove Theorem 4 (c). We use approximation theory.

Proof of Theorem 4 (c). We will show that $\Pr[D \in S_{m,c}] \leq \varepsilon$. Let $f \colon \{0,1\}^n \to \{0,1\}$ be the characteristic function of $S_{m,c}$. The proof amounts to exhibiting a real polynomial p in n variables of degree $d = C(n/m)(1/\varepsilon)^2$ such that $f(x) \leq p(x)$ for every $x \in \{0,1\}^n$, and $\mathbb{E}[p(U)] \leq \varepsilon$ for U uniform over $\{0,1\}^n$. To see that this suffices, note that $\mathbb{E}[p(U)] = \mathbb{E}[p(D)]$ for any distribution D that is d-wise uniform. Using this and the fact that f is non-negative, we can write

$$0 \leq \mathbb{E}[f(D)] \leq \mathbb{E}[p(D)] = \mathbb{E}[p(U)] \leq \varepsilon.$$

Hence, $\mathbb{E}[f(D)] \leq \varepsilon$. This is the method of sandwiching polynomials from [Baz09].

6

Let us write $f = g(\sum_i x_i/n)$, for $g \colon \{0, 1/n, \ldots, 1\} \to \{0, 1\}$. We exhibit a univariate polynomial q of degree d such that $g(x) \leq q(x)$ for every x, and the expectation of q under the binomial distribution is at most ε. The polynomial p is then $q(\sum_i x_i/n)$.

Consider the continuous, piecewise linear function $s \colon [-1, 1] \to [0, 1]$ defined as follows. The function is always 0, except at intervals of radius a/n around the inputs x where g equals 1, i.e., inputs x such that nx is congruent to c modulo m. In those intervals we define the following tent function: it is a piecewise linear function that goes up and down like a '$\wedge$', starting from 0 at $x - a/n$, and reaching the value of 1 linearly at x, then going back to 0 at $x + a/n$. We set $a = \varepsilon m/10$.

By Jackson's theorem (see e.g., [Car, Theorem 7.4] or [Che66, Chapter 6]), for $d = O(n\varepsilon^{-1}a^{-1}) = O(n\varepsilon^{-2}m^{-1})$, there exists a univariate polynomial q' of degree d that approximates s with pointwise error $\varepsilon/10$. Our polynomial q is defined as $q := q' + \varepsilon/10$.

It is clear that $g(x) \leq q(x)$ for every $x \in \{0, 1/n, \ldots, 1\}$. It remains to estimate $\mathbb{E}[q(U)]$.

As q' is a good approximation of s we have $\mathbb{E}[q(U)] \leq 2\varepsilon/10 + \mathbb{E}[s(U)]$. We noted in Remark 1 that the remainder modulo m of $\sum U_i$ is δ-close to uniform over $\{0, 1, \ldots, m-1\}$ for $\delta = \cos(\pi/2m)^n = e^{-\Omega(n/m^2)}$. Now the function s, as a function of $\sum x_i$, is a periodic function with period m. So we have

$$\mathbb{E}[s(U)] = \sum_{x=0}^{a} \Pr\left[\sum_i U_i \equiv x \pmod{m}\right] \cdot s(x/n)$$

$$\leq (1/m + \delta) \sum_{x=0}^{a} \frac{x}{a}$$

$$\leq (1/m + \delta) \cdot a$$

$$\leq (1/m + \delta) \cdot \varepsilon m/10.$$

It follows that if n is at least a large constant times $m^2 \log m$, we have $\mathbb{E}[s(U)] \leq 2\varepsilon/10$ and we conclude that $\mathbb{E}[q(U)] \leq 4\varepsilon/10$. $\qquad\square$

3   Tight lower bound on k-wise uniformity vs. mod m

In this section we prove Theorem 2. Throughout the rest of this paper, we will interpret 0^0 as 1. For convenience, from now on we will consider the space $\{-1, 1\}^n$ instead of $\{0, 1\}^n$. In particular, we will consider strings $x \in \{-1, 1\}^n$ that satisfy $\sum_i x_i \equiv c \pmod{m}$. One can translate results stated for $\{0, 1\}^n$ to results for $\{-1, 1\}^n$ and vice versa using the following fact. Recall that $S_{m,c} = \{x \in \{0, 1\}^n : \sum_i x_i \equiv c \pmod{m}\}$ and $S'_{m,c} = \{y \in \{-1, 1\}^n : \sum_i y_i \equiv c \pmod{m}\}$.

Fact 12. *Let $x \in \{0, 1\}^n$ and let $y \in \{-1, 1\}^n$ be the string obtained by replacing each x_i by $y_i = 1 - 2x_i$. Then x is in $S_{m,c}$ if and only if y is in $S'_{2m,n-2c}$.*

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be independent random variables chosen uniformly from $\{-1, 1\}$. Let B be the sum of all the X_j. The distribution of B is a shifted binomial distribution. Note that B has the same parity as n.

Theorem 13. *Let m and n be positive integers and c be an integer. Suppose that m is odd or n and c have the same parity. Let k be a positive integer such that $k \leq \frac{n}{8m^2}$. Then there is a probability distribution on the $c \bmod m$ integers that matches the first k moments of B. Furthermore, the support of the probability distribution is a subset of the support of B.*

Theorem 2 follows from applying Fact 12 to Theorem 13.

Our goal is to come up with a distribution supported on $c \bmod m$ so that its first k moments match the moments of B. We first start with a measure q on the $c \bmod m$ integers. Here q may not be a probability measure — its values may not sum to 1. However, we will show that we can turn q into a probability measure p by a small adjustment Δ on $k + 1$ appropriately chosen positive values of $q(x)$.

3.1   Defining $C_{m,c}(x)$

Let m be a positive integer (the modulus). Let c be an integer (the residue). We will assume that either m is odd or n and c have the same parity. We will use Iverson bracket notation: $[\![\text{true}]\!] = 1$ and $[\![\text{false}]\!] = 0$. Define the comb function $C_{m,c}$ on the integers by $C_{m,c}(x) = m[\![x \equiv c \pmod{m}]\!]$ if m is odd and $C_{m,c}(x) = \frac{m}{2}[\![x \equiv c \pmod{m}]\!]$ if m is even.

3.2   Defining $q(x)$

Define the function q on the integers by $q(x) = C_{m,c}(x)\Pr[B = x]$.[1] Note that q is nonnegative. Also if $q(x) \neq 0$, then x is $c \pmod{m}$ and in the support of B.

Lemma 14. *If f is a function on the integers, then*

$$\sum_x q(x)f(x) = \mathbb{E}[C_{m,c}(B)f(B)].$$

Proof. By the definition of expected value, we have

$$\sum_x q(x)f(x) = \sum_x \Pr[B = x]C_{m,c}(x)f(x) = \mathbb{E}[C_{m,c}(B)f(B)]. \qquad \square$$

3.3   Defining Lagrange polynomials

Let k be a positive integer. Let a_0, a_1, ..., a_k be $k + 1$ distinct integers that are $c \bmod m$, $n \bmod 2$, and as close to 0 as possible. Because they are as close to 0 as possible, we have $|a_j| \leq (k+1)m \leq 2km$. In our application, $2km$ will be at most n. So each a_j will be in the support of B.

Given an integer v such that $0 \leq v \leq k$, define the Lagrange polynomial L_v as follows:

$$L_v(x) = \prod_{\substack{0 \leq j \leq k \\ j \neq v}} (x - a_j).$$

Note that $L_v(a_w) = 0$ if and only if $v \neq w$. It's well known that $L_0, L_1, \ldots, L_k$ form a basis (the Lagrange basis) of the vector space of polynomials of degree at most k.

---

[1] Alternatively, we could have defined q to be the probability distribution of B conditioned on $B \equiv c \pmod{m}$. While that definition would work, it would make our argument slightly more complicated.

3.4 Defining $\Delta(x)$

Define the function Δ on the integers as follows. If x equals a_v (for some v), then

$$\Delta(a_v) = \frac{\mathbb{E}[C_{m,c}(B)L_v(B)] - \mathbb{E}[L_v(B)]}{L_v(a_v)}.$$

If $x \neq a_w$ for every w, then $\Delta(x) = 0$.[2]

Lemma 15. *If f is a polynomial of degree at most k, then*

$$\sum_x \Delta(x)f(x) = \mathbb{E}[C_{m,c}(B)f(B)] - \mathbb{E}[f(B)].$$

Proof. We will first prove the claim when f is a Lagrange polynomial L_v. If $\Delta(x) \neq 0$, then x is of the form a_w for some w. But if $L_v(a_w) \neq 0$, then $v = w$. So the sum has at most one nonzero term, corresponding to $x = a_v$. And the equation is true in this case by the definition of Δ.

We have proved the claim for Lagrange polynomials. But every polynomial of degree at most k is a linear combination of the Lagrange polynomials. This completes the proof. $\square$

3.5 Defining $p(x)$

Define the function p on the integers by $p(x) = q(x) - \Delta(x)$. Note that if $p(x) \neq 0$, then x is $c \bmod m$ and (assuming $2km \leq n$) in the support of B.

