Abstract

We prove several new results about bounded uniform and small-bias distributions. A main message is that, small-bias, even perturbed with noise, does not fool several classes of tests better than bounded uniformity. We prove this for threshold tests, small-space algorithms, and small-depth circuits. In particular, we obtain small-bias distributions that achieve an optimal lower bound on their statistical distance to any bounded-uniform distribution. This closes a line of research initiated by Alon, Goldreich, and Mansour in 2003, and improves on a result by O’Donnell and Zhao.

We have heavier tail mass than the uniform distribution. This answers a question posed by several researchers including Bun and Steinke.

We rule out a popular paradigm for constructing pseudorandom generators, originating in a 1989 work by Ajtai and Wigderson. This again answers a question raised by several researchers. For branching programs, our result matches a bound by Forbes and Kelley.

Our small-bias distributions above are symmetric. We show that the xor of any two symmetric small-bias distributions fools any bounded function. Hence our examples cannot be extended to the xor of two small-bias distributions, another popular paradigm whose power remains unknown. We also generalize and simplify the proof of a result of Bazzi.

1 Introduction

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; 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 [52], are closely related to universal hash functions [17], error-correcting codes (see e.g. [35]), and in their modern guise were introduced in the works [5, 23, 47].

$(\varepsilon, k)$-biased distributions behave like the uniform distribution in that several prominent tests cannot distinguish the two distributions.
Definition 1. A test \( f : \{-1, 1\}^n \to [-1, 1] \) is \( \delta \)-fooled by a distribution \( D \) we have \( |E[f(U)] - E[f(D)]| \leq \delta \), where \( U \) is the uniform distribution.

At the same time, \( (\epsilon, k) \)-biased distributions can be sampled efficiently from a short seed. The combination of these facts enables many applications in algorithm design, coding theory, pseudorandomness, and more. For background we refer the reader to [60, 35, 64], where bounds on seed lengths are also discussed.

To generate \( k \)-wise uniformity, seed length \( ck \log n \) is sufficient, and necessary for \( k < n^\epsilon \); while for \( \epsilon \)-biased seed length \( c \log(n/\epsilon) \) is sufficient and necessary. In this paper, as in [64], 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-Oh notation.

It is known that any \( \epsilon \)-biased distribution is close to a \( k \)-wise uniform distribution in total variation (a.k.a. statistical, \( L_1 \), etc.) distance [6, 4, 50].

Lemma 2 (Theorem 1.1 [50]). Any \( (\epsilon, k) \)-biased distribution is \( (\frac{\epsilon^3 n}{4})^{k/2} \)-close to a \( k \)-wise uniform distribution in total variation distance.

Hence, any property enjoyed by \( k \)-wise uniform distributions is inherited by distributions with bias \( n^{-ck} \). Unsurprisingly, that is precisely the bias for which the seed length of the latter matches that of the former, as discussed above. The question arises as to which tests can be fooled with a larger bias, which would result in shorter seed length.

Question 3. Which tests can distinguish some \( \epsilon \)-biased distribution from every \( k \)-wise uniform distribution, for an \( \epsilon \) suitably larger than the bound in Lemma 2?

For a concrete setting, one can think e.g. \( k = 10 \log n \) and \( \epsilon = n^{-100} \), or any \( \epsilon = n^{-o(\log n)} \).

Question 3 is a computational version of the classic question of the statistical distance between \( \epsilon \)-biased and \( k \)-wise uniform distributions, studied in [6, 4, 50]. Lemma 2 shows that for small \( \epsilon \), no test, efficient or not, can distinguish the distributions. Those works also give lower bounds in various ranges of parameters, which means that in those ranges, there exists some \( \epsilon \)-biased distribution such that every \( k \)-wise uniform distribution can be distinguished from it by some test. However, the arguments in these papers either do not apply to the tests we consider below or for bias \( \epsilon \) larger than \( n^{-k} \), which is the regime of interest here. These works are discussed more below.

A trivial test which cannot distinguish between small-bias and \( k \)-wise uniform distributions in the sense of Question 3 is parity. By definition, the bias of parity is at most \( \epsilon \), which is \( \epsilon \)-close to the bias of the uniform distribution, which is 0. And the uniform distribution is in particular \( k \)-wise uniform.

However, the answer to Question 3 was not known for various other classes of tests of interest. To our knowledge, it was not known for symmetric or even threshold tests (a.k.a. tail, deviation, concentration bounds, etc.). In particular, Bun and Steinke posed the following question in [16].

In this work, we focused on understanding the limits of \( k \)-wise independent distributions. Gopalan et al. [31] gave a much more sophisticated generator with nearly optimal seed length. But could simple, natural pseudorandom distributions, such as small-bias spaces, give strong tail bounds themselves?

More concretely, the following question has been asked by several researchers. We use \( 1^\top x \) to denote the sum \( \sum_{i=1}^n x_i \) of \( x \in \{-1, 1\}^n \), and \( B \) to denote the binomial distribution \( 1^\top U \).
Question 4. Is it true that for every \( a \), there exists \( b \) such that
\[
\Pr[1^T D \geq t] \leq \Pr[B \geq t] + \frac{1}{n^a}
\]
for every \( n^{-b} \)-biased distribution \( D \) and \( t = \sqrt{n \log n} \)?

The answer to Question 4 was known to be negative for \( t = 0 \) (corresponding to the majority function): One can take \( D \) to be uniform on strings of weight 0 modulo 3, see [8]. The answer was also known to be positive when \( t > n^{1/2 + \varepsilon} \) for a constant \( \varepsilon \) because all the relevant quantities are small enough; formally combine Corollary 28 with Lemma 2. But for other values of \( t \) closer to \( \sqrt{n} \) the answer was less clear.

Smoothed tests and distributions

A main focus of this paper is on smoothed tests and smoothed distributions, which are tests and distributions perturbed by noise.

Definition 5. \( N_\rho \) 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 \( \{0, 1\} \). Note \( x \cdot N_1 = x \) and \( x \cdot N_0 = U \), for any \( x \).

For a test \( f \) and a distribution \( D \), a smoothed distribution is defined as \( D \cdot N_\rho \) and a smoothed test is defined as \( T_\rho f(x) := \mathbf{E}[f(x \cdot N_\rho)] \), for some retention rate \( \rho \in [0, 1] \). Note that \( \mathbf{E}[f(D \cdot N_\rho)] = \mathbf{E}[T_\rho f(D)] \) and we will use both viewpoints interchangeably throughout.

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 tests and 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 and are discussed next; for additional background, we refer the readers to the recent monograph [35].

Small-bias plus (pseudorandom) noise

This paradigm goes back to Ajtai and Wigderson [3], but saw no further work until it was revived by Gopalan, Meka, Reingold, Trevisan, and Vadhan [32]. It has been used in a number of subsequent works including [30, 54, 56, 34, 42, 21, 29, 45, 40, 28, 22]. In particular, this paradigm gave rise to PRGs with near-optimal seed lengths for several well-studied classes of tests, including combinatorial rectangles [32, 40] and read-once AC\(^0\) formulas [27, 28].

The Ajtai–Wigderson paradigm comprises several steps. A main step in this paradigm requires fooling (the average of) a random restriction of tests with a pseudorandom distribution. (This can also be viewed as constructing a fractional pseudorandom generator [19, 20, 18].) The works by Haramaty, Lee, and Viola [41, 34, 42, 40] have reinterpreted the notion of “random restrictions” as perturbing or xor-ing a small-bias distribution with noise. The perspective of noise has proved influential and is maintained in several following works, including the present one.

This perspective of noise has been used to prove a variety of new results in areas ranging from communication complexity [34], coding theory [34, 55], Turing machines [63], and one-way small-space computation [29, 45].

In particular, building on the proof in [34], Forbes and Kelley [29] significantly improved the parameters in [34] and obtained pseudorandom generators with seed length \( c \log^2 n \) that fool one-way logspace computation. The main new feature of their result over the classic generator by Nisan [48] is that the order in which the input is read by the computation is
arbitrary. A main step in the result in [29] is showing that $c \log n$-wise uniformity xor-ed with noise fools logspace. After their work, a natural question, asked independently by several researchers, is whether one can improve the seed length to $o(\log^3 n)$ by replacing $c \log n$-wise uniformity with polynomial bias.

**Question 6.** Does $1/\text{poly}(n)$-bias plus noise fool one-way logspace?

A positive answer would give improved generators for small-space algorithms from $c \log^3 n$ to $c \log^2 n$, bringing the parameters of the result in [29], which works in any order, in alignment with the classic fixed-order result of Nisan [48].

In fact, the answer to Question 6 was not known even for the special case of one-way logspace algorithms which compute symmetric tests; or for the even more restricted class of threshold tests that have the form $1(1^T x \geq t)$, for any bias larger than $n^{-\log n}$.

**Xor-ing small-bias distributions**

Starting with [13], researchers have considered the bit-wise xor of several independent copies of small-bias distributions. The work [41] draws a connection with the previous paradigm, showing that for a special class of small-bias distributions, the paradigms are equivalent.

These distributions – the xor of several small-bias distributions – appear to be significantly more powerful than a single small-bias distribution, while retaining a modest seed length. We refer to [49, 62, 35, 64] for background.

Despite several attempts [12, 46, 41], no definitive counterexample to this paradigm has been bound; its power remains unknown.

### 1.1 Our results

In this work we prove several new results on $(\varepsilon, k)$-biased distributions. A main message is that, for several natural classes of tests, small-bias distributions are no better than bounded uniformity, i.e., we provide new information about Question 3, and answer Question 4 and Question 6.

To set the stage, we start with showing that $k$-wise uniformity plus noise does fool symmetric functions with error $2^{-ck}$. Note that noise is necessary, for parity is not fooled even by $(n-1)$-wise uniformity. And even for threshold tests, the error would be polynomial [26, 10] rather than exponential in $k$.

**Theorem 7.** Let $D$ be a distribution on $\{-1, 1\}^n$ that is either

(i) $(2k)$-wise uniform, or

(ii) $(ck/n)^k$-biased.

Let $f: \{-1, 1\}^n \to [-1, 1]$ be symmetric. Then $|\mathbb{E}[f(U)] - \mathbb{E}[f(D \cdot N_\rho)]| \leq c \cdot (e\rho)^{k/2}$.

Theorem 7.i follows from [29] when $k \geq c \log n$, but their proof does not apply to smaller $k$. Our result applies to any $k$, and this will be critical.

Theorem 7.ii follows from Theorem 7.i via the following simple extension of Lemma 2, which we establish by taking noise into account. (A direct application of Lemma 2 would give a larger error of $2^{-ck}$.)

**Lemma 8.** Let $D$ be an $(\varepsilon, k)$-biased distribution on $\{-1, 1\}^n$. Then $D \cdot N_\rho$ is $((\frac{e \rho n}{k})^{k/2} \varepsilon)$-close to a $k$-wise uniform distribution in total variation distance.

A natural question is whether larger bias suffices in Theorem 7.ii. A main result in this work is that it does not, even for threshold tests. The best possible bound for small-bias distributions is in fact obtained by combining Theorem 7.i with the generic Lemma 8.
Theorem 9. There exists a $((ck)/n)^k$-biased distribution $D$ such that $\Pr[1^T (D \cdot N_\rho)] \geq 2\sqrt{kn} \geq \Pr[B \geq 2\sqrt{kn}] + (c\rho)^{2k}$ for every $\rho \in [0, 1]$.

In fact, the distribution $D$ in Theorem 9 (and in Theorem 10 below) is simultaneously $(2k - 1)$-wise uniform.

