On Closeness to k-Wise Uniformity

Ryan O'Donnell
Carnegie Mellon University, Pittsburgh, PA, USA
odonnell@cs.cmu.edu

Yu Zhao
Carnegie Mellon University, Pittsburgh, PA, USA
yuzhao1@cs.cmu.edu

Abstract

A probability distribution over $\{-1,1\}^n$ is (ϵ,k)-*wise uniform* if, roughly, it is ϵ-close to the uniform distribution when restricted to any k coordinates. We consider the problem of how far an (ϵ,k)-wise uniform distribution can be from any globally k-wise uniform distribution. We show that every (ϵ,k)-wise uniform distribution is $O(n^{k/2}\epsilon)$-close to a k-wise uniform distribution in total variation distance. In addition, we show that this bound is optimal for all even k: we find an (ϵ,k)-wise uniform distribution that is $\Omega(n^{k/2}\epsilon)$-far from any k-wise uniform distribution in total variation distance. For $k=1$, we get a better upper bound of $O(\epsilon)$, which is also optimal.

One application of our closeness result is to the sample complexity of testing whether a distribution is k-wise uniform or δ-far from k-wise uniform. We give an upper bound of $O(n^k/\delta^2)$ (or $O(\log n/\delta^2)$ when $k=1$) on the required samples. We show an improved upper bound of $\tilde{O}(n^{k/2}/\delta^2)$ for the special case of testing fully uniform vs. δ-far from k-wise uniform. Finally, we complement this with a matching lower bound of $\Omega(n/\delta^2)$ when $k=2$.

Our results improve upon the best known bounds from [3], and have simpler proofs.

2012 ACM Subject Classification Theory of computation → Design and analysis of algorithms

Keywords and phrases k-wise independence, property testing, Fourier analysis, Boolean function

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2018.54

Related Version A full version of this paper is available at `https://arxiv.org/abs/1806.03569`.

Funding Supported by NSF grants CCF-1618679, CCF-1717606. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

1 Introduction

1.1 k-wise uniformity and almost k-wise uniformity

We say that a probability distribution over $\{-1,1\}^n$ is k-*wise uniform* if its marginal distribution on every subset of k coordinates is the uniform distribution. For Fourier analysis of the Hamming cube, it is convenient to identify the distribution with its density function $\varphi : \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ satisfying

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\{-1,1\}^n}[\varphi(\boldsymbol{x})] = 1.$$

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018).
Editors: Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer; Article No. 54; pp. 54:1–54:19

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

We write $\boldsymbol{x} \sim \varphi$ to denote that $\boldsymbol{x}$ is a random variable drawn from the associated distribution with density φ:

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}[\boldsymbol{x} = x] = \frac{\varphi(x)}{2^n}$$

for any $x \in \{-1,1\}^n$. Then a well-known fact is that a distribution is k-wise uniform if and only if the Fourier coefficient of φ is 0 on every subset $S \subseteq [n]$ of size between 1 and k:

$$\widehat{\varphi}(S) = \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\varphi}\left[\prod_{i\in S} \boldsymbol{x}_i\right] = 0.$$

k-wise uniformity is an essential tool in theoretical computer science. Its study dates back to work of Rao [25]. They studied k-wise uniform sets, which are special cases of k-wise uniform distribution. A subset of $\{-1,1\}^n$ is *a k-wise uniform set* if the uniform distribution on this subset is k-wise uniform. Rao gave constructions of a pairwise-uniform set of size $n+1$ (when $n = 2^r - 1$ for any integer r), a 3-wise uniform set of size $2n$ (when $n = 2^r$ for any integer r), and a lower bound (reproved in [4, 14]) that a k-wise uniform set on $\{-1,1\}^n$ requires size at least $\Omega(n^{\lfloor k/2 \rfloor})$. An alternative proof of the lower bound for even k is shown in [6] using a hypercontractivity-type technique, as opposed to the linear algebra method. Coding theorists have also heavily studied k-wise uniformity, since MacWilliams and Sloane showed that linear codes with dual minimum distance $k+1$ correspond to k-wise uniform sets in [21]. The importance in theoretical computer science of k-wise independence for derandomization arose simultaneously in many papers, with [17, 20] emphasizing derandomization via the most common pairwise-uniformity case, and [4, 14] emphasizing derandomization based on k-wise independence more generally.