Lemma 16. *If f is a polynomial of degree at most k, then*

$$\sum_x p(x)f(x) = \mathbb{E}[f(B)].$$

Proof. By Lemmas 14 and 15, we have

$$\sum_x p(x)f(x) = \sum_x q(x)f(x) - \sum_x \Delta(x)f(x)$$
$$= \mathbb{E}[C_{m,c}(B)f(B)] - \Big(\mathbb{E}[C_{m,c}(B)f(B)] - \mathbb{E}[f(B)]\Big)$$
$$= \mathbb{E}[f(B)]. \qquad \square$$

To show that p is nonnegative, we will show that $|\Delta(x)| \leq \frac{1}{2}q(x)$. First we bound above $|\Delta(x)|$. Then we bound below $q(x)$.

Recall that $\Delta(x) = \frac{\mathbb{E}[C_{m,c}(B)L_v(B)] - \mathbb{E}[L_v(B)]}{L_v(a_v)}$ if x equals a_v for some v (and 0 otherwise). Lemmas 19 and 21 below give upper and lower bounds on the numerator and denominator, respectively.

---

[2]Why this definition of Δ? We want Δ to be nonzero only at the $k + 1$ values $a_0, a_1, \ldots, a_k$. We also want Lemma 15 to hold. Those two constraints force the given definition of Δ.

3.6 Bounding numerator of Δ

We start this subsection by stating a useful lemma.

Lemma 17. *Let r be an integer such that $0 \leq r \leq \frac{n}{8}$. If $0 < \theta \leq \frac{\pi}{2}$, then*

$$\left|\mathbb{E}[B^r e^{i\theta B}]\right| \leq 2\left(\frac{8r}{\theta}\right)^r e^{-\theta^2 n/4}.$$

Informally, this lemma says that the rth moment of B, after being twisted in the complex plane, is exponentially small. We defer its proof to Section 5.

Our next lemma says that B is exponentially close to being uniformly distributed mod m even when weighted by moments.

Lemma 18. *Let r be an integer such that $0 \leq r \leq \frac{n}{8}$. Let m be an integer such that $1 \leq m \leq \sqrt{n}$. Suppose that m is odd or n and c have the same parity. Then*

$$\left|\mathbb{E}[B^r C_{m,c}(B)] - \mathbb{E}[B^r]\right| \leq 8(3rm)^r e^{-2n/m^2}.$$

Proof. Let m' be m if m is odd and $m/2$ if m is even. Let α be π/m'. For now, we will assume that n and c have the same parity. At the end, we will show how to adjust the proof when n and c have the opposite parity.

Because B and n have the same parity, $B - c$ is even. So we have the key identity

$$\sum_{j=1}^{m'-1} e^{ij\alpha(B-c)} = C_{m,c}(B) - 1.$$

Hence, by the triangle inequality, we have

$$\left|\mathbb{E}[B^r C_{m,c}(B)] - \mathbb{E}[B^r]\right| = \left|\mathbb{E}\left[B^r \sum_{j=1}^{m'-1} e^{ij\alpha(B-c)}\right]\right|$$

$$= \left|\sum_{j=1}^{m'-1} e^{-ij\alpha c} \mathbb{E}[B^r e^{ij\alpha B}]\right|$$

$$\leq \sum_{j=1}^{m'-1} \left|\mathbb{E}[B^r e^{ij\alpha B}]\right|.$$

The sum is symmetric: the terms corresponding to $j = \ell$ and $j = m' - \ell$ are equal. So we can double its first half:

$$\left|\mathbb{E}[B^r C_{m,c}(B)] - \mathbb{E}[B^r]\right| \leq 2 \sum_{j=1}^{\lfloor m'/2 \rfloor} \left|\mathbb{E}[B^r e^{ij\alpha B}]\right|.$$

10

Therefore, by Lemma 17, we have

$$\left| \mathbb{E}[B^r C_{m,c}(B)] - \mathbb{E}[B^r] \right| \le 4 \sum_{j=1}^{\lfloor m'/2 \rfloor} \left(\frac{8r}{j\alpha} \right)^r e^{-j^2 \alpha^2 n/4}$$

$$\le 4 \left(\frac{8r}{\alpha} \right)^r \sum_{j=1}^{\lfloor m'/2 \rfloor} e^{-j^2 \alpha^2 n/4}$$

$$\le 4 \left(\frac{8rm}{\pi} \right)^r \sum_{j=1}^{\lfloor m'/2 \rfloor} e^{-j^2 \pi^2 n/(4m^2)}$$

$$\le 4(3rm)^r \sum_{j=1}^{\lfloor m'/2 \rfloor} e^{-2j^2 n/m^2}$$

$$\le 4(3rm)^r \sum_{j=1}^{\lfloor m'/2 \rfloor} e^{-2jn/m^2}.$$

The sum is a geometric series whose common ratio is less than $\frac{1}{2}$, so we can bound it by twice its first term:

$$\left| \mathbb{E}[B^r C_{m,c}(B)] - \mathbb{E}[B^r] \right| \le 8(3rm)^r e^{-2n/m^2}.$$

The proof above assumed that n and c have the same parity. When n and c have the opposite parity, we can adjust the proof as follows. From the parity hypothesis in the theorem, we know that m is odd. In particular, $m' = m$. Because B and n have the same parity, $B - c$ is odd. So we have the identity

$$\sum_{j=1}^{m'-1} (-1)^j e^{ij\alpha(B-c)} = C_{m,c}(B) - 1.$$

It's the same identity as before except for the factor of $(-1)^j$. We can now continue with the remainder of the proof. The factor of $(-1)^j$ goes away as soon as we apply the triangle inequality. Hence we obtain the same bound. $\square$

Our final lemma of this subsection says that B is exponentially close to being uniformly distributed mod m even when weighted by Lagrange polynomials. It provides the upper bound on the numerator of Δ that we were seeking.

Lemma 19. *Let m be an integer such that $1 \le m \le \sqrt{n}$. Suppose that m is odd or n and c have the same parity. Suppose that $k \le \frac{n}{8}$. If v is an integer such that $0 \le v \le k$, then*

$$\left| \mathbb{E}[L_v(B) C_{m,c}(B)] - \mathbb{E}[L_v(B)] \right| \le 8(5km)^k e^{-2n/m^2}.$$

Proof. Given a subset A of $\{1, 2, \ldots, n\}$, define the monomial X^A to be the product of the X_j for which j is in A. By expanding the product, we have

$$L_v(B) = \prod_{j \ne v} (B - a_j) = \sum_A (-1)^{|A|} a^A B^{k-|A|},$$

11

where A ranges over every subset of $\{0, 1, \ldots, k\} - \{v\}$. Therefore, by the triangle inequality, Lemma 18, and the binomial theorem, we have

$$
\begin{aligned}
|\mathbb{E}[L_v(B)C_{m,c}(B)] - \mathbb{E}[L_v(B)]| &= \left| \mathbb{E}\left[[C_{m,c}(B) - 1]L_v(B) \right] \right| \\
&= \left| \mathbb{E}\left[[C_{m,c}(B) - 1] \sum_A (-1)^{|A|} a^A B^{k-|A|} \right] \right| \\
&= \left| \sum_A (-1)^{|A|} a^A \, \mathbb{E}\left[[C_{m,c}(B) - 1] B^{k-|A|} \right] \right| \\
&\leq \sum_A \left| a^A \, \mathbb{E}\left[[C_{m,c}(B) - 1] B^{k-|A|} \right] \right| \\
&\leq \sum_A (2km)^{|A|} \left| \mathbb{E}\left[[C_{m,c}(B) - 1] B^{k-|A|} \right] \right| \\
&\leq \sum_A (2km)^{|A|} \cdot 8(3(k - |A|)m)^{k-|A|} e^{-2n/m^2} \\
&= 8e^{-2n/m^2} \sum_A (2km)^{|A|} (3(k - |A|)m)^{k-|A|} \\
&\leq 8e^{-2n/m^2} \sum_A (2km)^{|A|} (3km)^{k-|A|} \\
&= 8e^{-2n/m^2} (5km)^k. \qquad \square
\end{aligned}
$$

3.7   Bounding denominator of Δ

Our lower bound on $|L_v(a_v)|$ follows from the following claim.