Theorem 9 gives a negative answer to Question 4. Specifically, setting $\rho$ to be a constant and $k = \log n$ we obtain bias $1/n^\omega(1)$ but the error is $\geq 1/n^{c\rho}$.

Note that our negative answer holds even with noise, while an answer to Question 4 was not known even for plain small-bias distributions. This makes our results stronger. Moreover, we do not know of a simpler proof if one does not care about noise. Indeed, we obtained several different proofs of essentially Theorem 9, see [25]. In all these proofs (including the one presented here) the small-bias distribution $D$ itself can be written as $D := D' \cdot N_\rho$, that is, by adding noise to another distribution. Further adding noise to $D$ then comes at little cost, as already pointed out in [41], see Claim 22. We also mention that some of these proofs cover wider range of parameters, and provide new information even for bounded uniformity. We refer to [25] for more on this.

Combining Theorem 9 with Theorem 7, one immediately obtains a smoothed threshold test which distinguishes some $n^{-k}$-bias distribution from any $ck$-wise uniform distribution, answering Question 3 for such tests.

For general symmetric tests and the same distribution $D$, we prove a stronger result improving on the classic line of works in [6, 4, 50] and finally matching Lemma 8.

Theorem 10. There exists a $((ck)/n)^k$-biased distribution $D$ such that for every $\rho \in [0, 1]$ the following holds. There exists a symmetric function $f : \{-1, 1\}^n \rightarrow \{0, 1\}$ such that for every $(2k)$-wise uniform distribution $D_{2k}$,

$$\mathbb{E}[f(1^T (D \cdot N_\rho))] \geq \mathbb{E}[f(D_{2k})] + \left(\frac{c\rho}{\log(1/2\rho)}\right)^{2k}.$$

Again, this result was not known even without noise. Note that Theorem 10 implies the same separation without noise simply setting $\rho := 1$. But the other way around is not clear.

An interesting question is whether one could prove a single result that implies both Theorem 10 and Theorem 9.

From Theorem 9, we derive several consequences on small-space computation and small-depth circuits.

One-way small space. We give a negative answer to Question 6.

Corollary 11. For any $\rho \in (0, 1]$, there is a distribution $D$ on $\{-1, 1\}^n$ that is $n^{-c\log_1/\rho}n^{-k}$-biased and a threshold-of-thresholds $T : \{-1, 1\}^n \rightarrow \{0, 1\}$ such that $\mathbb{E}[(T(U)) - \mathbb{E}[T(D \cdot N_\rho)] \geq 1/3$. In particular, there is a read-once branching program $T$ of width $n^{c\rho}$ for which the inequality holds.

Corollary 11, in combination with Lemma 8, matches a result in [29], which shows that the error is $\rho^{2k}$ for $k$-wise uniform when $k \geq c\log_1/\rho n$.

Proof of Corollary 11 from Theorem 9. We divide the input into $\sqrt{n}$ blocks, and in each block sample an independent copy of the $n^{-k/2}$-biased distribution from Theorem 9 on $\sqrt{n}$ bits. The resulting distribution has the required properties, since the bias of a test that spans multiple blocks equals the product of the biases in each block.
In each block, a suitable threshold tells \( D \cdot N_\rho \) from uniform with advantage \((c \rho)^k \geq n^{-0.1}\) for \( k = c \log_{1/\rho} n \). A threshold of \( \sqrt{n} \) such blocks is sufficient to boost the advantage to constant.

Finally, this threshold-of-thresholds computation can be implemented with \( c \log n \) bits of space, by simply maintaining two counters.

What may have made this problem harder is that it was not clear what distinguishing bound one should expect in Theorem 9. One may be tempted to aim for larger advantage, perhaps independent from \( k \). But as we showed in Theorem 7, this is false: \( k \)-wise uniformity plus noise fools thresholds with error \( 2^{-ck} \). One can then ask if \( k \)-wise uniformity fools with error \( 2^{-ck} \) more general classes of tests, like threshold of thresholds. Corollary 11 shows this is also false.

**Constant-depth circuits.** Next we discuss a negative result for fooling the circuit class \( \text{AC}^0 \). It is known that polylogarithmic independence or quasi-polynomial bias fools \( \text{AC}^0 \) [7, 53, 15, 57], and these bounds are nearly tight. But despite attempts [41] it was not known if logarithmic uniformity plus noise, or polynomial bias plus noise suffices. We show that bias \( n^{-\omega(1)} \) is necessary.

**Corollary 12.** For any \( \rho \in (0, 1] \) there is a distribution \( D \) on \( \{-1, 1\}^n \) that is \( n^{-c_\rho \log \log n} \)-biased and an \( \text{AC}^0 \) circuit \( C \) of size \( n^c \) and depth \( c \) such that \( \mathbb{E}[C(U)] - \mathbb{E}[C(D \cdot N_\rho)] \geq 1/3 \).

The proof is similar to before, except we take blocks of polylogarithmic length, and set \( k = c_\rho \log \log n \). The threshold in each block can be computed in \( \text{AC}^0 \) since it’s only on polylogarithmic number of bits. By our setting of \( k \), the advantage in each block is polylogarithmic, and so computing approximate majority [2] (cf. [61]) suffices to have constant advantage.

**Sum of small-bias distributions.** A next natural question is whether our counterexamples can be extended to the xor of two small-bias distributions. We show that they cannot. Specifically, our small-bias distributions are symmetric, and we show that the sum of two such distributions fools any function (symmetric or not).

**Theorem 13.** Let \( D_1 \) and \( D_2 \) be two independent \( n^{-20k} \)-biased, symmetric distributions on \( \{-1, 1\}^n \). Then \( |\mathbb{E}[f(D_1 \cdot D_2)] - \mathbb{E}[f]| \leq c_k (n^{-0.3k}) \) for any function \( f : \{-1, 1\}^n \rightarrow [-1, 1] \).

In fact, we prove stronger results. We show that to fool any symmetric function it suffices for one of the two distributions to be symmetric (Corollary 33). In fact, this holds even if one of the two distributions is any fixed string \( x \) with \( 1^Tx \leq n^{0.99} \) (Theorem 34); and we complement this with a result showing that the result is false if \( 1^Tx \) is large. This is in Section 4.

**Typical shifts.** We generalize and simplify the proof of a result by Bazzi [9]. We first discuss his result. Let \( C \subseteq \{0, 1\}^n \) be a binary linear code with minimum distance \( k + 1 \) and maximum distance \( n - k - 1 \). Let \( U_{C^\perp} \) be the uniform distribution on the dual code of \( C \), and \( u \sim \{0, 1\}^n \) be a uniform string. Bazzi [9] showed that for most shifts \( u \), the distribution \( u + U_{C^\perp} \) fools any symmetric function \( f : \{0, 1\}^n \rightarrow \{0, 1\} \):

\[
\mathbb{E}_{u}[|\mathbb{E}[f(u + U_{C^\perp})] - \mathbb{E}[f]|] \leq (k/n)^c .
\]
It follows from the distance properties of \( C \) that \( U_C \) fools all parity tests of size at most \( k \) and at least \( n - k \) (with no error). We show that in fact the conclusion above holds for every distribution \( D \) that fools such parity tests, without requiring the distribution to be linear.

\[ \text{Theorem 14.} \quad \text{Let } D \text{ be a distribution on } \{-1,1\}^n \text{ such that } E[D^S] = 0 \text{ for every subset } S \in [1,k] \cup [n-k,n], \text{ and } u \sim \{0,1\}^n \text{ be a uniform string. For every symmetric function } f : \{-1,1\}^n \to [-1,1], \]

\[ E_u[|E[f(u \cdot D)] - E[f]|] \leq 6(k/n)^{k-1}. \]

For context, we note that the condition on fooling large parity tests is necessary, as otherwise, one can consider the uniform distribution \( D \) on strings with the parity 0 (say), which is \((n-1)\)-wise uniform, and for every shift \( u \), the parity of \( u + D \) (which is a symmetric function) is simply the parity of \( u \).

Also, note that no fixed shift \( u \) suffices, for else one can shift \( D \) by this \( u \) and give a counterexample. This does not contradict the results discussed above about shifting symmetric small-bias distributions because the shift of a symmetric distribution is not in general symmetric.

### 1.2 Krawtchouk polynomials

All our results rely on bounds for the (shifted) Krawtchouk polynomials \( \mathcal{K} \), which can be defined by

\[ \mathcal{K}(k,t) := \sum_{|S|=k} z^S, \]

where \( z \in \{-1,1\}^n \) is any string such that \( 1^\top z = t \), and \( z^S \) is the product of the bits of \( z \) indexed by \( S \). It can be shown that this is a degree-\( k \) polynomial in \( t \).

This is a classic quantity (cf. [39]) and the bounds we need do not seem well known.

To illustrate the bounds we find it convenient to define the normalized version of \( \mathcal{K} \),

\[ N\mathcal{K}(k,t) := \frac{\mathcal{K}(k,t)}{\binom{n}{k}} \]

and its “with replacement” counterpart

\[ NR(k,t) := \frac{E}{f : [k] \to [n]} \left[ \prod_{i \in [k]} z_{f(i)} \right] = \frac{(t/n)^k}{(t/n)^k}, \]

where \( f \) is uniform.

Note that \( N\mathcal{K} \) is the same as \( NR \) conditioned on \( f \) having no collisions – via the correspondence \( S = \{ f(i) : i \in [k] \} \) – which is the same as saying that the images of \( f \) are picked from \([n]\) without replacement.

The bounds on \( N\mathcal{K} \) (and hence \( \mathcal{K} \)) can now be understood as approximations to \( NR(k,t) \).

First we prove a lower bound, needed for Theorem 9. The proof is short and follows from known results on Krawtchouk polynomials. However, we are unable to find the result we need in the literature.

\[ \text{Claim 15.} \quad N\mathcal{K}(k,t) \geq \left( \frac{t}{2n} \right)^k = NR(k,t)/2^k \text{ for } t \geq 2\sqrt{k(n-k)}. \]
For Theorem 10 we need an upper bound. We could use a bound which to our knowledge appeared first in [11]. Since the proof in the latter is somewhat technical, we also give a new simple proof of a stronger bound, stated next, building on the recent work by Tao [59]. We could also use [11] for Theorem 7, but we would get a bound of the form $(a_0)^k$ for an unspecified constant $a$. The stronger bound in Corollary 16 proved here gives a better dependence on $\rho$ and gets us closer to the natural bound of $\rho^k$, which is currently not clear.

**Corollary 16.** For every $1 \leq k \leq n$, we have $|N\mathcal{K}(k,t)| \leq \left( \frac{k}{n} + \frac{\alpha}{\beta} \right)^{k/2}$.

Note this is similar to $NR(k,t)$ except for the extra term $k/n$.

For other results we need additional bounds which hold in regimes where the above bounds are loose, such as when $k$ is close to $n/2$ and $t$ is close to 0. To illustrate, let $n$ be even and $1^T x = t := 0$, corresponding to $x \in \{-1,1\}^n$ being a balanced string. Note that $\mathcal{K}(k,0)$ is the $k$-th coefficient of the polynomial $(1 - x^2)^{n/2}$, which is $(-1)^{k/2}(\frac{n/2}{k/2})1(k \text{ is even})$. In this case, Corollary 16 gives an upper bound of $\binom{n}{k}(k/n)^{k/2}$. In particular, when $k = n/2$, the bound is roughly $2^{n/4}$. By contrast, the bound given next by Proposition 17 is $2^{\frac{1}{2}H(k/n)}$, which when $k = n/2$ becomes $2^{n/2}$.