A distribution is "almost k-wise uniform" if its marginal distribution on every k coordinates is very close to the uniform distribution. Typically we say two distributions φ, ψ are *δ-close*, if the total variation distance between φ and ψ is at most δ; and we say they are *δ-far*, if the total variation distance between them is more than δ. However the precise notion of "close to uniform" has varied in previous work. Suppose ψ is the density function for the marginal distribution of φ restricted to some specific k coordinates and $\mathbf{1}$ is the density function for the uniform distribution. Several standard ways are introduced in [6, 3] to quantify closeness to uniformity, corresponding to the L_1, L_2, L_∞ norms:

- (L_1 norm): $\|\psi - \mathbf{1}\|_1 = 2d_{\mathrm{TV}}(\psi, \mathbf{1}) \le \epsilon$, where d_{TV} denotes total variation distance;
- (L_2 norm): $\|\psi - \mathbf{1}\|_2 = \sqrt{\chi^2(\psi, \mathbf{1})} = \sqrt{\sum_{S\neq\emptyset} \widehat{\psi}(S)^2} \le \epsilon$, where $\chi^2(\psi, \mathbf{1})$ denotes the χ^2-divergence of ψ from the uniform distribution;
- (L_∞ norm): $\|\psi - \mathbf{1}\|_\infty \le \epsilon$, or in other words, for any $x \in \{-1,1\}^n$,

$$\left|\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\psi}[\boldsymbol{x} = x] - 2^{-k}\right| \le 2^{-k}\epsilon.$$

Note the following: First, closeness in L_1 norm is the most natural for algorithmic derandomization purposes: it tells us that the algorithm cannot tell ψ is different from the uniform distribution up to ϵ error. Second, these definitions of closeness are in increasing order of strength. On the other hand, we also have that $\|\psi - \mathbf{1}\|_1 \le \|\psi - \mathbf{1}\|_\infty \le 2^k \|\psi - \mathbf{1}\|_1$; thus all these notions are within a factor of 2^k. We generally consider k to be constant (or at worst, $O(\log n)$), so that these notions are roughly the same.

A fourth reasonable notion, proposed by Naor and Naor in [22], is that the distribution has a small bias over every non-empty subset of at most k coordinates. We say density

function φ is (ϵ, k)-*wise uniform* if for non-empty set $S \subseteq [n]$ with size at most k,

$$|\widehat{\varphi}(S)| = \left| \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}\left[\prod_{i\in S} \boldsymbol{x}_i = 1\right] - \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\varphi}\left[\prod_{i\in S} \boldsymbol{x}_i = -1\right]\right| \leq \epsilon.$$

Here we also have $\epsilon = 0$ if and only if φ is exactly k-wise uniform. Clearly if the marginal density of φ over every k coordinates is ϵ-close to the uniform distribution in total variation distance, then φ is (ϵ, k)-wise uniform. On the other hand, if φ is (ϵ, k)-wise uniform, then the marginal density of φ over every k coordinates is $2^{k/2}\epsilon$-close to uniform distribution in total variation distance. Again, if k is considered constant, this bias notion is also roughly the same as previous notions. In the rest of paper we prefer this (ϵ, k)-wise uniform notion for "almost k-wise uniform" because of its convenience for Fourier analysis.

The original paper about almost k-wise uniformity is [22], which is concerned with derandomization; e.g., they use (ϵ, k)-wise uniformity for derandomizing the "set balancing (discrepancy)" problem. Alon et al. give a further discussion of the relationship between almost k-wise uniformity and derandomization in [6]. The key idea is the following: In many cases of randomized algorithms, the analysis only relies on the property that the random bits are k-wise uniform, as opposed to fully uniform. Since there exists an efficiently samplable k-wise uniform distribution on a set of size at most $O(n^{\lfloor k/2 \rfloor})$, one can reduce the number of random unbiased bits used in the algorithm down to $O(k \log n)$. To further reduce the number of random bits used, a natural line of thinking is to consider distributions which are "almost k-wise uniformity". Alon et al. [5] showed that we can deterministically construct (ϵ, k)-wise uniform sets that are of size $\mathrm{poly}(2^k, \log n, 1/\epsilon)$, much smaller than exact k-wise uniform ones (roughly $\Omega(n^{\lfloor k/2 \rfloor})$ size). Therefore we can use substantially fewer random bits by taking random strings from an almost k-wise uniform distribution.