Claim 20. *Let $a_0 < \cdots < a_k$ be $k + 1$ points such that for every $j \in \{1, \ldots, k\}$ we have $a_j - a_{j-1} \geq d$. Then for any integer t such that $0 \leq t \leq k$,*

$$
\prod_{j \neq t} |a_t - a_j| \geq \left(\frac{kd}{2e} \right)^k.
$$

Proof. First observe that we have $|a_t - a_j| \geq d|t - j|$. So we have

$$
\prod_{j \neq t} |a_t - a_j| \geq \prod_{j \neq t} d|t - j| = d^k t!(k - t)!.
$$

Note that $t!(k - t)! \geq k!/2^k$ since $\binom{k}{t} \leq 2^k$. Using Stirling's formula $x! \geq (x/e)^x$ for positive integer x, we have

$$
\prod_{j \neq t} |a_t - a_j| \geq d^k t!(k - t)! \geq d^k \frac{k!}{2^k} \geq d^k \left(\frac{k}{2e} \right)^k = \left(\frac{kd}{2e} \right)^k. \qquad \square
$$

Lemma 21. *If v is an integer such that $0 \leq v \leq k$, then*

$$
|L_v(a_v)| \geq \left(\frac{km}{6} \right)^k.
$$

Proof. Without loss of generality, assume that the a_j are in sorted order. Then this lemma follows from Claim 20 with d replaced by m. $\qquad \square$

3.8 Conclude upper bound on Δ

Lemma 22. *Let m be an integer such that $1 \leq m \leq \sqrt{n}$. Suppose that m is odd or n and c have the same parity. Suppose that $k \leq \frac{n}{8}$. If x is an integer, then*

$$|\Delta(x)| \leq 8(30)^k e^{-2n/m^2}.$$

Proof. If x is different from a_v for every v, then $\Delta(x) = 0$. So we may assume that $x = a_v$ for some v. By Lemmas 19 and 21, we have

$$
\begin{aligned}
|\Delta(x)| &= \frac{|\mathbb{E}[C_{m,c}(B)L_v(B)] - \mathbb{E}[L_v(B)]|}{|L_v(a_v)|} \\
&\leq \frac{8(5km)^k e^{-2n/m^2}}{(km/6)^k} \\
&= 8(30)^k e^{-2n/m^2}.
\end{aligned}
$$
$\square$

3.9 Bounding below $q(x)$

Lemma 23. *If a is an integer such that $|a| \leq n$ and $a \equiv n \pmod 2$, then*

$$\Pr[B = a] \geq 2^{-a^2/n} \frac{1}{2\sqrt{n}} \, .$$

Proof. The event $B = a$ is equivalent to $\frac{n+a}{2}$ of the X_j being 1 and the other $\frac{n-a}{2}$ being -1. Hence

$$\Pr[B = a] = \frac{1}{2^n} \binom{n}{(n+a)/2}.$$

Using a precise form of Stirling's formula due to Robbins [Rob55], we can bound the binomial coefficient by

$$\binom{n}{(n+a)/2} \geq 2^{H(\frac{1}{2} + \frac{a}{2n})n} \frac{2\sqrt{2\pi}}{e^2 \sqrt{n}} \, ,$$

where H is the binary entropy function $H(x) = -x \log_2 x - (1-x) \log_2(1-x)$. By using the Taylor series of H at $\frac{1}{2}$, we obtain the inequality $H(\frac{1+\varepsilon}{2}) \geq 1 - \varepsilon^2$. Plugging that inequality into our bound on the binomial coefficient gives the desired result. $\square$

3.10 Conclude $|\Delta(x)| \leq q(x)$

Lemma 24. *Let m and n be positive integers and c be an integer. Suppose that m is odd or n and c have the same parity. Let k be a positive integer such that $k \leq \frac{n}{8m^2}$. If x is an integer, then $|\Delta(x)| \leq \frac{1}{2}q(x)$.*

Proof. If x is different from a_v for every v, then $\Delta(x) = 0$. So we may assume that $x = a_v$ for some v. By Lemma 22, we have

$$|\Delta(x)| \leq 8(30)^k e^{-2n/m^2} \leq \frac{1}{4} e^{8k} e^{-2n/m^2} \leq \frac{1}{4} e^{n/m^2} e^{-2n/m^2} = \frac{1}{4} e^{-n/m^2}.$$

13

By the definition of q and Lemma 23, we have

$$q(x) = C_{m,c}(x) \Pr[B = x] \geq \frac{m}{2} \Pr[B = x] \geq 2^{-x^2/n} \frac{m}{4\sqrt{n}} .$$

We know that

$$|x| = |a_v| \leq 2km \leq \frac{n}{4m} .$$

So

$$q(x) \geq 2^{-n/(16m^2)} \frac{m}{4\sqrt{n}} \geq e^{-n/(16m^2)} \frac{m}{4\sqrt{n}} .$$

Applying the inequality $x \leq e^{x/e}$ (to $x = 4n/m^2$), we have

$$\frac{4n}{m^2} \leq e^{4n/(em^2)} \leq e^{3n/(2m^2)}.$$

Thus

$$\frac{m}{4\sqrt{n}} = \frac{1}{2}\left(\frac{4n}{m^2}\right)^{-1/2} \geq \frac{1}{2} e^{-3n/(4m^2)}.$$

Hence

$$q(x) \geq \frac{1}{2} e^{-3n/(4m^2)} e^{-n/(16m^2)} \geq \frac{1}{2} e^{-n/m^2}.$$

Comparing our bounds for Δ and q, we see that $|\Delta(x)| \leq \frac{1}{2} q(x)$. $\qquad \square$

3.11 Conclude lower bound

Proof of Theorem 13. Recall the function p from Lemma 16 is defined as $p(x) = q(x) - \Delta(x)$. We will show that p is the desired probability distribution. From the definition of p and Lemma 24, we get $p(x) \geq \frac{1}{2} q(x)$; in particular, p is nonnegative. Applying Lemma 16 to the constant function 1 (the zeroth moment), we see that the sum of the $p(x)$ is 1. In other words, p is indeed a probability distribution. Applying Lemma 16 to the other monomials (namely x, x^2, ..., x^k), we see that p matches the first k moments of B. This completes the proof. $\qquad \square$

4 Tight lower bound on k-wise uniformity vs. threshold

In this section we prove Theorem 6. Like the last section, we will work with $\{-1, 1\}^n$ and translate the results back to $\{0, 1\}^n$ using the following fact.

Fact 25. *Let $x \in \{0, 1\}^n$ and $y \in \{-1, 1\}^n$ be the string obtained by replacing each x_i by $y_i = 1 - 2x_i$. Then $|\sum_i x_i - n/2| \leq t$ if and only if $|\sum_i y_i| \leq 2t$.*

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be independent random variables chosen uniformly from $\{-1, 1\}$. Let B be the sum of all the X_j. The distribution of B is a shifted binomial distribution. Note that B has the same parity as n.

Theorem 26. *Let n and t be positive integers such that $t \leq n$. Let k be a positive integer such that $k \leq \frac{t^2}{200n}$. Then there is a probability distribution on the integers with absolute value at most t that matches the first k moments of B. Furthermore, the support of the probability distribution is a subset of the support of B.*

Theorem 6 follows from applying Fact 25 to Theorem 26.

4.1   Defining $T(x)$

Let m be an odd integer between $\frac{n}{3t}$ and $\frac{n}{t}$. In Theorem 13, we constructed a probability distribution (call it p') on the 0 mod m integers that matches the first $\frac{t^2}{8n}$ moments of B. Furthermore, the support of p' is a subset of the support of B. Looking at the proof, we see that $p'(x) \geq \frac{1}{2}C_{m,0}(x)\Pr[B = x]$ for all x. Let $C'(x)$ be $p'(x)/\Pr[B = x]$ if $\Pr[B = x] > 0$ and $C_{m,0}(x)$ otherwise. We have $p'(x) = C'(x)\Pr[B = x]$ for all x. Because p' matches the first few moments of B, we have $\mathbb{E}[C'(B)f(B)] = \mathbb{E}[f(B)]$ for every polynomial f of degree at most $\frac{t^2}{8n}$. Also, $C'(x) \geq \frac{1}{2}C_{m,0}(x)$ for all x.

Given an integer x, let $T(x)$ be $[\![\, |x| \leq t\,]\!]C'(x)$.

4.2   Defining $q(x)$

Given an integer x, define $q(x)$ to be $T(x)\Pr[B = x]$. Note that q is nonnegative. Also if $q(x) \neq 0$, then $|x| \leq t$ and x is in the support of B.

Lemma 27. *If f is a function on the integers, then*

$$\sum_x q(x)f(x) = \mathbb{E}[T(B)f(B)].$$

Proof. By the definition of expected value, we have

$$\sum_x q(x)f(x) = \sum_x \Pr[B = x]T(x)f(x) = \mathbb{E}[T(B)f(B)]. \qquad \square$$

4.3   Defining Lagrange polynomials

Let $a_0, a_1, \ldots, a_k$ be $k + 1$ distinct integers that are 0 mod m, n mod 2, and as close to 0 as possible. Because they are as close to 0 as possible, we have $|a_j| \leq (k + 1)m \leq 2km$. Because $k \leq \frac{t^2}{200n}$ and $m \leq \frac{n}{t}$, we have $2km \leq t \leq n$. So $|a_j| \leq t$ and a_j is in the support of B.

Given an integer v such that $0 \leq v \leq k$, define the Lagrange polynomial L_v as follows:

$$L_v(x) = \prod_{\substack{0 \leq j \leq k \\ j \neq v}} (x - a_j).$$

Note that $L_v(a_w) = 0$ if and only if $v \neq w$. It's well known that $L_0, L_1, \ldots, L_k$ form a basis (the Lagrange basis) of the vector space of polynomials of degree at most k.

4.4   Defining $\Delta(x)$

Define the function Δ on the integers as follows. If x equals a_v (for some v), then

$$\Delta(a_v) = \frac{\mathbb{E}[T(B)L_v(B)] - \mathbb{E}[L_v(B)]}{L_v(a_v)}.$$

For $x \neq a_w$ for any w, then $\Delta(x) = 0$.

Lemma 28. *If f is a polynomial of degree at most k, then*

$$\sum_x \Delta(x)f(x) = \mathbb{E}[T(B)f(B)] - \mathbb{E}[f(B)].$$

Proof. We will first prove the claim when f is a Lagrange polynomial L_v. If $\Delta(x) \neq 0$, then x is of the form a_w for some w. But if $L_v(a_w) \neq 0$, then $v = w$. So the sum has at most one nonzero term, corresponding to $x = a_v$. And the equation is true in this case by the definition of Δ.