**Proposition 17.** Let $k = \beta n$ and $t = (1 - 2\alpha)n$. We have $\log_2 |\mathcal{K}(k,t)| \leq \frac{1}{2} \left( 1 + H(\beta) - H(\alpha) \right)$, where $H(\alpha) = -\alpha \log_2(\alpha) - (1 - \alpha) \log_2(1 - \alpha)$ is the binary entropy function.

A similar bound also appears in [51, Lemma 2.1]. Using the estimate $H(1/2 + \gamma) \geq 1 - 4\gamma^2$ for $\gamma \in [0,1/2]$, we have the following corollary.

**Corollary 18.** $|\mathcal{K}(k,t)| \leq 2^{\frac{1}{2}(H(\frac{k}{n}) + \frac{\beta}{2})}$.

In Section 6 we prove bounds more general than the above.

## 2 Small-bias plus noise is far from bounded uniformity

In this section we prove Theorems 9 and 10. We build on the work by O’Donnell and Zhao [50]. In particular, we use the same distribution $D$. However, jumping ahead, our analyses differ from [50] in three ways, each of which is critical for us:

1. while we analyze the same symmetric test in Theorem 10, we use a new and explicit threshold test in Theorem 9;
2. the distinguishing advantages in Theorems 9 and 10 are explicit and stronger. This relies on our use of (and bounds for) Krawtchouk polynomials, instead of the Hermite approximation in [50];
3. we take noise into account.

We now define $D$ and derive some properties of it. Then in the next subsections the theorems are proved in turn.

**Definition 19.** For a parameter $\alpha \in [0, \frac{1}{5\epsilon}]$, define $D_{\alpha} : \{-1,1\}^n \to \mathbb{R}$ to be

$$D_{\alpha}(x) := 2^{-n} \left( 1 + \alpha^k \binom{n}{2k} - \frac{\alpha}{2k} \sum_{|S|=2k} x^S \right) \text{ for every } x \in \{-1,1\}^n.$$  

Note that the right hand side is the Fourier transform of $D_{\alpha}$, and thus $2^n \hat{D}_{\alpha}(\emptyset) = \sum_{x \in \{-1,1\}^n} D_{\alpha}(x) = 1$. We now show that for $\alpha \leq 1/(5\epsilon)$, we have $D_{\alpha}(x) \geq 0$ for every $x \in \{-1,1\}^n$ and thus it is a distribution.
\[\text{Claim 20.} \quad \text{For } \alpha \leq 1/(5e), \text{ we have } D_\alpha(x) \geq 0 \text{ for every } x.\]

**Proof.** The key observation is that as a degree- \((2k)\) polynomial in \(t\), the zeros of \(K(2k, t)\) all lies within \(|t| \leq 2\sqrt{(2k - 1)(n - 2k + 2)} \leq 2\sqrt{2kn} \) (see also [39, Section 5]). As \(2k\) is even, we know that when \(x\) is the all-1 or all-(-1) string (i.e., \(1^T x \in \{n, -n\}\)), we have \(K(2k, 1^T x) := \sum_{|S|=2k} x^S > 0\). So \(K(2k, 1^T x)\) can only be negative when \(|1^T x| \leq 2\sqrt{2nk}\).

In this interval, using Corollary 16 and \(K\) is even, we know that when \(|1^T x| \leq 2\sqrt{2nk}\), we have

\[
\alpha^k \left( \frac{n}{2k} \right)^{-\frac{1}{2}} \left| \sum_{|S|=2k} x^S \right| \leq \alpha^k \left( \frac{n}{2k} \right)^{-\frac{1}{2}} \left( \frac{10k}{n} \right)^k \leq (5e\alpha)^k \leq 1.
\]

By the above, \(D_\alpha\) is a well-defined distribution whenever \(\alpha \leq 1/(5e)\). The following claim is immediate.

\[\text{Claim 21.} \quad \text{For } \alpha \leq 1/(5e), D_\alpha \text{ is a distribution that is } (2k-1)\text{-wise uniform, } \alpha^k \left( \frac{n}{2k} \right)^{-1/2}\text{-biased, and } (\alpha e^3/2)^k\text{-close to } (2k)\text{-wise uniform.}\]

**Proof.** The first two properties follow directly from the definition of \(D_\alpha\), that is, for every nonempty \(S\), we have \(|E[D_\alpha]| = 2^n |\hat{D}_\alpha(S)| = \alpha^k \left( \frac{n}{2k} \right)^{-1/2} \mathbb{1}(|S| = 2k)\). The closeness to \((2k)\)-wise uniform follows directly from Lemma 2.

Observe that the family \(\{D_\alpha : \alpha \geq 0\}\) is closed under adding noise, as shown in the following claim.

\[\text{Claim 22.} \quad D_\alpha \cdot N_\rho = D_{\alpha \rho^2} \text{ for every } \rho \in [0, 1].\]

**Proof.** Observe that \(N_\rho\) dampens each size- \((2k)\) (Fourier) coefficient of \(D_\alpha\) by a factor of \(\rho^{2k}\).

To see this, note that \(N_\rho(x) = \frac{1}{2}(1 + \rho x_i)\), and thus

\[N_\rho(x) = 2^{-n} \left( 1 + \sum_{S} \rho^{S(x)} \right).
\]

By Plancherel’s theorem, each Fourier coefficient of the convolution \(D_\alpha \cdot N_\rho\) is the product of the coefficient of \(D_\alpha\) and \(N_\rho\). So we have

\[(D_\alpha \cdot N_\rho)(x) = 2^{-n} \left( 1 + \rho^{2k} \alpha^k \left( \frac{n}{2k} \right)^{-\frac{1}{2}} \sum_{|S|=2k} x^S \right) = D_{\alpha \rho^2}(x).
\]

### 2.1 Distinguishing \(D_\alpha\) from uniform with a threshold

We now show that a specific threshold distinguishes \(D_\alpha\) from the uniform distribution. First, we establish the following claim showing that \(D_\alpha\) always puts more mass than \(U\) on unbalanced strings.

\[\text{Claim 23.} \quad \Pr[1^T D_\alpha = t] \geq \Pr[B = t] \cdot (1 + \left(\frac{\alpha t^2}{4kn}\right)^k) \text{ for every } t \geq 2\sqrt{kn} \text{ and } \rho \in [0, 1].\]

**Proof.** By our lower bound on Krawtchouk polynomials (Claim 15), we have

\[
\Pr[1^T D_\alpha = t] = \Pr[B = t] \left( 1 + \alpha^k \left( \frac{n}{2k} \right)^{-1/2} K(2k, t) \right)
\]

\[
\geq \Pr[B = t] \left( 1 + \alpha^k \left( \frac{n}{2k} \right)^{1/2} \left( \frac{t}{2^n} \right)^{2k} \right)
\]

\[
\geq \Pr[B = t] \left( 1 + \left(\frac{\alpha t^2}{4kn}\right)^k \right).
\]
Theorem 9 then easily follows from Claim 23 by summing over all the points at the tail, and then setting \( \alpha \) to be \( \rho^2/(5\epsilon) \).

**Proof of Theorem 9.** From Claim 23, it follows that

\[
\Pr\left[1^T D_\alpha \geq 2\sqrt{kn}\right] \geq \sum_{t \geq 2\sqrt{kn}} \Pr[B = t] \cdot \left(1 + \left(\frac{\alpha t^2}{4kn}\right)^k\right)
\]

\[
\geq \Pr[B \geq 2\sqrt{kn}] \cdot \left(1 + \left(\frac{\alpha(2\sqrt{kn})^2}{4kn}\right)^k\right)
\]

\[
\geq \Pr[B \geq 2\sqrt{kn}] + 2^{-ck} \cdot \alpha^k,
\]

where the last inequality is because by tail bounds for the binomial distribution (cf. [1]) we have \( \Pr[B \geq 2\sqrt{kn}] \geq 2^{-ck} \). The theorem then follows from setting \( \alpha \) to \( \rho^2/(5\epsilon) \), and noting that \( D_{1/(5\epsilon)} \cdot N_\rho = D_{\rho^2/(5\epsilon)} \) by Claim 22.

\[\blacksquare\]

## 2.2 Distinguishing \( D_\alpha \) from bounded uniformity with a symmetric test

In this section, we prove Theorem 10. We start with a claim showing that it suffices to consider bounded symmetric functions instead of Boolean symmetric test.

\[\triangleright \text{Claim 24.} \quad \text{Let } D_1, D_2 \text{ be any distributions on } \{-1,1\}^n. \text{ Suppose there is a symmetric function } f: \{-1,1\}^n \to [-1,1] \text{ such that } \mathbb{E}[f(D_1)] \geq \mathbb{E}[f(D_2)] + \epsilon. \text{ Then there exists a symmetric Boolean function } f': \{-1,1\}^n \to \{-1,1\} \text{ such that } \mathbb{E}[f'(D_1)] \geq \mathbb{E}[f'(D_2)] + \epsilon.\]

Proof. Define \( g: \{-n, \ldots, n\} \to [-1,1] \) so that \( g(1^T x) := f(x) \). Considering the randomized function \( g: \{-n, \ldots, n\} \to \{-1,1\} \) defined by

\[
g(w) := \begin{cases} 
1 & \text{with probability } \frac{1 + g(w)}{2} \\
-1 & \text{with probability } \frac{1 - g(w)}{2}.
\end{cases}
\]

As \( f \) is symmetric, we have

\[
\mathbb{E}_g\left[\mathbb{E}[g(1^T D_1)]\right] = \mathbb{E}[f(D_1)] \geq \mathbb{E}[f(D_2)] + \epsilon = \mathbb{E}_g\left[\mathbb{E}[g(1^T D_2)]\right] + \epsilon,
\]

and so by averaging, there must be a choice \( g' \) of \( g \) such that \( \mathbb{E}[g'(1^T D_1)] \geq \mathbb{E}[g'(1^T D_2)] + \epsilon \). Defining \( f': \{-1,1\}^n \to \{-1,1\} \) by \( f'(x) := g'(1^T x) \) proves the claim.

\[\triangleright\]

We now define our symmetric test. For a sufficiently small constant \( \alpha \), let \( \beta := \frac{100}{\log(1/\alpha)} \). Define the homogeneous degree-\( k \) polynomial \( p_\beta: \{-1,1\}^n \to \mathbb{R} \) by

\[
p_\beta(x) := \beta^k \left(\frac{n}{2k}\right)^{-\frac{1}{2}} \sum_{|S| = 2k} x^S = 2^k D_\beta(x) - 1.
\]

Let \( f_\beta \) be its truncation so that it is bounded by 1, that is, we define \( f_\beta: \{-1,1\}^n \to [-1,1] \) by \( f_\beta(x) := \min\{1, p_\beta(x)\} \). As \( \alpha \) is sufficiently small, so is \( \beta \). Thus, by Claim 20, we have \( f_\beta(x) \geq -1 \) and so \( f_\beta(x) \in [-1,1] \) for every \( x \in [-1,1] \).

\[\triangleright \text{Claim 25.} \quad \mathbb{E}[f_\beta(D_\alpha)] - \mathbb{E}[f_\beta(D)] \geq (\alpha \beta)^k/2 \text{ for any } k\text{-wise uniform distribution } D.\]
Proof. As \(p_\beta\) has degree-(2k), for any (2k)-wise uniform distribution \(D\), we have \(\mathbb{E}[f_\beta(D)] \leq \mathbb{E}[p_\beta(D)] = \mathbb{E}[p_\beta(U)]=0\). Note that we can write \(f_\beta(x)\) as \(p_\beta(x)-(p_\beta(x)-1)1(p_\beta(x)>1)\), and so
\[
\mathbb{E}[f_\beta(D_\alpha)] = \mathbb{E}[p_\beta(D_\alpha)] - \mathbb{E}[(p_\beta(D_\alpha)-1)1(p_\beta(D_\alpha)>1)].
\]
(1)
To bound \(\mathbb{E}[f_\beta(D_\alpha)]\) from below, we will compute \(\mathbb{E}[p_\beta(D_\alpha)]\) and then bound \(\mathbb{E}[(p_\beta(D_\alpha)-1)1(p_\beta(D_\alpha)>1)]\) from above.