However we need to ensure that the original analysis of the randomized algorithm still holds under the almost k-wise uniform distribution. This is to say that if the randomized algorithm behaves well on a k-wise uniform distribution, it may also work as well with an (ϵ, k)-wise uniform distribution, when the parameter ϵ is small enough.

1.2 The Closeness Problem

For the analysis of derandomization, it would be very convenient if (ϵ, k)-wise uniformity – which means that "every k-local view looks close to uniform" – implies global δ-closeness to k-wise uniformity. A natural question that arises, posed in [6], is the following:

How small can δ be, such that the following is true: For every (ϵ, k)-wise uniform distribution φ on $\{-1, 1\}^n$, φ is δ-close to some k-wise uniform distribution?

In this paper, we will refer to this question as *the Closeness Problem*.

1.2.1 Previous work and applications

On one hand, the main message of [6] showed a lower bound: For every even constant $k > 4$, they gave an (ϵ, k)-wise uniform distribution with $\epsilon = O(1/n^{k/4-1})$, yet which is $\frac{1}{2}$-far from every k-wise uniform distribution in total variation distance.

On the other hand, [6] proved a very simple theorem that $\delta \leq O(n^k\epsilon)$ always holds. Despite simplicity, this upper bound has been used many times in well known results.

One application is in circuit complexity. [6]'s upper bound is used for fooling disjunctive normal formulas (DNF) [11] and AC^0 [12]. In these works, once the authors showed that k-wise uniformity suffices to fool DNF/AC^0, they deduced that $(O(1/n^k), k)$-uniform distributions

suffice, and hence $O(1/n^k)$-biased sets sufficed trivially. [6]'s upper bound is also used as a tool for the construction of two-source extractors for a similar reason in [13, 19].

The notions of pairwise-uniformity, k-wise uniformity, and δ-closeness to k-wise uniformity are also important for hardness of constraint satisfactory problems (CSPs). Austrin and Mossel [7] shows that one can obtain integrality gaps and UGC-hardness for CSPs based on k-wise uniform distributions of small support size. If a predicate is k-wise uniform, Kothari et al. [18] showed that one can get SOS-hardness of refuting random instances of it when there are around $n^{(k+1)/2}$ constraints. Indeed, [18] shows that if we have a predicate that is δ-close to k-wise uniform, then with roughly $n^{(k+1)/2}$ random constraints, SOS cannot refute that a $(1 - O(\delta))$-fraction of constraints are satisfiable. This also motivates studying δ-closeness to k-wise uniformity, and how it relates to Fourier coefficients. δ-closeness to k-wise uniformity is also discussed in [2] on hardness of random CSP.

Alon et al. [3] investigated the Closeness Problem further by improving the upper bound to $O((n \log n)^{k/2}\epsilon)$. Indeed, they showed a strictly stronger fact that a distribution is $O\left(\sqrt{\mathbf{W}^{1\dots k}[\varphi]} \log^{k/2} n\right)$-close to some k-wise uniform, where $\mathbf{W}^{1\dots k}[\varphi] = \sum_{1\le|S|\le k} \widehat{\varphi}(S)^2$. Rubinfeld and Xie [27] generalized some of these results to non-uniform k-wise independent distributions over larger product spaces.

Let us briefly summarize the method [3] used to prove their upper bounds. Given an (ϵ, k)-wise uniform φ, They first try to generate a k-wise uniform "pseudo-distribution" φ' by forcing all Fourier coefficients at degree at most k to be zero. It is a "pseudo-distribution" because some points might have negative density. After this, they use a fully uniform distribution and k-wise uniform distributions with small support size to try to mend all points to be non-negative. They bound the weight of these mending distributions to upper-bound the distance incurred by mending process. This mending process uses the fully uniform distribution to mend the small negative weights and uses k-wise uniform distributions with small support size to correct the large negative weights point by point. By optimizing the threshold between small and large weights it introduces a factor of $(\log n)^{k/2}$.

Though they did not mention it explicitly, they also give a lower bound for the Closeness Problem of $\delta \ge \Omega\left(\frac{n^{(k-1)/2}}{\log n}\epsilon\right)$ for $k > 2$ by considering the uniform distribution on a set of $O(n^k)$ random chosen strings. No previous work gave any lower bound for the most natural case of $k = 2$.

1.2.2 Our result

In this paper, we show sharper upper and lower bounds for the Closeness Problem, which are tight for k even and $k = 1$. Comparing to the result in [3], we get rid of the factor of $(\log n)^{k/2}$.