We have proved the claim for Lagrange polynomials. But every polynomial of degree at most k is a linear combination of the Lagrange polynomials. This completes the proof. $\square$

4.5 Defining $p(x)$

Define the function p on the integers by $p(x) = q(x) - \Delta(x)$. Note that if $p(x) \neq 0$, then $|x| \leq t$ and x is in the support of B.

Lemma 29. *If f is a polynomial of degree at most k, then*

$$\sum_x p(x)f(x) = \mathbb{E}[f(B)].$$

Proof. By Lemmas 27 and 28, we have

$$\sum_x p(x)f(x) = \sum_x q(x)f(x) - \sum_x \Delta(x)f(x)$$
$$= \mathbb{E}[T(B)f(B)] - \Big(\mathbb{E}[T(B)f(B)] - \mathbb{E}[f(B)]\Big)$$
$$= \mathbb{E}[f(B)]. \qquad \square$$

We will show that $|\Delta(x)| \leq \frac{1}{2}q(x)$. First we bound above $|\Delta(x)|$. Then we bound below $q(x)$. Recall that $\Delta(x) = \frac{\mathbb{E}[L_v(B)T(B)] - \mathbb{E}[L_v(B)]}{L_v(a_v)}$ if x equals a_v for some v and 0 otherwise. Lemmas 33 and 34 below give upper and lower bounds on the numerator and denominator, respectively.

4.6 Bounding numerator of Δ

We start this subsection by bounding the central moments of a binomial distribution. The following fact is a special case of Khintchine's inequality [Khi23, Haa81]. For completeness, we provide a short, self-contained proof. Recall that we interpret 0^0 as 1.

Fact 30. *Let d be a nonnegative integer. Then the $(2d)$th moment $\mathbb{E}[B^{2d}]$ is at most $\frac{(2d)!}{2^d d!}n^d$.*

Proof. The odd moments of each X_j are all 0. The even moments of X_j are all 1. Let g_1, g_2, ..., g_n be independent standard Gaussians (with mean zero and unit variance). The odd moments of g_j are all 0. If c is a nonnegative integer, then the $(2c)$th moment of g_j is known to be $\frac{(2c)!}{2^c c!}$, the product of the positive odd integers less than $2c$. In particular, the

even moments of g_j are all at least 1. So the moments of X_j are at most the corresponding moments of g_j.

Let G be the sum of the g_j. When we expand B^{2d} and G^{2d}, each gives a sum of n^{2d} terms. By the previous paragraph, the expectation of each term of B^{2d} is at most the expectation of the corresponding term of G^{2d}. Hence the $(2d)$th moment of B is at most the $(2d)$th moment of G. But G is a Gaussian with mean zero and variance n. In particular, $G/\sqrt{n}$ is a standard Gaussian. So we have

$$\mathbb{E}[B^{2d}] \leq \mathbb{E}[G^{2d}] = n^d\,\mathbb{E}[(G/\sqrt{n})^{2d}] = \frac{(2d)!}{2^d d!}n^d. \qquad \square$$

Corollary 31. *If d is a nonnegative integer, then $\mathbb{E}[B^{2d}]$ is at most $\sqrt{2}\,(2dn/e)^d$.*

Proof. Fact 30 says that

$$\mathbb{E}[B^{2d}] \leq \frac{(2d)!}{2^d d!}n^d.$$

Using the form of Stirling's formula due to Robbins [Rob55], we can bound the factorials by

$$\frac{(2d)!}{2^d d!} < \sqrt{2}\left(\frac{2d}{e}\right)^d. \qquad \square$$

Lemma 32. *Let n and t be positive integers such that $t \leq n$ and $t^2 \geq 200n$. Let r be an integer such that $0 \leq r \leq \frac{t^2}{9n}$. Then*

$$\left|\mathbb{E}[B^r T(B)] - \mathbb{E}[B^r]\right| \leq 2t^r e^{-t^2/(6n)}.$$

Proof. Let s be a nonnegative integer such that $r + s$ is an even number between $\frac{t^2}{9n}$ and $\frac{t^2}{8n}$. (Because $t^2 \geq 200n$, there is such an s.) We have

$$[\![\,|B| > t\,]\!] \leq t^{-s}|B|^s.^3$$

Hence, by the moment-matching property of C', the definition of T, the triangle inequality, and Corollary 31, we have

$$
\begin{aligned}
\left|\mathbb{E}[B^r] - \mathbb{E}[B^r T(B)]\right| &= \left|\mathbb{E}[B^r C'(B)] - \mathbb{E}[B^r T(B)]\right| \\
&= \left|\mathbb{E}\!\left[B^r C'(B)[\![\,|B| > t\,]\!]\right]\right| \\
&\leq \mathbb{E}\!\left[|B|^r C'(B)[\![\,|B| > t\,]\!]\right] \\
&\leq t^{-s}\,\mathbb{E}\!\left[|B|^{r+s}C'(B)\right] \\
&= t^{-s}\,\mathbb{E}\!\left[B^{r+s}C'(B)\right] \\
&= t^{-s}\,\mathbb{E}\!\left[B^{r+s}\right] \\
&\leq 2t^{-s}\left(\frac{(r+s)n}{e}\right)^{(r+s)/2} \\
&= 2t^r\left(\frac{(r+s)n}{et^2}\right)^{(r+s)/2} \\
&\leq 2t^r\left(\frac{1}{8e}\right)^{t^2/(18n)} \\
&\leq 2t^r e^{-t^2/(6n)}. \qquad \square
\end{aligned}
$$

---

[3] Bounding a hard threshold by a smooth function is a standard idea in concentration inequalities, used for example in the proofs of Markov's inequality and Chebyshev's inequality.

Lemma 33. *Let n and t be positive integers such that $t \leq n$. Let k be a positive integer such that $k \leq \frac{t^2}{200n}$. If v is an integer such that $0 \leq v \leq k$, then*

$$\left| \mathbb{E}[L_v(B)T(B)] - \mathbb{E}[L_v(B)] \right| \leq 2(2t)^k e^{-t^2/(6n)}.$$

Proof. Recall that for a subset A of $\{1, 2, \ldots, n\}$, we define X^A to be the product of the X_j for which j is in A.

By expanding the product, we have

$$L_v(B) = \prod_{j \neq v}(B - a_j) = \sum_A (-1)^{|A|} a^A B^{k-|A|},$$

where A ranges over every subset of $\{0, 1, \ldots, k\} - \{v\}$. Therefore, by the triangle inequality, Lemma 32, and the binomial theorem, we have

$$
\begin{aligned}
|\mathbb{E}[L_v(B)T(B)] - \mathbb{E}[L_v(B)]| &= \left| \mathbb{E}\Big[[T(B) - 1]L_v(B) \Big] \right| \\
&= \left| \mathbb{E}\Big[[T(B) - 1]\sum_A (-1)^{|A|} a^A B^{k-|A|} \Big] \right| \\
&= \left| \sum_A (-1)^{|A|} a^A \, \mathbb{E}\Big[[T(B) - 1]B^{k-|A|} \Big] \right| \\
&\leq \sum_A \left| a^A \, \mathbb{E}\Big[[T(B) - 1]B^{k-|A|} \Big] \right| \\
&\leq \sum_A t^{|A|} \left| \mathbb{E}\Big[[T(B) - 1]B^{k-|A|} \Big] \right| \\
&\leq \sum_A t^{|A|} \cdot 2t^{k-|A|} e^{-t^2/(6n)} \\
&= 2e^{-t^2/(6n)} \sum_A t^k \\
&= 2e^{-t^2/(6n)}(2t)^k. \qquad \square
\end{aligned}
$$

4.7 Bounding denominator of Δ

Lemma 34. *If v is an integer such that $0 \leq v \leq k$, then*

$$|L_v(a_v)| \geq \left(\frac{kn}{9t} \right)^k.$$

Proof. Without loss of generality, assume that the a_j are in sorted order. Then this lemma follows from Claim 20 (with d replaced by $2m$) and the bound $m \geq \frac{n}{3t}$. $\square$

4.8 Conclude upper bound on Δ

Lemma 35. *Let n and t be positive integers such that $t \leq n$. Let k be a positive integer such that $k \leq \frac{t^2}{200n}$. If x is an integer, then*

$$|\Delta(x)| \leq \frac{1}{50} e^{-t^2/(12n)}.$$

Proof. If x is different from a_v for every v, then $\Delta(x) = 0$. So we may assume that $x = a_v$ for some v. By Lemmas 33 and 34, we have

$$
\begin{aligned}
|\Delta(x)| &= \frac{|\mathbb{E}[T(B)L_v(B)] - \mathbb{E}[L_v(B)]|}{|L_v(a_v)|} \\
&\leq \frac{2(2t)^k e^{-t^2/(6n)}}{(kn/(9t))^k} \\
&= 2\left(\frac{18t^2}{kn}\right)^k e^{-t^2/(6n)} \\
&\leq \frac{1}{50}\left(\frac{1800t^2}{kn}\right)^k e^{-t^2/(6n)}.
\end{aligned}
$$

The expression $(\frac{1800t^2}{kn})^k$ is an increasing function of k on the interval $(0, \frac{1800t^2}{en}]$. Because $k \leq \frac{t^2}{200n}$, we have

$$
\left(\frac{1800t^2}{kn}\right)^k \leq (1800 \cdot 200)^{t^2/(200n)} \leq e^{t^2/(12n)}.
$$