Observe that
\[
\mathbb{E}\left[\sum_{|S|=2k} U^S\right] = \sum_{|S|=2k} \mathbb{E}[U^S] = 0
\]
\[
\mathbb{E}\left[\left(\sum_{|S|=2k} U^S\right)^2\right] = \sum_{|S|=2k} \mathbb{E}[U^{S\triangle T}] = \binom{n}{2k}
\]
\[
\mathbb{E}\left[\left(\sum_{|S|=2k} U^S\right)^3\right] = \sum_{|S|=2k} \mathbb{E}[U^{S\triangle T\triangle R}] = \binom{n}{2k} \frac{2k}{k} \binom{n-2k}{k},
\]
where the last equality is because the number of subsets \(S, T, R \subseteq [n]\) of size 2k that satisfy \(S \triangle T = R\) is \(\binom{2k}{2k} \binom{n-2k}{k}\).

We have
\[
\mathbb{E}[p_\beta(D_\alpha)] = \sum_{x \in \{-1,1\}^n} D_\alpha(x) \mathbb{E}[p_\beta(x)]
\]
\[
= \sum_{x \in \{-1,1\}^n} 2^{-n} \left(1 + \alpha k \binom{n}{2k} - \frac{1}{2} \sum_{|S|=2k} x^S\right) \left(\beta k \binom{n}{2k} - \frac{1}{2} \sum_{|S|=2k} x^S\right)
\]
\[
= (\alpha \beta) k \binom{n}{2k}^{-1} \mathbb{E}\left[\left(\sum_{|S|=2k} U^S\right)^2\right] 
\]
\[
= (\alpha \beta) k \binom{n}{2k}^{-1} \left(\binom{n}{2k} - \binom{n-2k}{k}\right). \tag{2}
\]
We also have
\[
\mathbb{E}[p_\beta(D_\alpha)^2] = 2^{-n} \sum_{x \in \{-1,1\}^n} \left(1 + \alpha k \binom{n}{2k} - \frac{1}{2} \sum_{|S|=2k} x^S\right)^2 \left(\beta k \binom{n}{2k} - \frac{1}{2} \sum_{|S|=2k} x^S\right)^2
\]
\[
= \beta^{2k} \binom{n}{2k}^{-1} \mathbb{E}\left[\left(\sum_{|S|=2k} U^S\right)^2\right] + (\alpha \beta^2) k \binom{n}{2k}^{-1} \mathbb{E}\left[\left(\sum_{|S|=2k} U^S\right)^3\right]
\]
\[
= \beta^{2k} + (\alpha \beta^2) k \binom{2k}{k} \binom{n-2k}{k} 
\]
\[
\leq \beta^{2k} + (8 \alpha \beta^2)^k,
\]
\[
\leq 1, \tag{3}
\]
where the first inequality is because \(\binom{n-2k}{k} \leq \binom{n}{k} \binom{n-k}{k}^2 = \binom{n}{k}^2 \binom{2k}{k}^2\), using the identity \(\binom{n}{m} \binom{n-m}{k-m} = \binom{n}{k} \binom{k}{m}\). The last inequality is for small enough \(\alpha\).
We rewrite and bound the second term using Cauchy–Schwarz as follows:

\[ \Pr[p_\beta(D_\alpha) > 1]^{\frac{1}{2}} \leq 1 \cdot \Pr[p_\beta(D_\alpha) > 1]^{\frac{1}{2}}. \]  

(4)

We will show that

\[ \Pr[p_\beta(D_\alpha) > 1]^{\frac{1}{2}} \leq e^{-\left(\frac{1}{2\beta} - 1\right)\frac{1}{2}}. \]  

(5)

Therefore

\[ \Pr[p_\beta(D_\alpha)] - \E[f_\beta(D)] \geq (\alpha\beta)^k / 2 \]

for any \((2k)\)-wise uniform \(D\), proving the claim assuming (5) holds.

It remains to prove (5). Suppose \(|p_\beta(x)| > 1\). Then by its definition and Corollary 16, it must be the case that

\[ 1 \leq \beta^k \left( n \left( \frac{2k}{n} \right)^{\frac{3}{2}} \sum_{|S|=2k} x^S \right) \leq \beta^k \left( n \left( \frac{2k}{n} + \frac{(1^T x)^2}{n^2} \right)^k \right) \leq (\epsilon \beta)^k \left( 1 + \frac{(1^T x)^2}{kn} \right)^k, \]

which implies \(x \in E_\beta := \{ x \in \{-1, 1\}^n : (1^T x)^2 \geq (\frac{1}{c^2} - 1)2kn \}\). Jumping ahead, we will use below that by Hoeffding’s inequality, we have \(\Pr[U \in E_\beta] \leq e^{-k(\frac{1}{\epsilon^2} - 1)}\). In the meanwhile, we use the implication just noted to write

\[ \Pr[p_\beta(D_\alpha) > 1] \leq \Pr[D_\alpha \in E_\beta] = 2^{-n} \sum_{x \in E_\beta} \left( 1 + \alpha^k \left( n \left( \frac{2k}{n} \right)^{\frac{3}{2}} \sum_{|S|=2k} x^S \right) \right) \]

\[ = \Pr[U \in E_\beta] + \alpha^k \left( n \left( \frac{2k}{n} \right)^{\frac{3}{2}} 2^{-n} \sum_{x \in E_\beta} \sum_{|S|=2k} x^S. \right) \]

(6)

We rewrite and bound the second term using Cauchy–Schwarz as follows:

\[ \alpha^k \left( n \left( \frac{2k}{n} \right)^{\frac{3}{2}} \E \left[ \sum_{|S|=2k} U^S \cdot 1(U \in E_\beta) \right] \leq \alpha^k \left( n \left( \frac{2k}{n} \right)^{\frac{3}{2}} \E \left[ \left( \sum_{|S|=2k} U^S \right)^2 \right]^{\frac{1}{2}} \cdot \Pr[U \in E_\beta] \right)^{\frac{1}{2}} \]

\[ = \alpha^k \cdot \Pr[U \in E_\beta]^{\frac{1}{2}}. \]

Therefore

\[ (6) \leq \Pr[U \in E_\beta]^{\frac{1}{2}} \left( \Pr[U \in E_\beta]^{\frac{1}{2}} + \alpha^k \right) \leq e^{-\left(\frac{1}{2\beta} - 1\right)\frac{1}{2}} \cdot (1/2 + 1/2). \]

This proves (5).

\(<\)

**Proof of Theorem 10.** For any \(\rho \in (0, 1]\), by Claim 22 we have \(D_\alpha \cdot N_\rho = D_{\rho^2}\). So we can take \(\alpha\) to be \(c\rho^2\), and thus \(\beta = c / \log(1/2\rho)\). By Claim 25, the distinguishing advantage is at least \((c\rho^2 / \log(1/2\rho))^k\).

\(<\)
3 Bounded uniformity plus noise fools symmetric tests

Here we prove Theorem 7. The starting observation for the proof of this theorem (and also of Theorems 13 and 14) is that the Fourier expansion of any symmetric function is a linear combination of the Krawtchouk polynomials $\overline{K}(\ell, 1^\top x) := \sum_{|S| = \ell} x^S$ weighted by the coefficients $\hat{f}(S)$. As $k$-wise uniformity fools all parities of size at most $k$, it suffices to consider the $\ell > k$ terms. While $\overline{K}(\ell, 1^\top x)$ can be as large as $\binom{n}{\ell}$ on the all-1 string, it follows from Cauchy–Schwarz and Parseval’s identity, we have $\overline{E}[\overline{K}(\ell, 1^\top x)]^2$ is bounded by $\binom{n}{\ell}^{-1/2}$. Moreover, a simple argument (Fact 26) shows that $|\hat{f}(\ell)|$ is bounded by $\binom{n}{\ell}^{-1/2}$, the reciprocal of the upper bound on $\overline{E}[K(\ell, 1^\top x)]$, and so their product is at most 1, which is then dampened to $\rho^\ell \leq \rho^k$ by noise.

To make this argument go through, we use Corollary 16 to show that $\overline{K}(\ell, 1^\top x)$ is nearly-balanced, which holds with high probability under $k$-wise uniformity (Corollary 28).

We start by proving a few useful facts about symmetric functions and distributions.

Fact 26. Let $f : \{-1, 1\}^n \rightarrow [-1, 1]$ be any symmetric function. For every $S \subseteq [n]$ of size $\ell$, we have (1) $\overline{E}[f(S)] = \hat{f}(\ell)$ and (2) $|\hat{f}(\ell)| \leq \binom{n}{\ell}^{-1/2}$.

Proof. (1) is clear. To see (2), by Cauchy–Schwarz and Parseval’s identity, we have

$$\left(\frac{n}{\ell}\right)|\hat{f}(\ell)| = \left|\sum_{|S| = \ell} \hat{f}(S)\right| \leq \left(\sum_{|S| = \ell} \hat{f}(S)^2\right)^{1/2} \leq \left(\frac{n}{\ell}\right)^{1/2} \overline{E}[f(U)^2] \leq \left(\frac{n}{\ell}\right)^{1/2}. \ \Box$$

We also need the following well-known moment bounds for $k$-wise uniform distributions. For a short proof see [14, Lemma 32].

Lemma 27. Let $D$ be a $(2k)$-wise uniform distribution on $\{-1, 1\}^n$. Then $\overline{E}[(\sum_{i=1}^n D_i)^{2k}] \leq \sqrt{2} \binom{2kn}{k}$.

By Markov’s inequality, this implies the following tail bound.

Corollary 28. Let $D$ be a $(2k)$-wise uniform distribution on $\{-1, 1\}^n$. For every integer $\ell > 0$, we have

$$\Pr[|1^\top D| \geq \ell] \leq \sqrt{2} \left(\frac{2kn}{\ell^2}\right)^k.$$

The following fact says that a distribution remains close to itself after conditioning on any high probability event.

Fact 29. Let $D$ be any distribution on $\{-1, 1\}^n$ and $E$ be any event. Then the conditional distribution $D \mid E$ is $(1 - \Pr[E])$-close to $D$.

Proof. Let $\overline{E}$ be the complement of $E$. For every Boolean test $g : \{-1, 1\}^n \rightarrow \{0, 1\}$, we have

$$\overline{E}[g(D)] = \overline{E}[g(D \mid E)](1 - \Pr[\overline{E}]) + \overline{E}[g(D \mid \overline{E})] \Pr[\overline{E}] = \overline{E}[g(D \mid E)] + (\overline{E}[g(D \mid E)] - \overline{E}[g(D \mid E)]) \Pr[\overline{E}].$$

So $|\overline{E}[g(D)] - \overline{E}[g(D \mid E)]| \leq \Pr[\overline{E}]$, as $|\overline{E}[g(D \mid E)] - \overline{E}[g(D \mid E)]| = |\overline{E}[g(D \mid E)]| = 1$.
Proof of Theorem 7. Define \( G := \{ x \in \{ -1, 1 \}^n : |\sum_{i=1}^n x_i| \leq \sqrt{nk/3p} \} \). We write \( f := f_{\leq k} + f_{>k} \), where \( f_{\leq k}(x) = \sum_{|S| \leq k} \hat{f}(S)x^S \), and \( f_{>k}(x) := f(x) - f_{\leq k}(x) = \sum_{|S| > k} \hat{f}(S)x^S \).