▶ **Theorem 1.** *Any density φ over $\{-1, 1\}^n$ is δ-close to some k-wise uniform distribution, where*

$$\delta \le e^k \sqrt{\mathbf{W}^{1\dots k}[\varphi]} = e^k \sqrt{\sum_{1\le|S|\le k} \widehat{\varphi}(S)^2}.$$

Consequently, if φ is (ϵ, k)-wise uniform, i.e., $|\widehat{\varphi}(S)| \le \epsilon$ for every non-empty set S with size at most k, then

$$\delta \le e^k n^{k/2} \epsilon.$$

For the special case $k = 1$, the corresponding δ can be further improved to $\delta \le \epsilon$.

Our new technique is trying to mend the original distribution to be k-wise uniform all at once. We want to show that some mixture distribution $(\varphi + w\psi)$ is k-wise uniform with small mixture weight w. The distance between the final mixture distribution and the original distribution φ is bounded by $O(w)$. Therefore we only need to show that the mending distribution ψ exists for some small weight w. Showing the existence of such a distribution ψ can be written as the feasibility of a linear program (LP). We upper bound w by bounding the dual LP, using the hypercontractivity inequality.

Our result is sharp for all even k, and is also sharp for $k = 1$. We state the matching lower bound for even k:

▶ **Theorem 2.** *For any n and even k, and small enough ϵ, there exists some (ϵ, k)-wise uniform distribution φ over $\{-1, 1\}^n$, such that φ is δ-far from every k-wise uniform distribution in total variation distance, where*

$$\delta \geq \Omega\left(\frac{1}{k}\right)^k n^{k/2}\epsilon.$$

Our method for proving this lower bound is again LP duality. Our examples in the lower bound are symmetric distributions with Fourier weight only on level k. The density functions then can be written as binary Krawtchouk polynomials which behave similar to Hermite polynomials when n is large. Our dual LP bounds use various properties of Krawtchouk and Hermite polynomials.

Interestingly both our upper and lower bound utilize LP-duality, which we believe is the most natural way of looking at this problem.

We remark that we can derive a lower bound for odd k from Theorem 2 trivially by replacing k by $k-1$. There exists a gap of $\sqrt{n}$ between the resulting upper and lower bounds for odd k. We believe that the lower bound is tight, and the upper bound may be improvable by a factor of $\sqrt{n}$, as it is in the special case $k = 1$. We leave it as a conjecture for further work:

▶ **Conjecture 3.** *Suppose the distribution φ over $\{-1, 1\}^n$ is (ϵ, k)-wise uniform. Then φ is δ-close to some k-wise uniform distribution in total variation distance, where*

$$\delta \leq O(n^{\lfloor k/2 \rfloor}\epsilon).$$

1.3 The Testing Problem

Another application of the Closeness Problem is to property testing of k-wise uniformity. Suppose we have sample access from an unknown and arbitrary distribution; we may wonder whether the distribution has a certain property. This question has received tremendous attention in the field of statistics. The main goal in the study of property testing is to design algorithms that use as few samples as possible, and to establish lower bound matching these sample-efficient algorithms. In particular, we consider the property of being k-wise uniform:

Given sample access to an unknown and arbitrary distribution φ on $\{-1, 1\}^n$, how many samples do we need to distinguish between the case that φ is k-wise uniform versus the case that φ is δ-far from every k-wise uniform distribution?

In this paper, we will refer to this question as *the Testing Problem*.

We say a testing algorithm is *a δ-tester for k-wise uniformity* if the algorithm outputs "Yes" with high probability when the distribution φ is k-wise uniform, and the algorithm outputs "No" with high probability when the distribution φ is δ-far from any k-wise uniform distribution (in total variation distance).

Property testing is well studied for Boolean functions and distributions. Previous work studied testing related properties of distribution, including uniformity [16, 9, 26] and independence [8, 10, 1, 15].

[6, 3, 28] discussed about testing k-wise uniformity. [6] constructed a δ-tester for k-wise uniformity with sample complexity $O(n^{2k}/\delta^2)$, and [3] improved it to $O(n^k \log^{k+1} n/\delta^2)$. As for lower bounds, [3] show that $\Omega(n^{(k-1)/2}/\delta)$ samples are necessary, albeit only for $k > 2$. This lower bound is in particular for distinguishing the uniform distribution from δ-far-from-k-wise.