Plugging this bound into our previous inequality for $|\Delta(x)|$ completes the proof. $\qquad\square$

4.9 Conclude $|\Delta(x)| \leq q(x)$

Lemma 36. *Let n and t be positive integers such that $t \leq n$. Let k be a positive integer such that $k \leq \frac{t^2}{200n}$. If x is an integer, then $|\Delta(x)| \leq \frac{1}{2}q(x)$.*

Proof. If x is different from a_v for every v, then $\Delta(x) = 0$. So we may assume that $x = a_v$ for some v. By Lemma 35, we have $|\Delta(x)| \leq \frac{1}{50}e^{-t^2/(12n)}$. By the definition of q and Lemma 23, we have

$$
q(x) = C'(x)\Pr[B = x] \geq \frac{1}{2}C_{m,0}(x)\Pr[B = x] \geq \frac{n}{6t}\Pr[B = x] \geq 2^{-x^2/n}\frac{\sqrt{n}}{12t}.
$$

We know that

$$
|x| = |a_v| \leq 2km \leq 2 \cdot \frac{t^2}{200n} \cdot \frac{n}{t} = \frac{t}{100}.
$$

So

$$
q(x) \geq 2^{-t^2/(10000n)}\frac{\sqrt{n}}{12t} \geq e^{-t^2/(10000n)}\frac{\sqrt{n}}{12t}.
$$

Applying the inequality $x \leq e^{x/e}$ (to $x = \frac{t^2}{4n}$), we have

$$
\frac{t^2}{4n} \leq e^{t^2/(4en)} \leq e^{t^2/(10n)}.
$$

Thus

$$
\frac{\sqrt{n}}{12t} = \frac{1}{24}\left(\frac{t^2}{4n}\right)^{-1/2} \geq \frac{1}{24}e^{-t^2/(20n)}.
$$

Hence

$$
q(x) \geq \frac{1}{24}e^{-t^2/(20n)}e^{-t^2/(10000n)} \geq \frac{1}{24}e^{-t^2/(12n)}.
$$

Comparing our bounds for Δ and q, we see that $|\Delta(x)| \leq \frac{1}{2}q(x)$. $\qquad\square$

4.10 Conclude lower bound

Proof of Theorem 26. Recall the function p from Lemma 29 is defined as $p(x) = q(x) - \Delta(x)$. We will show that p is the desired probability distribution. From the definition of p and Lemma 36, we get $p(x) \geq \frac{1}{2}q(x)$; in particular, p is nonnegative. Applying Lemma 29 to the constant function 1 (the zeroth moment), we see that the sum of the $p(x)$ is 1. In other words, p is indeed a probability distribution. Applying Lemma 29 to the other monomials (namely x, x^2, ..., x^k), we see that p matches the first k moments of B. This completes the proof. $\qquad\square$

5 Twisted moments: proof of Lemma 17

In this section, we will restate and then prove the following lemma that was used in subsection 3.6. Recall that X_1, X_2, ..., X_n are independent random variables chosen uniformly from $\{-1, 1\}$, and B is the sum of all the X_j. Also, we interpret 0^0 as 1.

Lemma 17. *Let r be an integer such that $0 \leq r \leq \frac{n}{8}$. If $0 < \theta \leq \frac{\pi}{2}$, then*

$$\left| \mathbb{E}[B^r e^{i\theta B}] \right| \leq 2\left(\frac{8r}{\theta}\right)^r e^{-\theta^2 n/4}.$$

Informally, this lemma says that the rth moment of B, after being twisted in the complex plane, is exponentially small. Showing that this "twisted moment" is exponentially small is delicate. We have to somehow argue that lots of cancellation is taking place in the complex plane. Our approach will be to bound the delicate twisted moment by an expected value for which no cancellation is needed. That's exactly what Lemma 40(b) below does.

Recall that for a subset A of $\{1, 2, \ldots, n\}$, we define the monomial X^A to be the product of the X_j for which j is in A.

We will need the hyperbolic functions. Recall that $\cosh z$ is $(e^z + e^{-z})/2$, $\sinh z$ is $(e^z - e^{-z})/2$, $\tanh z$ is $\sinh z / \cosh z$, $\coth z$ is $\cosh z / \sinh z$, and $\operatorname{sech} z$ is $1/\cosh z$. Also, we have the identities $\cosh(i\theta) = \cos(\theta)$ and $\sinh(i\theta) = i\sin(\theta)$.

Our first lemma will express an exponential moment in terms of hyperbolic functions.

Lemma 37. *If A is a subset of $\{1, 2, \ldots, n\}$ and z is a complex number, then*

$$\mathbb{E}[X^A e^{zB}] = (\sinh z)^{|A|}(\cosh z)^{n-|A|}.$$

Proof. Independence of the X_j makes the proof routine. Because B is the sum of the X_j,

we have

$$\mathbb{E}[X^A e^{zB}] = \mathbb{E}\Big[\prod_{j\in A} X_j \prod_{j=1}^n e^{zX_j}\Big]$$

$$= \mathbb{E}\Big[\prod_{j\in A} X_j e^{zX_j} \prod_{j\notin A} e^{zX_j}\Big]$$

$$= \prod_{j\in A} \mathbb{E}[X_j e^{zX_j}] \prod_{j\notin A} \mathbb{E}[e^{zX_j}]$$

$$= \prod_{j\in A} \sinh z \prod_{j\notin A} \cosh z$$

$$= (\sinh z)^{|A|}(\cosh z)^{n-|A|}. \qquad \square$$

Our next lemma shows how to express a monomial twisted in the complex plane as a real quantity. This lemma will play a key role in bounding twisted moments.

Lemma 38. *Let A be a subset of $\{1, 2, \ldots, n\}$. Let θ be a real number such that $0 \le \theta < \frac{\pi}{4}$. Then $|\mathbb{E}[X^A e^{i\theta B}]| = (\cos 2\theta)^{n/2}\,\mathbb{E}[X^A e^{\lambda B}]$, where λ is $\frac{1}{2}\ln\frac{1+\tan\theta}{1-\tan\theta}$.*

Proof. The idea is that both expected values can be expressed in terms of hyperbolic functions using Lemma 37. So all we have to do is connect the hyperbolic function of an imaginary number with the hyperbolic function of a real number.

Because $0 \le \theta < \frac{\pi}{4}$, we have $0 \le \tan\theta < 1$, so λ is well defined. Also $\cos 2\theta > 0$. From our choice of λ, we have

$$\tanh\lambda = \frac{e^{2\lambda}-1}{e^{2\lambda}+1} = \frac{(1+\tan\theta)-(1-\tan\theta)}{(1+\tan\theta)+(1-\tan\theta)} = \tan\theta.$$

It follows that

$$\cosh\lambda = \frac{1}{\sqrt{1-\tanh^2\lambda}} = \frac{1}{\sqrt{1-\tan^2\theta}} = \frac{\cos\theta}{\sqrt{\cos^2\theta - \sin^2\theta}} = \frac{\cos\theta}{\sqrt{\cos 2\theta}}.$$

Hence

$$\sinh\lambda = \tanh\lambda\cosh\lambda = \tan\theta\frac{\cos\theta}{\sqrt{\cos 2\theta}} = \frac{\sin\theta}{\sqrt{\cos 2\theta}}.$$

Therefore, applying Lemma 37 twice, we have

$$|\mathbb{E}[X^A e^{i\theta B}]| = |\sinh i\theta|^{|A|}|\cosh i\theta|^{n-|A|}$$

$$= (\sin\theta)^{|A|}(\cos\theta)^{n-|A|}$$

$$= (\cos 2\theta)^{n/2}(\sinh\lambda)^{|A|}(\cosh\lambda)^{n-|A|}$$

$$= (\cos 2\theta)^{n/2}\,\mathbb{E}[X^A e^{\lambda B}]. \qquad \square$$

Let r be a nonnegative integer. Let f be a function from $\{1, 2, \ldots, r\}$ to $\{1, 2, \ldots, n\}$. Define $\mathrm{odd}(f)$, the *odd image* of f, to be the set of j in $\{1, 2, \ldots, n\}$ such that $|f^{-1}(j)|$ is odd. Note that $|\mathrm{odd}(f)| \le r$ and $|\mathrm{odd}(f)| \le n$.

Our next simple lemma expresses B^r as a sum of monomials.

21

Lemma 39. *If r is a nonnegative integer, then $B^r = \sum_f X^{\mathrm{odd}(f)}$, where the sum is over every function f from $\{1, \ldots, r\}$ to $\{1, \ldots, n\}$.*

Proof. The idea is to expand B^r and exploit the constraint that each X_j is ± 1. We have

$$B^r = \sum_f X_{f(1)} \cdots X_{f(r)}$$

$$= \sum_f \prod_{j=1}^n X_j^{|f^{-1}(j)|}$$

$$= \sum_f \prod_{j=1}^n X_j^{|f^{-1}(j)| \bmod 2}$$

$$= \sum_f \prod_{j \in \mathrm{odd}(f)} X_j$$

$$= \sum_f X^{\mathrm{odd}(f)}. \qquad \square$$

As mentioned at the start of this subsection, our next key lemma will bound a delicate twisted moment by a simpler real quantity. Part (a) is not surprising; part (b) is subtler. Which bound is more useful depends on the angle θ. For θ small (near zero), part (b) is better, whereas for θ large ($\theta \geq \frac{\pi}{4}$), part (b) can't even be applied.