For convenience let \( Z := D \cdot N_p \). As \( Z \) is \((2k)\)-wise uniform, we have

\[
E[f] = E[f_{\leq k}(Z)] = E[f_{\leq k}(Z) 1(D \in G)] + E[f_{\leq k}(Z) 1(D \notin G)].
\]

So we can bound the error by

\[
|E[f(Z)] - E[f]| = |E[f_{\leq k}(Z) 1(D \in G)] + E[f_{\leq k}(Z) 1(D \notin G)] - E[f]| \\
\leq |E[f_{\leq k}(Z) 1(D \in G)] - E[f]| + |E[f_{\leq k}(Z) 1(D \notin G)] + E[f_{>k}(Z) 1(D \in G)]| + Pr[D \notin G] \\
\leq |E[f_{\leq k}(Z) 1(D \notin G)]| + |E[f_{>k}(Z) 1(D \in G)]| + Pr[D \notin G]. \tag{7}
\]

We now bound each term individually. By Corollary 28, we have

\[
Pr[D \notin G] \leq \sqrt{2} \cdot \left( \frac{2 \cdot 3p}{e} \right)^k \leq \sqrt{2} \cdot (ep)^k. \tag{8}
\]

We now bound the first term, As \( f_{\leq k}^2 \) has degree \( 2k \), by Parseval’s identity and \((2k)\)-wise uniformity of \( Z \), we have

\[
E[f_{\leq k}(Z)^2] = E[f_{\leq k}(U)^2] = E[f(U)^2] \leq 1.
\]

By Cauchy–Schwarz, the first term in (7) is at most

\[
|E[f_{\leq k}(Z) 1(D \notin G)]| \leq E[f_{\leq k}(Z)^2]^{1/2} Pr[D \notin G]^{1/2} \leq 2^{1/4} \cdot (ep)^{k/2}. \tag{9}
\]

It remains to bound the second term in (7). For every \( x \in G \), we will show that

\[
|E[f_{>k}(x \cdot N_p)]| \leq 7 \cdot (ep)^{k/2}. \tag{10}
\]

Plugging (8)–(10) into (7) gives an error bound of at most \( 10(ep)^{k/2} \), as desired.

We now show (10). As \( E[N_p^S] = \rho^{|S|} \), we have

\[
|E[f_{>k}(x \cdot N_p)]| = \left| \sum_{|S| > k} \rho^{|S|} \hat{f}(S)x^S \right| = \left| \sum_{|S| = \ell} \rho^\ell \sum_{|S| = \ell} \hat{f}(S)x^S \right|.
\]

Applying Fact 26 and Corollary 16, and using the inequality \( (\ell^n) \leq (en/\ell)^\ell \), we have

\[
\left| \sum_{\ell = k+1}^n \rho^\ell \sum_{|S| = \ell} \hat{f}(S)x^S \right| \leq \left| \sum_{\ell = k+1}^n \rho^\ell \cdot \hat{f}(\ell) \sum_{|S| = \ell} x^S \right| \\
\leq \sum_{\ell = k+1}^n \rho^\ell \cdot |\hat{f}(\ell)| \cdot \left| \sum_{|S| = \ell} x^S \right| \\
\leq \sum_{\ell = k+1}^n \rho^\ell \cdot \frac{n}{\ell} \left( \frac{\ell}{n} + \frac{k}{3pn} \right)^{\ell/2} \\
\leq \sum_{\ell = k+1}^n \rho^\ell \cdot e^{\ell/2} \left( 1 + \frac{k}{3pe} \right)^{\ell/2} \\
\leq \sum_{\ell = k+1}^n \left( \rho \left( pe + \frac{e}{3} \right) \right)^{\ell/2} \\
\leq 7 \cdot (ep)^{k/2}
\]

where the last inequality is because we can assume \( \rho \leq 1/e \), as otherwise the conclusion is trivial, and so we have \( pe + e/3 \leq 1 + e/3 \leq 2 \). This shows (10).
4 Shifted symmetric small-bias fools symmetric tests

In this section we prove Theorem 13. The proof follows a similar high-level idea to the proof of Theorem 7, but we trade symmetry for noise, because xor-ing the uniform permutation of a string has a similar effect to adding noise (see Claim 31).

As mentioned in the introduction, we actually prove stronger results about fooling symmetric functions. One can then obtain Theorem 13 by combining Corollary 33 below and the following claim.

Claim 30. Let \( D \) be a symmetric distribution on \( \{0,1\}^n \). If \( |D| \) is \( \varepsilon \)-close to the binomial distribution \( \text{Bin}(n, 1/2) \), then \( D \) is \( \varepsilon \)-close to the uniform distribution.

Proof. We have

\[
\sum_{w=0}^{n} \sum_{|x|=w} \left| 2^{-n} - D(w) \binom{n}{w} \right| = \sum_{w=0}^{n} \left( \binom{n}{w} \right) 2^{-n} - D(w) \binom{n}{w} \right| = \sum_{w=0}^{n} 2^{-n} \binom{n}{w} - D(w) \leq \varepsilon.
\]

In turn, Corollary 33 follows from Theorem 32, showing that any symmetric small-bias distribution xor-ed a nearly-balanced string fools symmetric functions. Then we prove Theorem 34, which is a generalization of Theorem 32 that covers a more general settings of parameters. We complement Theorem 34 with a lower bound (Claim 35).

First we show that the bias of the uniform permutation of a string on parity tests is equal to the normalized Krawtchouk polynomials.

Claim 31. 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

\[
|E[W_t^{[n]\setminus S}]| = |E[W_t^S]| = \frac{K(\ell,t)}{2^{\ell}}.
\]

Proof. The first equality follows from \( |\sum_{|S|=n-\ell} z^S| = |z^n| \sum_{|S|=\ell} z^S| = |\sum_{|S|=\ell} z^S| \). To prove the second inequality, first fix a string \( z \) with \( \sum_{i=1}^{n} z_i = t \). Observe that by symmetry we have \( \sum_{x \sum_{i=1}^{n} x_i = t} x^S = \sum_{x \sum_{i=1}^{n} x_i = t} x^{[\ell]} \) for any \( S \subseteq [n] \) of size \( \ell \), and \( \sum_{|S|=\ell} x^S = \sum_{|S|=\ell} z^S \) for any \( x \in \{-1,1\}^n \) with \( \sum_{i=1}^{n} x_i = t \). Hence,

\[
\binom{n}{\ell} \sum_{x \sum_{i=1}^{n} x_i = t} x^{[\ell]} = \sum_{|S|=\ell} \sum_{x \sum_{i=1}^{n} x_i = t} x^S = \sum_{x \sum_{i=1}^{n} x_i = t} \sum_{|S|=\ell} x^S = \binom{n}{\ell} \sum_{|S|=\ell} z^S.
\]

Rearranging gives

\[
|E[W_t^S]| = \frac{1}{2^{\ell}} \sum_{x \sum_{i=1}^{n} x_i = t} x^{[\ell]} = \frac{1}{2^{\ell}} \sum_{|S|=\ell} z^S = \frac{|K(\ell,t)|}{2^{\ell}}.
\]

4.1 Proof of Corollary 33

Corollary 33 is a straightforward corollary of Theorem 32, which we now prove.

Theorem 32. Let \( D_{\text{sym}} \) be a symmetric \( n^{-20k} \)-biased distribution on \( \{-1,1\}^n \) and \( z \in \{-1,1\}^n \) be any string with \( |\sum_{i=1}^{n} z_i| \leq n^{0.6} \). Then \( |E[f(z \cdot D_{\text{sym}})] - E[f]| \leq O_k(n^{-0.3k}) \).

Note that it is crucial that \( D_{\text{sym}} \) is symmetric, as small-bias distributions are closed under shifts; so every small-bias distribution \( D \) is also a shifted small-bias distribution.
Corollary 33. Let $D_{sym}$ and $D$ be two independent $n^{-20k}$-biased distributions on $\{-1, 1\}^n$, where $D_{sym}$ is symmetric. Then $|\mathbb{E}[f(D_{sym} + D)] - \mathbb{E}[f]| \leq c_k(n^{-0.3k})$ for every symmetric function $f : \{-1, 1\}^n \rightarrow [-1, 1]$.

Also note that $D + D_{sym}$ itself is not necessarily a symmetric distribution.

Proof of Corollary 33. By Lemma 2, $D$ is $n^{-10k}$-close to $(10k)$-wise uniform. So by Corollary 28,

$$\Pr\left[\left|\sum_{i=1}^{n} D_i \right| \geq n^{0.6}\right] \leq \left(\frac{10kn}{n^{1.2}}\right)^{5k} + n^{-10k} \leq O_k(n^{-k}).$$

It follows from Theorem 32 that the error is $O_k(n^{-k}) + O_k(n^{-0.3k}) = O_k(n^{-0.3k})$.

Proof of Theorem 32. Let $\varepsilon := n^{-20k}$ be the bias of $D_{sym}$ and $t := n^{0.6}$. Define $G := \{x \in \{-1, 1\}^n : \sum_{i=1}^{n} x_i \leq t\}$. As $D_{sym}$ is $\varepsilon$-biased, by Lemma 2, it is $\delta$-close to $(30k)$-wise uniform, where $\delta \leq n^{-4k}$. Applying Corollary 28 with our choice of $t = n^{0.6}$ in the definition of $G$, we have

$$\Pr[D_{sym} \notin G] \leq \left(\frac{30kn}{n^{1.2}}\right)^{15k} + \delta \leq O_k(n^{-3k}).$$

Let $D'_{sym}$ be the distribution of $D_{sym}$ conditioned on $D_{sym} \in G$. Note that $D'_{sym}$ remains a symmetric distribution and by Fact 29 is $\Pr[D_{sym} \notin G]$-close to $D_{sym}$, and thus is $\varepsilon'$-biased, where $\varepsilon' := \varepsilon + \Pr[D_{sym} \notin G] \leq O_k(n^{-3k})$. We now write $f := f_{mid} + f_{ends}$, where $f_{mid}(x) := \sum_{k<|S|<n-k} \hat{f}(S)x^S$, and $f_{ends}(x) := f(x) - f_{mid}(x) = \sum_{|S|\in[0,k]|\cup|n-k,n]} \hat{f}(S)x^S$. By the triangle inequality, we have

$$|\mathbb{E}[f(z : D_{sym})] - \mathbb{E}[f]| \leq |\mathbb{E}[f(z : D'_{sym})] - \mathbb{E}[f]| + \Pr[D_{sym} \notin G] + |\mathbb{E}[f_{ends}(z : D'_{sym})]| + O_k(n^{-3k}).$$

We now bound each of the two terms on the right hand side. As $z : D'_{sym}$ is $\varepsilon'$-biased, we have

$$|\mathbb{E}[f_{ends}(z : D'_{sym})] - \mathbb{E}[f]| \leq \sum_{|S|\in[1,k]|\cup|n-k,n]} |\hat{f}(S)|\varepsilon' \leq 2n^k\varepsilon' \leq O_k(n^{-2k}).$$

To bound $|\mathbb{E}[f_{mid}(z : D'_{sym})]|$, let $S$ be any subset of size $\ell$. As $f$ is symmetric, we have $\hat{f}(S) = \hat{f}(|\ell|)$. As $D'_{sym}$ is also symmetric, we have $\mathbb{E}[D'_{sym}] = \varepsilon_\ell$ for some $\varepsilon_\ell$ which only depends on the size of $S$. Hence,

$$|\mathbb{E}[f_{mid}(z : D'_{sym})]| \leq \sum_{\ell=k+1}^{n-k-1} \sum_{|S|\ell} \hat{f}(S) \mathbb{E}[D'_{sym}] z^S = \sum_{\ell=k+1}^{n-k} \hat{f}(|\ell|) \varepsilon_\ell \sum_{|S|\ell} z^S.$$

As $|\sum_{|S|\ell} z^S| = |z|^n |\sum_{|S|\ell} z^S| = |\sum_{|S|\ell} z^S|$, by Fact 26 and Claim 31 we have

$$\sum_{\ell=k+1}^{n-k} \hat{f}(|\ell|) \varepsilon_\ell \sum_{|S|\ell} z^S \leq \sum_{\ell=k+1}^{n-k+1} \left(\hat{f}(|\ell|) |\varepsilon_\ell| \sum_{|S|\ell} z^S\right) \leq 2 \sum_{\ell=k+1}^{n/2} \frac{\mathbb{K}(|\ell|, n^{0.6})^2}{(|\ell|)^{n/2}}.$$

We first bound the partial sum over $\ell$ from $n/4$ to $n/2$. Note that the binary entropy function $H(x) := x \log_2(1/x) + (1-x) \log_2(1/(1-x))$ is increasing on $[0, 1/2]$. In particular, we have
H(1/4) ≥ 4/5 and so 2H(1/4) ≥ 6/5. By Stirling’s approximation, we have \((n) ≥ \frac{1}{e}2^{nH(\frac{1}{2})}\) (see [24] for a proof). Applying Corollary 18 with these facts, we have

\[
\sum_{\ell = \max\{n/4,k\} + 1}^{n/2} \mathcal{K}(\ell, n^{0.6})^2 \left(\binom{n}{\ell}^{3/2}\right) \leq \frac{n}{4} \cdot n^2 \cdot 2^{-n\left(\frac{2}{3}H(\frac{1}{3}) - 1 - n^{-0.8}\right)} \leq 2^{-n/10}.
\] (14)

We now bound the remaining sum (i.e., the partial sum from \(\ell = k + 1\) to \(n/4\)). Using Corollary 16 and the inequality \((n) ≤ \left(\frac{en}{3}\right)\ell\), we have

\[
\sum_{\ell = k + 1}^{n/4} \mathcal{K}(\ell, n^{0.6})^2 \left(\binom{n}{\ell}^{3/2}\right) \leq \sum_{\ell = k + 1}^{n/4} \left(\binom{n}{\ell}^{1/2}\right) \left(\ell n + n^{1.2}\right) \leq \sum_{\ell = k + 1}^{n/4} \left(\frac{en}{3}\right)^{\ell/2} \left(\frac{\ell}{n} + \frac{1}{n^{0.8}}\right)^{\ell}.
\]

Observe that each term in the sum is at most 1/2 of its previous term, and so this sum is bounded by twice the first term, which is at most \(O_k(n^{-0.3k})\). Therefore,

\[
\mathbb{E}[f_{\text{mid}}(z \cdot D_{\text{sym}}^t)] \leq 2^{-n/10} + O_k(n^{-0.3k}) \leq O_k(n^{-0.3k}).
\] (15)

Plugging (13) and (15) in (12) completes the proof.

### 4.2 General case

Theorem 32 is stated for a nearly-balanced shift. We now prove a general bound that holds for any shifts.

**Theorem 34.** There exists a constant \(C\) such that the following holds. Let \(D_{\text{sym}}\) be a symmetric \(\varepsilon\)-biased distribution on \([-1,1]^n\) and \(z \in \{-1,1\}^n\) be any string. Let \(s := \sum_{i=1}^n z_i\). For every positive integer \(k\) and every symmetric function \(f : \{-1,1\}^n \to \{-1,1\}\), we have

\[
\left|\mathbb{E}[f(z \cdot D_{\text{sym}})] - \mathbb{E}[f]\right| \leq C \left(\left(\frac{11 \max\{s, \sqrt{kn}\}}{n}\right)^{k/2} + \left(\frac{e^3 n}{2k}\right)^{k/2} \varepsilon \right).
\]

The following lower bound shows that the dependence on \(s\) in Theorem 34 is necessary.

**Claim 35.** There exists a constant \(C > 0\) such that the following holds. For every integer \(m ≥ 3\), there is a symmetric \(e^{-cn/m^2}\)-biased distribution \(D\) on \(\{0,1\}^n\) such that for every string \(z \in \{0,1\}^n\) of Hamming weight at most \(|m/2| - 1\), there exists a symmetric function \(f : \{0,1\}^n \to \{0,1\}\) such that \(f(D) = 0\) always and \(\Pr[f(U) = 1] ≥ 1/m - e^{-cn/m^2}\).

**Proof of Claim 35.** Let \(D\) be the uniform distribution on \(\{x \in \{0,1\}^n : \sum_i x_i \equiv 0 \mod m\}\). It is known that \(D\) is \(2^{-cn/m^2}\)-biased (see Claim 18 and Lemma 19 in [11] for a proof). Let \(z\) be any string of weight \(|z| ≤ |m/2| - 1\). Consider the symmetric function \(f(x) := \text{1}(|x| ≡ [(m+1)/2] \mod m\). By the triangle inequality, we have \(|D| - |z| ≤ |D + z| ≤ |D| + |z|\), and so \(|D| \neq [(m+1)/2] \mod m\).

On the other hand, it is known that \(\Pr[\text{Bin}(n, 1/2) \equiv [(m+1)/2]] ≥ 1/m - e^{-cn/m^2}\) (see, again, Claim 18 in [11] for a proof).

**Proof of Theorem 34.** We may assume \(k ≤ n/16\) and \(s ≤ n/120\), as otherwise the bound given in the theorem is at least 1. Define \(G := \{x \in \{-1,1\}^n : \sum_{i=1}^n x_i ≤ t\}\), where \(t := t(n,k,s)\) is a parameter to be chosen. As \(D_{\text{sym}}\) is \(\varepsilon\)-biased, by Lemma 2, it is \(\delta\)-close to \((2k)\)-wise uniform, where \(\delta := \left(\frac{e^3 n}{2k}\right)^{k/2} \varepsilon\). Applying Corollary 28, we have

\[
\Pr[D_{\text{sym}} \not\in G] ≤ \sqrt{2} \cdot \left(\frac{2nk}{e^3 n}\right)^{k/2} + \delta.
\] (16)
We write \( f := f_{\text{mid}} + f_{\text{ends}} \), where \( f_{\text{mid}}(x) := \sum_{k+1 < |S| < n-k} \hat{f}(S)x^S \) and \( f_{\text{ends}}(x) := f(x) - f_{\text{mid}}(x) = \sum_{|S| \in [0,k] \cup [n-k,n]} \hat{f}(S)x^S \). For convenience, let \( Z := z \cdot D_{\text{sym}} \). As \( Z \) is \( \varepsilon \)-biased, we have

\[
|E[f_{\text{ends}}(Z)] - E[f]| \leq \sum_{|S| \in [1,k] \cup [n-k,n]} |\hat{f}(S)|\varepsilon \leq 2\left(\frac{\epsilon n^k}{k}\right)^{k/2} \varepsilon \leq \delta.
\]

By the triangle inequality, we have

\[
|E[f_{\text{ends}}(Z) \mathbb{1}(D_{\text{sym}} \in G)] - E[f]| \\
\leq |E[f_{\text{ends}}(Z) \mathbb{1}(D_{\text{sym}} \in G)] - E[f_{\text{ends}}(Z)]| + |E[f_{\text{ends}}(Z)] - E[f]| \\
\leq |E[f_{\text{ends}}(Z) \mathbb{1}(D_{\text{sym}} \in G)] - E[f_{\text{ends}}(Z)]| + \delta \\
= |E[f_{\text{ends}}(Z) \mathbb{1}(D_{\text{sym}} \notin G)]| + \delta. \tag{17}
\]

As \( Z \) is \( \varepsilon \)-biased,

\[
E[f_{\text{ends}}(Z)^2] = \sum_{|S|,|T| \in [0,k] \cup [n-k,n]} \hat{f}(S)\hat{f}(T)E[Z^{S \Delta T}] \\
\leq \sum_{|S| \in [0,k] \cup [n-k,n]} |\hat{f}(S)|^2 + \sum_{|S| \notin [0,k] \cup [n-k,n]} |\hat{f}(S)||\hat{f}(T)|\varepsilon \\
\leq 1 + 2\binom{n}{k} \varepsilon \leq 1 + \delta.
\]

By Cauchy–Schwarz,

\[
|E[f_{\text{ends}}(Z) \mathbb{1}(D_{\text{sym}} \notin G)]| \leq E[f_{\text{ends}}(Z)^2]^{1/2} Pr[D_{\text{sym}} \notin G]^{1/2} \leq 2 Pr[D_{\text{sym}} \notin G]^{1/2}. \tag{18}
\]

We now use (17) and (18) to bound the error as follows:

\[
|E[f(Z)] - E[f]| = |E[f(Z) \mathbb{1}(D_{\text{sym}} \in G)] + E[f(Z) \mathbb{1}(D_{\text{sym}} \notin G)] - E[f]| \\
\leq |E[f_{\text{ends}}(Z) \mathbb{1}(D_{\text{sym}} \in G)]| + |E[f_{\text{ends}}(Z) \mathbb{1}(D_{\text{sym}} \notin G)]| + Pr[D_{\text{sym}} \notin G] \\
\leq |E[f_{\text{ends}}(Z) \mathbb{1}(D_{\text{sym}} \notin G)]| + 3 Pr[D_{\text{sym}} \notin G] + \delta \\
\leq E[f_{\text{mid}}(Z) \mathbb{1}(D_{\text{sym}} \in G)] + 3 Pr[D_{\text{sym}} \notin G] + \delta \tag{19}
\]

We will bound the first term in (19) by

\[
|E[f_{\text{mid}}(z \cdot D_{\text{sym}})]| \leq \begin{cases} O(1) \left(\frac{120k}{n}\right)^{k/4} & \text{if } s \leq \sqrt{kn} \text{ and } t = (kn^3)^{1/4} \\ O(1) \left(\frac{120s^2}{n}\right)^{k/4} & \text{if } s \geq \sqrt{kn} \text{ and } t = \left(\frac{k^4n^4}{s^2}\right)^{1/4}. \end{cases} \tag{20}
\]

Plugging (16) and (20) into (19) gives us an error of

\[
O(1) \left(\frac{120\max\{s,\sqrt{kn}\}}{n}\right)^{k/2} + \delta
\]

as desired.