Lemma 40. *Let r be a nonnegative integer. Let θ be a real number such that $0 \leq \theta \leq \frac{\pi}{2}$.*

(a) *If $r \leq n$, then $|\mathbb{E}[B^r e^{i\theta B}]| \leq n^r (\cos \theta)^{n-r}$.*

(b) *If $\theta < \frac{\pi}{4}$, then $|\mathbb{E}[B^r e^{i\theta B}]| \leq (\cos 2\theta)^{n/2} \, \mathbb{E}[B^r e^{\lambda B}]$, where λ is $\frac{1}{2} \ln \frac{1+\tan\theta}{1-\tan\theta}$.*

Proof. By Lemma 39 and the triangle inequality, we have

$$|\mathbb{E}[B^r e^{i\theta B}]| = \left| \mathbb{E}\left[\sum_f X^{\mathrm{odd}(f)} e^{i\theta B} \right] \right| = \left| \sum_f \mathbb{E}[X^{\mathrm{odd}(f)} e^{i\theta B}] \right| \leq \sum_f |\mathbb{E}[X^{\mathrm{odd}(f)} e^{i\theta B}]|.$$

We will use this inequality in both parts.

(a) By Lemma 37, we have

$$|\mathbb{E}[B^r e^{i\theta B}]| \leq \sum_f |\mathbb{E}[X^{\mathrm{odd}(f)} e^{i\theta B}]|$$

$$= \sum_f (\sin \theta)^{|\mathrm{odd}(f)|} (\cos \theta)^{n - |\mathrm{odd}(f)|}$$

$$= (\cos \theta)^{n-r} \sum_f (\sin \theta)^{|\mathrm{odd}(f)|} (\cos \theta)^{r - |\mathrm{odd}(f)|}$$

$$\leq (\cos \theta)^{n-r} \sum_f 1$$

$$= n^r (\cos \theta)^{n-r}.$$

22

(b) By Lemmas 38 and 39, we have

$$|\mathbb{E}[B^r e^{i\theta B}]| \le \sum_f |\mathbb{E}[X^{\mathrm{odd}(f)} e^{i\theta B}]|$$

$$= \sum_f (\cos 2\theta)^{n/2}\, \mathbb{E}[X^{\mathrm{odd}(f)} e^{\lambda B}]$$

$$= (\cos 2\theta)^{n/2} \sum_f \mathbb{E}[X^{\mathrm{odd}(f)} e^{\lambda B}]$$

$$= (\cos 2\theta)^{n/2}\, \mathbb{E}[B^r e^{\lambda B}]. \qquad \square$$

To apply Lemma 40(b), we need an upper bound on $\mathbb{E}[B^r e^{\lambda B}]$. Fortunately, we can obtain such a bound using straightforward analysis, as the next lemma shows.

Lemma 41. *If $r \ge 0$ and $\lambda > 0$, then*

$$\mathbb{E}[|B|^r e^{\lambda B}] \le 2\Big(\frac{4r}{\lambda}\Big)^r \Big(\cosh \frac{11}{10}\lambda\Big)^n.$$

Proof. First we will bound $|B|^r$ by an exponential. Let a be $\frac{10r}{e\lambda}$. Let's temporarily assume that $r \ne 0$. Applying the inequality $x \le e^{x/e}$ (to $x = |B|/a$), we have $|B| \le a e^{|B|/(ea)}$. Raising both sides to the rth power gives $|B|^r \le a^r e^{r|B|/(ea)}$. Plugging in the definition of a, we have $|B|^r \le a^r e^{|\lambda B|/10}$. This inequality is true for $r = 0$ too.

Now we are ready to prove the desired inequality. By Lemma 37 (with $A = \emptyset$), we have

$$\mathbb{E}[|B|^r e^{\lambda B}] \le \mathbb{E}[|B|^r e^{|\lambda B|}]$$

$$\le a^r\, \mathbb{E}[e^{|\lambda B|/10} e^{|\lambda B|}]$$

$$= a^r\, \mathbb{E}[e^{11|\lambda B|/10}]$$

$$\le a^r\, \mathbb{E}[e^{11\lambda B/10} + e^{-11\lambda B/10}]$$

$$= 2a^r \Big(\cosh \frac{11}{10}\lambda\Big)^n$$

$$= 2\Big(\frac{10r}{e\lambda}\Big)^r \Big(\cosh \frac{11}{10}\lambda\Big)^n$$

$$\le 2\Big(\frac{4r}{\lambda}\Big)^r \Big(\cosh \frac{11}{10}\lambda\Big)^n. \qquad \square$$

Soon we will need three standard inequalities on hyperbolic functions and trigonometric functions. We state them below without proof.

Fact 42. *If $\lambda \ge 0$, then $\tanh \lambda \le \lambda$.*

Fact 43. *If $0 \le \theta < \frac{\pi}{2}$, then $\tan \theta \ge \theta$.*

Fact 44. *If $0 \le \theta \le \frac{\pi}{2}$, then $\cos \theta \le e^{-\theta^2/2}$.*

We will also need two nonstandard inequalities on hyperbolic cosine and (ordinary) cosine. We state and prove them below.

Lemma 45. *If $\lambda \geq 0$ and $c \geq 1$, then $\cosh c\lambda \leq (\cosh \lambda)^{c^2}$.*

Proof. We will first prove that $\tanh cx \leq c \tanh x$ for every $x \geq 0$. The derivative of $\tanh$ is sech^2. Because $\cosh$ is increasing on $[0, \infty)$, it follows that $\operatorname{sech}^2 ct \leq \operatorname{sech}^2 t$ for every $t \geq 0$. Integrating both sides (from 0 to x) gives $\frac{1}{c} \tanh cx \leq \tanh x$. Multiplying by c gives $\tanh cx \leq c \tanh x$.

Next we will prove the $\cosh$ inequality. The derivative of $\ln \cosh$ is $\tanh$. By the previous paragraph, we have $\tanh cx \leq c \tanh x$ for every $x \geq 0$. Integrating both sides (from 0 to λ) gives $\frac{1}{c} \ln \cosh c\lambda \leq c \ln \cosh \lambda$. Multiplying by c and exponentiating gives the desired inequality. $\qquad\square$

Lemma 46. *If θ is a real number such that $\cos^2 \theta \geq \frac{\sqrt{5}-1}{2}$, then $\cos 2\theta \geq \cos^6 \theta$.*

Proof. From the hypothesis, we have

$$\cos^2 \theta + \cos^4 \theta = \cos^2 \theta(1 + \cos^2 \theta) \geq \frac{\sqrt{5}-1}{2} \cdot \frac{\sqrt{5}+1}{2} = 1.$$

Therefore, we have

$$\begin{aligned}
\cos^6 \theta &= 1 - (1 - \cos^2 \theta)(1 + \cos^2 \theta + \cos^4 \theta) \\
&= 1 - \sin^2 \theta(1 + \cos^2 \theta + \cos^4 \theta) \\
&\leq 1 - 2\sin^2 \theta \\
&= \cos 2\theta. \qquad\qquad\square
\end{aligned}$$

5.1   Conclude proof of twisted moments lemma

Finally we have all the ingredients needed to prove our upper bound on twisted moments. We will restate the bound again and then prove it.

Lemma 17. *Let r be an integer such that $0 \leq r \leq \frac{n}{8}$. If $0 < \theta \leq \frac{\pi}{2}$, then*

$$\left| \mathbb{E}[B^r e^{i\theta B}] \right| \leq 2\left(\frac{8r}{\theta}\right)^r e^{-\theta^2 n/4}.$$

Proof. We will actually prove the stronger bound

$$\left| \mathbb{E}[B^r e^{i\theta B}] \right| \leq 2(8r \cot \theta)^r (\cos \theta)^{n/2}.$$

The lemma will then follow from Facts 43 and 44.

We will consider two cases: θ large (namely $\cos^2 \theta \leq e^{-1/e}$) and θ small ($\cos^2 \theta \geq \frac{\sqrt{5}-1}{2}$). Because $\frac{\sqrt{5}-1}{2} < e^{-1/e}$, these two cases cover all possible θ. For large θ we will use Lemma 40(a), whereas for small θ we will use Lemma 40(b).