It remains to prove (20). Let \( S \) be any subset of size \( \ell \). As \( f \) is symmetric, we have \( \hat{f}(S) = \hat{f}(|\ell|) \). Let \( D'_{\text{sym}} \) be the distribution of \( D_{\text{sym}} \) conditioned on \( D_{\text{sym}} \in G \). Note that \( D'_{\text{sym}} \) is also symmetric, and so we have \( E[D'_{\text{sym}}] = \varepsilon_{\ell} \) for some \( \varepsilon_{\ell} \) which only depends on the size of \( S \). Hence,
\[
\left| \mathbb{E}[f_{\text{mid}}(z \cdot D_{\text{sym}}) \cdot (D_{\text{sym}} \in G)] \right| \leq \left| \sum_{\ell=k+1}^{n-k-1} \sum_{|S|=\ell} \hat{f}(S) \mathbb{E}[D_{\text{sym}}^S] z^S \right| = \left| \sum_{\ell=k+1}^{n-k-1} \hat{f}(\ell) \varepsilon_{\ell} \sum_{|S|=\ell} z^S \right|.
\]

As \( \sum_{|S|=\ell} z^S = |z|^{\ell} = \sum_{|S|=n-\ell} z^S \) by Fact 26 and Claim 31, we have
\[
\left| \sum_{\ell=k+1}^{n-k-1} \hat{f}(\ell) \varepsilon_{\ell} \sum_{|S|=\ell} z^S \right| \leq \left( \sum_{\ell=k+1}^{n-k-1} \left| \hat{f}(\ell) \varepsilon_{\ell} \right| \sum_{|S|=\ell} z^S \right) \leq 2 \sum_{\ell=k+1}^{n-k-1} K(\ell, t) \hat{f}(\ell, s) \frac{(\ell)^{3/2}}{((\ell)!)^{3/2}}.
\]

We first bound the sum over \( \ell \) from \( n/4 \) to \( n/2 \). Note that the binary entropy function \( H(x) := x \log_2(1/x) + (1-x) \log_2(1/(1-x)) \) is increasing on \([0, 1/2]\). In particular, we have \( H(1/4) \geq 4/5 \) and so \( \frac{2}{3} H(1/4) \geq 6/5 \). By Stirling’s approximation, we have \( (n) \geq \frac{1}{\sqrt{2\pi n}} n^{n+1/2} \) (see [24] for a proof). Applying Corollary 18 with these facts along with \( s \leq n/120 \) and \( t \leq (kn^k)^{1/4} \leq n/2 \), we have
\[
\sum_{\ell=\max\{n/4, k\}+1}^{[n/2]} \frac{K(\ell, t) \hat{f}(\ell, s)}{(\ell)^{3/2}} \leq \sum_{\ell=\max\{n/4, k\}+1}^{[n/2]} \frac{1}{(n) \ell^{3/2}} \cdot 2^{n (2H(\frac{1}{4}) + \frac{2}{3} + \frac{2}{3\pi})} \leq n^2 \cdot 2^{-n H(\frac{1}{4}) - \frac{2}{3\pi}} \leq 2^{-n/15}.
\]

We now bound the remaining sum (i.e. from \( \ell = k+1 \) to \( n/4 \)). Using Corollary 16 and Claim 31, and the inequality \( (n) \leq \left( \frac{2}{\pi} \right)^{\ell} \), we have
\[
\sum_{\ell=k+1}^{[n/4]} \frac{K(\ell, t) \hat{f}(\ell, s)}{(\ell)^{3/2}} \leq \sum_{\ell=k+1}^{[n/4]} \left( \frac{n}{\ell} \right)^{3/2} \cdot \left( \frac{\ell}{n} + \frac{t^2}{n^2} \right)^{\ell/2} \left( \frac{n}{\ell} + \frac{s^2}{n^2} \right)^{\ell/2} \leq \sum_{\ell=k+1}^{[n/4]} \left( e \left( \frac{\ell}{n} + \frac{t^2}{n^2} + \frac{s^2}{n^2} + \frac{t^2 s^2}{n^3} \right) \right)^{\ell/2}.
\]

We now consider the two cases in (20).

**Case 1:** \( s \leq \sqrt{nk} \) and \( t = (n^3 k)^{1/4} \). In this case (23) is at most
\[
\sum_{\ell=k+1}^{[n/4]} \left( e \left( \frac{\ell}{n} + \sqrt{\frac{k}{n} + \frac{k}{n} + \sqrt{\frac{k}{n}} \right) \right)^{\ell/2} \leq \sum_{\ell=k+1}^{[n/4]} \left( e \left( \frac{\ell}{n} + 3\sqrt{\frac{k}{n}} \right) \right)^{\ell/2} \leq O(1) \left( \frac{120k}{n} \right)^{k/4},
\]
where the last inequality follows because each term in the sum is at most 9/10 of its previous term, and so the sum is bounded by 10 times the first term. Combining this with (21) proves the first case in (20).

**Case 2:** \( s \geq \sqrt{nk} \) and \( t = (k^2 n^3 / s^2)^{1/4} \). In this case (23) is at most
\[
\sum_{\ell=k+1}^{[n/4]} \left( e \left( \frac{\ell}{n} + \frac{k}{s} + \frac{s^2}{n^2} + \frac{s}{n} \right) \right)^{\ell/2} \leq \sum_{\ell=k+1}^{[n/4]} \left( e \left( \frac{\ell}{n} + 3\frac{s}{n} \right) \right)^{\ell/2} \leq O(1) \left( \frac{120s^2}{n^2} \right)^{k/4},
\]
where again the last inequality follows because each term in the sum is at most 9/10 of its previous term, and so the sum is bounded by 10 times the first term. Combining this with (21) proves the second case in (20).
5 Proof of Theorem 14

In this section, we prove Theorem 14, which is based on the same idea that was used in the previous sections. The difference is that here we use that a typical shift is nearly balanced, and so \( \mathcal{K}(\ell, 1^T x) \) is small.

**Proof of Theorem 14.** Applying Cauchy–Schwarz, Parseval’s identity (to the function \( g(u) := f(u \cdot D) \)), and the assumption that \( \mathbb{E}[D^S] = 0 \) for \( |S| \in [1, k] \cup [n-k, n] \), we have

\[
\mathbb{E} \left[ \left| \mathbb{E}[f(u \cdot D)] - \mathbb{E}[f] \right|^2 \right] \leq \mathbb{E} \left[ \left( \mathbb{E}[f(u \cdot D)] - \mathbb{E}[f] \right)^2 \right] = \sum_{S:|S| \in \{k, n-k\}} \hat{f}(S)^2 \mathbb{E}[\chi_S(D^2)],
\]

where \( D^2 \) is the sum of two independent copies of \( D \), which is also \( k \)-wise uniform. Let \( G := \{ x \in \{-1, 1\}^n : |\sum_{i=1}^n x_i| \leq (\frac{k}{2n})^{1/4} \} \), and \( D_G \) be the conditional distribution of \( D^2 \) supported on \( G \). By Fact 29, the distribution \( D_G \) is \( \mathbb{Pr}[D \not\in G] \)-close to \( D^2 \). As \( \sum_{S \subseteq [n]} \hat{f}(S)^2 \leq 1 \), we have

\[
\left| \sum_{S:|S| \in \{k, n-k\}} \hat{f}(S)^2 \mathbb{E}[(D^2)^S] \right| \leq \left( \sum_{S:|S| \in \{k, n-k\}} \hat{f}(S)^2 \mathbb{E}[D_G^S] \right) + \mathbb{Pr}[D \not\in G].
\]

Applying Corollary 28 (to the even integer \( k-1 \) or \( k \)), we have

\[
\mathbb{Pr}[D \not\in G] \leq \left( \frac{2k}{en} \right)^{k-1}. \quad (24)
\]

We now bound the first term on the right hand side as follows. Fix a string \( z \in G \). As \( |\sum_{|S|=\ell} z^S| = |z^0| \sum_{|S|=n-\ell} z^S| = |\sum_{|S|=n-\ell} z^S| \), by Fact 26,

\[
\left| \sum_{S:|S| \in \{k, n-k\}} \hat{f}(S)^2 z^S \right| = \sum_{\ell=k+1}^{n/2} \hat{f}(\ell)^2 \sum_{|S|=\ell} z^S \leq \sum_{\ell=k+1}^{n/2} \frac{1}{\ell} \sum_{|S|=\ell} z^S.
\]

We separate the sum into two parts depending on \( \ell \leq n/5 \) and bound each of them individually. First, using Corollary 16, we have

\[
\sum_{\ell=k+1}^{n/5} \frac{1}{\ell} \sum_{|S|=\ell} z^S \leq \sum_{\ell=k+1}^{n/5} \left( \frac{\ell}{n} + \sqrt{\frac{k}{2en}} \right)^{\ell/2} \leq 2 \left( 2\sqrt{\frac{k}{2en}} \right)^{k/2} \leq 2 \left( \frac{2k}{en} \right)^{k/4}, \quad (25)
\]

because each term in the sum is at most half its previous term, and so the sum can be bounded by twice the first term. For the remaining sum (from \( \ell = \max\{k, n/5\} + 1 \) to \( n/2 \)), note that the binary entropy function \( H(x) := x \log_2(1/x) + (1-x) \log_2(1/(1-x)) \) is increasing on \([0, 1/2] \). In particular, we have \( H(1/5) - \frac{1}{\sqrt{2\pi}} \geq 1/4 \). By Stirling’s approximation, we have \( \binom{n}{\ell} \geq \frac{1}{\sqrt{2\pi}} 2^{nH(\ell/n)} \) (see [24] for a proof). Applying Corollary 18 with these facts, we have

\[
\sum_{\ell=\max\{n/5, k\}+1}^{n/2} \frac{1}{\ell} \sum_{|S|=\ell} z^S \leq \sum_{\ell=\max\{n/5, k\}+1}^{n/2} n^2 \cdot 2^{-\frac{\ell}{2}H(\ell/n)} \leq 2^{-n/10}.
\]

Combining this with (24) and (25) gives an error of \( (\frac{2k}{en})^{k-1} + 2(2\frac{2k}{en})^k + 2^{-n/10} \) \( \leq 6(\frac{2k}{en})^{k-1} \).
6 Bounds on Krawtchouk polynomials

In this section, we prove our upper and lower bounds on Krawtchouk polynomials (Corollary 16, Proposition 17, , and Claim 15). Corollary 16 follows directly from Lemma 36, which is a general upper bound on the elementary symmetric polynomials $\sum_{|S|=\ell} y^S$ that holds for arbitrary real tuples $y \in \mathbb{R}^n$, not only for $y \in \{-1,1\}^n$.

Lemma 36. Let $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. For every $1 \leq \ell \leq n$, we have

$$\left| \sum_{S \subseteq [n]: |S| = \ell} y^S \right| \leq (n/\ell) \left( \frac{\ell - 1}{n - 1} \cdot \frac{\sum_{i=1}^n y_i^2}{n} + \left( 1 - \frac{\ell - 1}{n - 1} \right) \cdot \frac{(\sum_{i=1}^n y_i)^2}{n^2} \right)^{\frac{\ell}{2}},$$

with equality if and only if $y_1 = \cdots = y_n$ or $\ell = 1$.

Specializing to $y \in \{-1,1\}^n$, the elementary symmetric polynomial $\sum_{|S|=\ell} y^S$ is simply the degree-$\ell$ (shifted) Krawtchouk polynomial $K(\ell, |y|)$. In this case, we always have $\sum_{i=1}^n y_i^2 = n$, and hence we obtain Corollary 16.