Case 1: θ large ($\cos^2 \theta \leq e^{-1/e}$). Applying the inequality $x^x \geq e^{-1/e}$ for $x \geq 0$ (to $x = 8r/n$), we have

$$\cos^2 \theta \leq e^{-1/e} \leq \left(\frac{8r}{n}\right)^{8r/n}.$$

24

Hence, by Lemma 40(a), we have

$$
\begin{aligned}
\left|\mathbb{E}[B^r e^{i\theta B}]\right| &\leq n^r (\cos\theta)^{n-r} \\
&\leq n^r (\cos\theta)^{3n/4+r} \\
&= n^r (\cos^2\theta)^{n/8} (\cos\theta)^{n/2+r} \\
&\leq n^r \left(\frac{8r}{n}\right)^r (\cos\theta)^{n/2+r} \\
&= (8r\cos\theta)^r (\cos\theta)^{n/2} \\
&\leq (8r\cot\theta)^r (\cos\theta)^{n/2} \\
&\leq 2(8r\cot\theta)^r (\cos\theta)^{n/2}.
\end{aligned}
$$

Case 2: θ small $(\cos^2\theta \geq \frac{\sqrt{5}-1}{2})$. In particular, $\theta < \frac{\pi}{4}$. Let λ be $\frac{1}{2}\ln\frac{1+\tan\theta}{1-\tan\theta}$. In the proof of Lemma 38, we showed that $\tanh\lambda$ is $\tan\theta$ and $\cosh\lambda$ is $\cos\theta/\sqrt{\cos 2\theta}$. By Fact 42, we have $\lambda \geq \tanh\lambda = \tan\theta$. Hence, by Lemmas 40(b), 41, 45, and 46, we have

$$
\begin{aligned}
\left|\mathbb{E}[B^r e^{i\theta B}]\right| &\leq (\cos 2\theta)^{n/2}\,\mathbb{E}[B^r e^{\lambda B}] & \text{(Lemma 40(b))} \\
&\leq (\cos 2\theta)^{n/2}\,\mathbb{E}[|B|^r e^{\lambda B}] \\
&\leq 2(\cos 2\theta)^{n/2}\left(\frac{4r}{\lambda}\right)^r \left(\cosh\frac{11}{10}\lambda\right)^n & \text{(Lemma 41)} \\
&\leq 2(\cos 2\theta)^{n/2}(4r\cot\theta)^r \left(\cosh\frac{11}{10}\lambda\right)^n \\
&\leq 2(\cos 2\theta)^{n/2}(4r\cot\theta)^r (\cosh\lambda)^{121n/100} & \text{(Lemma 45)} \\
&\leq 2(\cos 2\theta)^{n/2}(4r\cot\theta)^r (\cosh\lambda)^{5n/4} \\
&= 2(4r\cot\theta)^r (\cos 2\theta)^{n/2}\left(\frac{\cos\theta}{\sqrt{\cos 2\theta}}\right)^{5n/4} \\
&= 2(4r\cot\theta)^r \frac{(\cos\theta)^{5n/4}}{(\cos 2\theta)^{n/8}} \\
&\leq 2(4r\cot\theta)^r \frac{(\cos\theta)^{5n/4}}{(\cos\theta)^{3n/4}} & \text{(Lemma 46)} \\
&= 2(4r\cot\theta)^r (\cos\theta)^{n/2} \\
&\leq 2(8r\cot\theta)^r (\cos\theta)^{n/2}. & \square
\end{aligned}
$$

6 Tight upper bound on k-wise uniformity vs. mod m

In this section we prove Theorem 3. It follows from Theorem 47 below by translating the statement for $\{-1,1\}^n$ back to $\{0,1\}^n$ using Fact 12. Recall that $S'_{m,c} = \{y \in \{-1,1\}^n : \sum_i y_i \equiv c \pmod{m}\}$.

Theorem 47. *Let m be a positive integer and let c be an integer. Let k be an integer greater than or equal to 4. Suppose there is a k-wise uniform distribution on $\{-1,1\}^n$ that is supported on $S'_{m,c}$. Then $k \leq \frac{140n}{m^2}$.*

We can restate the theorem using moments as follows. Let B be the sum of n independent random bits chosen uniformly from $\{-1, 1\}$. Let Y be a random variable that is always $c \bmod m$. Suppose the first k moments of Y match those of B. Then $k \le \frac{140n}{m^2}$.

The high-level idea of the proof is as follows. We compare $|\mathbb{E}[e^{2\pi i Y/m}]|$ and $|\mathbb{E}[e^{2\pi i B/m}]|$. The former is 1 because Y is always $c \bmod m$. We show that the latter is less than $e^{-\Omega(n/m^2)}$. We then take the Taylor approximations of the exponentials. The error is given by the kth term. Because the first k terms are equal by the moment-matching property of Y, the difference of the two expectations can be bounded in terms of the kth moment of B, which gives us an upper bound on k.

Proof of Theorem 47. If $m < 4$, then $k \le n \le \frac{16n}{m^2}$. So we may assume that $m \ge 4$. Let $\alpha = 2\pi/m$. Because $m \ge 4$, we have $0 < \alpha \le \pi/2$. For now, we will assume that k is even. At the end, we will handle the odd case.

Because Y is always $c \bmod m$, we have

$$\left|\mathbb{E}[e^{i\alpha Y}]\right| = \left|e^{i\alpha c}\right| = 1.$$

By Lemma 37 (with $A = \emptyset$) and Fact 44, we have

$$\left|\mathbb{E}[e^{i\alpha B}]\right| = |\cos\alpha|^n \le e^{-\alpha^2 n/2}.$$

The error in the Taylor approximation of an exponential has the following standard bound (see for example Feller [Fel71, Section XV.4]): for every real θ we have

$$\left|e^{i\theta} - \sum_{j=0}^{k-1} \frac{(i\theta)^j}{j!}\right| \le \frac{\theta^k}{k!}.$$

Hence by the triangle inequality we have

$$\left|\mathbb{E}[e^{i\alpha Y}] - \sum_{j=0}^{k-1} \frac{(i\alpha)^j}{j!}\mathbb{E}[Y^j]\right| \le \mathbb{E}\left[\left|e^{i\alpha Y} - \sum_{j=0}^{k-1}\frac{(i\alpha Y)^j}{j!}\right|\right] \le \frac{\alpha^k}{k!}\mathbb{E}[Y^k].$$

Similarly we have

$$\left|\mathbb{E}[e^{i\alpha B}] - \sum_{j=0}^{k-1} \frac{(i\alpha)^j}{j!}\mathbb{E}[B^j]\right| \le \frac{\alpha^k}{k!}\mathbb{E}[B^k].$$

Because the first k moments of Y match those of B, we get a ton of cancellation:

$$\left|\mathbb{E}[e^{i\alpha Y}] - \mathbb{E}[e^{i\alpha B}]\right| \le \frac{\alpha^k}{k!}\mathbb{E}[Y^k] + \frac{\alpha^k}{k!}\mathbb{E}[B^k] = 2\frac{\alpha^k}{k!}\mathbb{E}[B^k].$$

In particular, we have

$$1 = \left|\mathbb{E}[e^{i\alpha Y}]\right| \le \left|\mathbb{E}[e^{i\alpha B}]\right| + 2\frac{\alpha^k}{k!}\mathbb{E}[B^k] \le e^{-\alpha^2 n/2} + 2\frac{\alpha^k}{k!}\mathbb{E}[B^k].$$

Hence, using the moment bound of Fact 30, we have

$$1 \le e^{-\alpha^2 n/2} + 2\frac{\alpha^k}{2^{k/2}(k/2)!}n^{k/2} \le e^{-\alpha^2 n/2} + \frac{2}{(k/2)!}\left(\frac{\alpha^2 n}{2}\right)^{k/2}.$$

26

Let f be the function defined by $f(x) = e^{-x} + \frac{2}{(k/2)!}x^{k/2}$. Then the inequality above simplifies to $f(\alpha^2 n/2) \geq 1$. Note that $f(0) = 1$. Also f is convex on the interval $[0, \infty)$.

We claim that $f(k\sqrt{2}/8) < 1$. To prove it, we will show that the first term of f is less than $\frac{1}{2}$ and the second term is at most $\frac{1}{2}$. Because $k \geq 4$, the first term indeed satisfies

$$e^{-k\sqrt{2}/8} \leq e^{-\sqrt{2}/2} < \frac{1}{2} \,.$$

The second term is $\frac{2}{(k/2)!}(k\sqrt{2}/8)^{k/2}$. If $k = 4$, then this term is exactly $\frac{1}{2}$. Otherwise $k \geq 6$, and so by Stirling's formula we have

$$\frac{2}{(k/2)!}\left(\frac{k\sqrt{2}}{8}\right)^{k/2} \leq \frac{2}{(k/2)^{k/2}e^{-k/2}\sqrt{\pi k}}\left(\frac{k\sqrt{2}}{8}\right)^{k/2} = \frac{2}{\sqrt{\pi k}}\left(\frac{e\sqrt{2}}{4}\right)^{k/2} < \frac{2}{\sqrt{\pi k}} < \frac{1}{2} \,.$$

In either case, the second term is at most $\frac{1}{2}$. Hence $f(k\sqrt{2}/8) < 1$.

To summarize, f is convex on $[0, \infty)$, $f(0) = 1$, and $f(k\sqrt{2}/8) < 1$. It follows that f is less than 1 on the interval $(0, k\sqrt{2}/8]$. Because $f(\alpha^2 n/2) \geq 1$, we have $\alpha^2 n/2 > k\sqrt{2}/8$. Solving for k gives

$$k < 2\sqrt{2}\alpha^2 n = 2\sqrt{2}\left(\frac{2\pi}{m}\right)^2 n < \frac{112n}{m^2} \,.$$

So far, we assumed that k is even. Now suppose that k is odd. We can apply the proof above to $k-1$, which gives the bound $k - 1 < \frac{112n}{m^2}$. Because $k \geq 5$, we have

$$k \leq \frac{5}{4}(k-1) < \frac{5}{4} \cdot \frac{112n}{m^2} = \frac{140n}{m^2} \,. \qquad \square$$

7 Tight upper bound on k-wise uniformity vs. threshold

In this section we prove Theorem 7, which follows from Theorem 49 below by translating the statement for $\{-1,1\}^n$ to $\{0,1\}^n$ using Fact 25.

We will show that for any $k \geq 3$, any k-wise uniform distribution over $\{-1,1\}^n$ must put nonzero probability masses on strings x whose sums $\sum_{i=1}^n x_i$ are $-\Omega(\sqrt{nk})$ and $\Omega(\sqrt{nk})$ away from 0. This result shows that the lower bound we obtain in Theorem 6 is tight. We note that this is not true for $k = 2$, as when n is odd, there exists a pairwise uniform distribution supported on the all -1 vector and vectors with $(n+1)/2$ ones.

Let $X_1, X_2, \ldots, X_n$ be independent random variables chosen uniformly from $\{-1, 1\}$. Let B be the sum of all the X_j. The distribution of B is a shifted binomial distribution.

First we give a lower bound on the dth moment of B. Recall that we interpret 0^0 as 1.

Claim 48. *Let d be a nonnegative even integer such that $d \leq 2n$. Then $\mathbb{E}[B^d] \geq \left(\frac{nd}{e^2}\right)^{d/2}$.*

Proof. Let $d = 2r$. Consider expanding B^{2r}, which gives us a sum of n^{2r} terms. If for some index i, the variable X_i appears an odd number of times in a term, then this term has expectation zero. So the terms with nonzero expectation are the ones in which each X_i appears an even number of times. In particular, each such term has expectation 1. It suffices

to consider the terms in which either each X_i appears exactly twice or does not appear at all. There are $\binom{n}{r}$ ways of choosing the indices that appear twice in a term, and each term appears $(2r)!/2^r$ number of times in the n^{2r} terms. Hence we have

$$\mathbb{E}[B^{2r}] \geq \binom{n}{r} \frac{(2r)!}{2^r}.$$

Using the inequality $\binom{n}{r} \geq (n/r)^r$ and a crude form of Stirling's formula, $n! \geq (n/e)^n$, we have

$$\binom{n}{r}\frac{(2r)!}{2^r} \geq \left(\frac{n}{r}\right)^r \left(\frac{2r}{e}\right)^{2r} \frac{1}{2^r} = \left(\frac{2nr}{e^2}\right)^r,$$

proving the claim. $\qquad\square$

Theorem 49. *Let t^+ and t^- be two positive integers. Let Y be a random variable that is supported on $\{-1,1\}^n$ so that $\sum_i Y_i \geq -t^-$ and $\sum_i Y_i \leq t^+$. Let k be a positive integer such that $k \leq n$. Suppose that both the $(2k)$th and $(2k+1)$th moments of Y and B are equal. Then $\min\{t^-, t^+\} \geq \sqrt{nk}/3$.*

Remark 2. *The conclusion is false when Y matches only the first two moments of B. When n is odd, there exists a pairwise uniform distribution using Hadamard matrices that is supported on the all -1 vector and vectors with $(n+1)/2$ ones. Namely, with probability $\frac{1}{n+1}$ choose the all -1 vector and with the remaining probability $\frac{n}{n+1}$ choose a vector with $(n+1)/2$ ones uniformly at random.*

Proof. Let p^+ and p^- denote $\Pr[Y \geq 0]$ and $\Pr[Y < 0]$ respectively. Note that $\mathbb{E}[Y^{2k+1}] = \mathbb{E}[B^{2k+1}] = 0$. Thus,

$$p^+ \mathbb{E}[|Y|^{2k+1} \mid Y \geq 0] - p^- \mathbb{E}[|Y|^{2k+1} \mid Y < 0] = \mathbb{E}[Y^{2k+1}] = 0$$

$$p^+ \mathbb{E}[|Y|^{2k+1} \mid Y \geq 0] + p^- \mathbb{E}[|Y|^{2k+1} \mid Y < 0] = \mathbb{E}[|Y|^{2k+1}] \geq \mathbb{E}[Y^{2k}]^{\frac{2k+1}{2k}},$$

where the last inequality follows from Jensen's inequality. Summing the two relations, we have $2p^+ \mathbb{E}[|Y|^{2k+1} \mid Y \geq 0] \geq \mathbb{E}[Y^{2k}]^{\frac{2k+1}{2k}}$. Because the $(2k)$th moment of Y is equal to the $(2k)$th moment of B, together with Claim 48, we have $\mathbb{E}[Y^{2k}]^{\frac{2k+1}{2k}} = \mathbb{E}[B^{2k}]^{\frac{2k+1}{2k}} \geq \left(\frac{2nk}{e^2}\right)^{\frac{2k+1}{2}}$. Hence, there must be a point y in the support of Y such that $y^{2k+1} \geq (nk/9)^{\frac{2k+1}{2}}$, and so $y \geq \sqrt{nk}/3$. By symmetry, there is another point y' in the support of Y such that $y' \leq -\sqrt{nk}/3$. $\qquad\square$

Acknowledgments

References

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

[AW89]   Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant-depth circuits. *Advances in Computing Research — Randomness and Computation*, 5:199–223, 1989.

[Baz09]   Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009.

[BDT16]   Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Explicit two-source extractors for near-logarithmic min-entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:88, 2016.

[BHLV16]   Ravi Boppana, Johan Håstad, Chin Ho Lee, and Emanuele Viola. Bounded independence vs. moduli. In *Proc. 20th International Workshop on Randomization and Computation (RANDOM)*, volume 60 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 24:1–24:9. Schloss Dagstuhl, 2016.

[Bra10]   Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *J. of the ACM*, 57(5), 2010.

[Car]   Neal Carothers. A short course on approximation theory. Available at http://fourier.math.uoc.gr/ mk/approx1011/carothers.pdf.

[Che66]   Elliott Cheney. *Introduction to Approximation Theory*. McGraw-Hill, New York, New York, 1966.

[CRS00]   Suresh Chari, Pankaj Rohatgi, and Aravind Srinivasan. Improved algorithms via approximations of probability distributions. *J. Comput. System Sci.*, 61(1):81–107, 2000.

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

[CZ16]   Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *48th ACM Symp. on the Theory of Computing (STOC)*, pages 670–683, 2016.

[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.

[DKN10]   Ilias Diakonikolas, Daniel Kane, and Jelani Nelson. Bounded independence fools degree-2 threshold functions. In *51st IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 11–20. IEEE, 2010.

[EGL+92]   Guy Even, Oded Goldreich, Michael Luby, Noam Nisan, and Boban Velickovic. Approximations of general independent distributions. In *24th ACM Symp. on the Theory of Computing (STOC)*, pages 10–16, 1992.

[Fel71]   William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, second edition, 1971.

[GMR+12]   Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudorandom generators from milder pseudorandom restrictions. In *53rd IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 120–129, 2012.

[GOWZ10] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *25th IEEE Conf. on Computational Complexity (CCC)*, pages 223–234. IEEE, 2010.

[Haa81] Uffe Haagerup. The best constants in the Khintchine inequality. *Studia Mathematica*, 70(3):231–283, 1981.

[HLV18] Elad Haramaty, Chin Ho Lee, and Emanuele Viola. Bounded independence plus noise fools products. *SIAM J. on Computing*, 47(2):493–523, 2018.

[HS16] Prahladh Harsha and Srikanth Srinivasan. On polynomial approximations to AC^0. In *Proc. 20th International Workshop on Randomization and Computation (RANDOM)*, volume 60 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 32:1–32:14. Schloss Dagstuhl, 2016. Journal version to appear in *Random Structures & Algorithms*, 2019.

[Khi23] Aleksandr Khintchine. Über dyadische Brüche. *Mathematische Zeitschrift*, 18:109–116, 1923.

[LV17a] Chin Ho Lee and Emanuele Viola. More on bounded independence plus noise: Pseudorandom generators for read-once polynomials. Available at `http://www.ccs.neu.edu/home/viola/`, 2017.

[LV17b] Chin Ho Lee and Emanuele Viola. Some limitations of the sum of small-bias distributions. *Theory of Computing*, 13, 2017.

[MZ09] Raghu Meka and David Zuckerman. Small-bias spaces for group products. In *13th Workshop on Randomization and Computation (RANDOM)*, volume 5687 of *Lecture Notes in Computer Science*, pages 658–672. Springer, 2009.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. Available on line at `http://analysisofbooleanfunctions.net/`.

[Raz87] Alexander A. Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Akademiya Nauk SSSR. Matematicheskie Zametki*, 41(4):598–607, 1987. English translation in Mathematical Notes of the Academy of Sci. of the USSR, 41(4):333-338, 1987.

[Raz09] Alexander A. Razborov. A simple proof of Bazzi's theorem. *ACM Transactions on Computation Theory (TOCT)*, 1(1), 2009.

[Rob55] Herbert Robbins. A remark on Stirling's formula. *The American Mathematical Monthly*, 62(1):26–29, January 1955.

[RS10] Yuval Rabani and Amir Shpilka. Explicit construction of a small epsilon-net for linear threshold functions. *SIAM J. on Computing*, 39(8):3501–3520, 2010.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *19th ACM Symp. on the Theory of Computing (STOC)*, pages 77–82. ACM, 1987.

[Tal17]   Avishay Tal. Tight bounds on the Fourier spectrum of AC^0. In *32nd Computational Complexity Conference*, volume 79 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages 15:1–15:31. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2017.

[Vio17]   Emanuele Viola. Special topics in complexity theory. Lecture notes of the class taught at Northeastern University. Available at `http://www.ccs.neu.edu/home/viola/classes/spepf17.html`, 2017.

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