Corollary 16 appeared in [11] with an extra factor of $c^\ell$. Lemma 36 shows that the same inequality holds even when $y_1, \ldots, y_n$ are arbitrary real numbers. A similar-looking but incomparable inequality, first proved in [33], showed that

$$\left| \sum_{S \subseteq [n]: |S| = \ell} y^S \right| \leq O\left( \frac{k}{\ell} \right)^{\frac{\ell}{2}} \max_{k' \in \{k, k+1\}} \left( \left| \sum_{S \subseteq [n]: |S|=k'} y^S \right| \right)^{\frac{\ell}{2}},$$

Using a different approach, Tao [59] recently sharpened this inequality to

$$\left| \sum_{S \subseteq [n]: |S| = \ell} y^S \right| \leq O\left( \frac{1}{\ell} \right)^{\frac{\ell}{2}} \max_{k' \in \{k, k+1\}} \left( \left| \sum_{S \subseteq [n]: |S|=k'} y^S \right| \right)^{\frac{\ell}{2}},$$

confirming a conjecture made on MathOverflow, see https://mathoverflow.net/q/446254. Note that specializing to the case $k = 1$, and using the inequality $|\sum_{1 \leq i < j \leq n} y_i y_j| \leq \frac{1}{2} \sum_{i=1}^n y_i^2$, both (26) and (27) imply a weaker form of Lemma 36. In the other direction, Tao [58] observed that one cannot replace the quantity $\sum_{i=1}^n y_i^2$ in Lemma 36 with $|\sum_{1 \leq i < j \leq n} y_i y_j|$, as otherwise when $n$ is the square of an even number, for $y \in \{-1,1\}^n$ such that $\sum_{i=1}^n y_i = \sqrt{n}$, we have $\sum_{1 \leq i < j \leq n} y_i y_j = 0$ and the inequality fails at $\ell = n$.

We note that Lemma 36 can be obtained by a slight modification of both proofs in [33, 59]. Here we follow the approach taken in [59], as it gives a sharper constant and the argument is cleaner.

6.1 Proof of Lemma 36

Our approach is based on [59], which relies on several basic properties of real-rooted polynomials. We say an $(n+1)$-tuple of real numbers $(s_0, \ldots, s_n)$ is attainable if the polynomial

$$\sum_{k=0}^n (-1)^k \binom{n}{k} s_k z^{n-k}$$

is monic with all roots real. By its real-rootedness, we can factor the polynomial as

$$\sum_{k=0}^n (-1)^k \binom{n}{k} s_k(y) z^{n-k} = \prod_{i=1}^n (z - y_i)$$
for some real numbers \(y_1, \ldots, y_n\), where
\[
s_k(y) = \frac{1}{\binom{n}{k}} \sum_{|i| = k} y_i^2 = \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < \cdots < i_k \leq n} y_{i_1} \cdots y_{i_k}.
\]

Conversely, given an \(n\)-tuple of real numbers \(y = (y_1, \ldots, y_n)\), we can define \(s_k(y)\) as above to obtain an attainable tuple. We will use the following truncation property of attainable tuples.

**Fact 37 (Truncation).** Let \((s_0, \ldots, s_n)\) be an attainable tuple. Then \((s_0, \ldots, s_\ell)\) is attainable for every \(1 \leq \ell \leq n\).

**Proof.** It suffices to show that \((s_0, \ldots, s_{n-1})\) is attainable. Write \(s_k \coloneqq s_k(y_1, \ldots, y_n)\) for some real numbers \(y_1, \ldots, y_n\). By Rolle’s theorem, between every two consecutive real roots of a polynomial, there is a real root of its derivative. Thus the derivative of a real-rooted polynomial is also real-rooted. Therefore, the polynomial
\[
\frac{1}{n} \cdot \frac{d}{dz} \sum_{k=0}^{n} (-1)^k \binom{n}{k} s_k(y)z^{n-k} = \sum_{k=0}^{n-1} (-1)^k \frac{n-k}{n} \binom{n}{k} s_k(y)z^{n-1-k} = \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} s_k(y)z^{n-1-k}
\]
is monic and real-rooted, showing that \((s_0, \ldots, s_{n-1})\) is also attainable.

**Remark 38.** One should view \(s_\ell \) as \(s_\ell = \prod_{i=1}^{\ell} y'_i\) for some \(y'_1, \ldots, y'_\ell \in \mathbb{R}\), instead of \(s_\ell = \binom{n}{\ell}^{-1} \sum_{|i|=\ell} y^2\) for some \(y_1, \ldots, y_n \in \mathbb{R}\) such that \(s_n = \prod_{i=1}^{n} y_i\).

Lemma 36 relies on the following slight refinement in Tao’s argument.

**Lemma 39.** Let \((s_0, \ldots, s_n)\) be an attainable tuple. Then for every \(1 \leq \ell \leq n\),
\[
|s_\ell|^2 \leq (\ell - 1) \cdot (s_1^2 - s_2) + s_1^2.
\]

**Proof.** By the truncation property (Fact 37), it suffices to consider the case \(\ell = n\). Write \(s_k \coloneqq s_k(y_1, \ldots, y_n)\) for some \(y = (y_1, \ldots, y_n) \in \mathbb{R}^n\). By the AM-GM inequality, we have
\[
|s_n(y)|^2 = (y_1^2 \cdots y_n^2)^2 \leq \frac{1}{n} \sum_{i=1}^{n} y_i^2.
\]
By the Newton identity we have
\[
\sum_{i=1}^{n} y_i^2 = \left(\sum_{i=1}^{n} y_i\right)^2 - 2 \sum_{1 \leq i < j \leq n} y_i y_j = n^2 s_1(y_1, \ldots, y_n)^2 - 2 \binom{n}{2} s_2(y_1, \ldots, y_n).
\]
Therefore,
\[
|s_n(y)|^2 \leq n s_1(y)^2 - (n-1) s_2(y) = (n-1)(s_1(y)^2 - s_2(y)) + s_1(y)^2.
\]

Lemma 36 immediately follows from Lemma 39 by un-normalizing \(s_\ell, s_1\) and \(s_2\).
Proof of Lemma 36. Let \( S_k(y) := \binom{n}{k} s_k(y) = \sum_{|S|=k} y^S \). Applying Lemma 39, we have
\[
|S_k|^2 \leq \binom{n}{\ell} \left( (\ell - 1)(s_1^2 - s_2) + s_1^2 \right) = \binom{n}{\ell} \left( (\ell - 1)(\frac{S_1^2}{n^2} - \frac{2S_2}{n(n-1)}) + \frac{S_1^2}{n^2} \right) = \binom{n}{\ell} \left( \frac{\ell - 1}{n - 1} \left( \frac{S_1^2 - 2S_2}{n} - \frac{S_1^2}{n^2} \right) + \frac{S_1^2}{n^2} \right) = \binom{n}{\ell} \left( \frac{\ell - 1}{n - 1} \left( \frac{S_1^2 - 2S_2}{n} \right) + \left( 1 - \frac{\ell - 1}{n - 1} \right) \frac{S_1^2}{n^2} \right).
\]

Applying Newton’s identity, i.e., \( S_1^2 - 2S_2 = \sum_{i=1}^{n} y_i^2 \), completes the proof. ▶

We now prove Proposition 17. We note that a similar argument also appears in [51, Lemma 2.1]. For completeness we provide a self-contained proof here.

Proof of Proposition 17. First note
\[
(1 + z)^{(n+1)/2} (1 - z)^{(n-1)/2} = \sum_{\ell=0}^{n} K(\ell, t) z^\ell.
\]

Let \( r = |z| \). The logarithmic function is known to be concave:
\[
\alpha \log_2(u) + (1 - \alpha) \log_2(v) \leq \log_2(\alpha u + (1 - \alpha)v).
\]
for any positive \( u \) and \( v \). Using concavity and the observation \( |1 + z|^2 + |1 - z|^2 = 2 + 2|z|^2 = 1 + z + \overline{z} + |z|^2 + 1 - z - \overline{z} + |z|^2 = 2 + 2|z|^2 = 2 + 2r^2 \) gives
\[
H(\alpha) + \log_2 \left( |1 + z|^{2\alpha} |1 - z|^{2(1-\alpha)} \right) = \alpha \log_2 \left( \frac{|1 + z|^2}{\alpha} \right) + (1 - \alpha) \log_2 \left( \frac{|1 - z|^2}{1 - \alpha} \right) \\
\leq \log_2(|1 + z|^2 + |1 - z|^2) = \log_2(2 + 2r^2).
\]

For an integer \( \ell \) with \( 0 \leq \ell \leq n \) consider the Laurent polynomial
\[
p(z) = \frac{(1 + z)^{(n+1)/2} (1 - z)^{(n-1)/2}}{z^\ell}.
\]

If \( r^2 = \beta/(1 - \beta) \), then we have
\[
\log |p(z)| = \frac{n}{2} \left( \log_2 \left( |1 + z|^{2\alpha} |1 - z|^{2(1-\alpha)} \right) - \beta \log_2 \left( |z|^2 \right) \right) \leq \frac{n}{2} \left( \log_2(2 + 2r^2) - \beta \log_2(r^2) - H(\alpha) \right) \leq \frac{n}{2} \left( \log_2 \frac{2}{1 - \beta} - \beta \log_2 \frac{\beta}{1 - \beta} \right) = \frac{n}{2} \left( 1 + H(\beta) + H(\alpha) \right).
\]

The coefficient of \( 1 = z^0 \) in \( p(z) \) is \( K(\ell, t) \), so it follows that \( K(\ell, t) = \int_0^1 p(re^{2\pi i t}) \, dt \). We conclude that
\[
\log_2 |K(\ell, t)| \leq \log_2 \left( \int_0^1 |p(re^{2\pi i t})| \, dt \right) \leq \max_{|z|=r} \log_2 |p(z)| \leq \frac{n}{2} \left( 1 - H(\alpha) + H(\beta) \right). \quad ▶
\]
6.2 Lower bound on Krawtchouk polynomials

In this section, we prove Claim 15. It follows from an inequality on Krawtchouk polynomials which appears to be well known in the coding theory literature \cite{44, 36, 38, 37}, and essentially follows from Newton’s inequality.

For convenience we will work with the standard (non-shifted) definition of Krawtchouk polynomials $K(\ell, t) = \overline{K}(\ell, n - 2t)$. Note that in the claim below, we intentionally swap $t$ and $\ell$.

\begin{claim}
\text{Claim 40.} \quad K(n/2 - t, \ell) \geq \binom{n}{n/2 - t} \left( \frac{2t}{n} \right)^\ell \text{ for } t \geq \sqrt{\ell(n - \ell)}.
\end{claim}

Claim 40 follows from the fact that $K(t, 0) = \binom{n}{t}$ and then applying the following lemma iteratively $\ell$ times.

\begin{lemma}[Theorem 8 in \cite{38}]
\text{For } \ell, i \text{ such that } (n - 2i)^2 \geq 4\ell(n - \ell) \text{ (so that } s = \sqrt{(n - 2i)^2 - 4\ell(n - \ell)} \text{ is real and nonnegative),}
\end{lemma}

\begin{align*}
\frac{K(i, \ell + 1)}{K(i, \ell)} &> \frac{n - 2i + s}{2(n - \ell)} \geq \frac{n - 2i}{2n}.
\end{align*}

We can now prove Claim 15 using the following fact and translating the statement in terms of $\overline{K}(\cdot, \cdot)$.

\begin{fact}
\text{Fact 42.} \quad \binom{n}{\ell} K(\ell, t) = \binom{n}{t} K(t, \ell).
\end{fact}

Proof of Claim 15. We have

\begin{align*}
\overline{K}(\ell, t) &= K\left( \ell, \frac{n}{2} - \frac{t}{2} \right) = \binom{n}{\frac{n}{2} - \frac{t}{2}} K\left( \frac{n}{2} - \frac{t}{2}, \ell \right) \geq \binom{n}{\ell} \left( \frac{t}{2n} \right)^\ell.
\end{align*}

References


Harm Derksen, Peter Ivanov, Chin Ho Lee, and Emanuele Viola. Pseudorandomness, symmetry, smoothing: II, 2024.