We show a better upper bound for sample complexity:

▶ **Theorem 4.** *There exists a δ-tester for k-wise uniformity of distributions on $\{-1,1\}^n$ with sample complexity $O\left(\frac{1}{k}\right)^{k/2} \frac{n^k}{\delta^2}$. For the special case of $k = 1$, the sample complexity is $O\left(\frac{\log n}{\delta^2}\right)$.*

A natural δ-tester of k-wise uniformity is mentioned in [3]: Estimate all Fourier coefficients up to level k from the samples. If they are all smaller than ϵ then output "Yes". In fact this algorithm is exactly attempting to check whether the distribution is (ϵ, k)-wise uniform. Hence the sample complexity depends on the upper bound for the Closeness Problem. Therefore we can reduce the sample complexity of this algorithm down to $O\left(\frac{n^k \log n}{\delta^2}\right)$ via our improved upper bound for the Closeness Problem. One $\log n$ factor remains because we need to union-bound over the $O(n^k)$ Fourier coefficients up to level k. To further get rid of the last $\log n$ factor, we present a new algorithm that estimates the Fourier weight up to level k, $\sum_{1 \le |S| \le k} \widehat{\varphi}^2(S)$, rather than estimating these Fourier coefficients one by one.

Unfortunately, a lower bound for the Closeness Problem does not imply a lower bound for the Testing Problem directly. In [3], they showed that a uniform distribution over a random subset of $\{-1,1\}^n$ of size $O(\frac{n^{k-1}}{\delta^2})$, is almost surely δ-far from any k-wise uniform distribution. On the other hand, by the Birthday Paradox, it is hard to distinguish between the fully uniform distribution on all strings of length n and a uniform distribution over a random set of such size. This gives a lower bound for the Testing Problem as $\Omega(n^{(k-1)/2}/\delta)$. Their result only holds for $k > 2$; there was no previous non-trivial lower bound for testing pairwise uniformity. We show a lower bound for the pairwise case.

▶ **Theorem 5.** *Any δ-tester for pairwise uniformity of distributions on $\{-1,1\}^n$ needs at least $\Omega(\frac{n}{\delta^2})$ samples.*

For this lower bound we analyze a symmetric distribution with non-zero Fourier coefficients only on level 2. We prove that it is hard to distinguish a randomly shifted version of this distribution from the fully uniform distribution. This lower bound is also better than [3] in that we have a better dependence on the parameter δ ($\frac{1}{\delta^2}$ rather than $\frac{1}{\delta}$). Unfortunately we are unable to generalize our lower bound for higher k.

Notice that for our new upper and lower bounds for k-wise uniformity testing, there still remains a quadratic gap, for $k \ge 2$, indicating that the upper bound might be able to be improved. Both the lower bound in our paper and that in [3] show that it is hard to distinguish between the fully uniform distribution and some specific sets of distributions that are far from k-wise uniform. We show that if one wants to improve the lower bound, one will need to use a distribution in the "Yes" case that is *not* fully uniform, because we give a sample-efficient algorithm for distinguishing between fully uniform and δ-far from k-wise uniform:

Table 1 Summary of our results.

	Upper bound		Lower bound	
	[3]	Our paper	[3]	Our paper
Closeness Problem	$O(n^{k/2}(\log n)^{k/2}\epsilon)$	$O(n^{k/2}\epsilon)$ $O(\epsilon)$ for $k=1$	$\Omega\left(\frac{n^{(k-1)/2}}{\log n}\epsilon\right)$	$\Omega(n^{\lfloor k/2\rfloor}\epsilon)$
Testing k-wise vs. far from k-wise	$O\left(\frac{n^k(\log n)^{k+1}}{\delta^2}\right)$	$O\left(\frac{n^k}{\delta^2}\right)$ $O\left(\frac{\log n}{\delta^2}\right)$ for $k=1$	$\Omega\left(\frac{n^{(k-1)/2}}{\delta}\right)$ for $k>2$	$\Omega\left(\frac{n}{\delta^2}\right)$ for $k=2$
Testing n-wise vs. far from k-wise	$O\left(\frac{n^k(\log n)^{k+1}}{\delta^2}\right)$	$O\left(\frac{n^{k/2}}{\delta^2}(\log\frac{n}{\delta})^{k/2}\right)$ $O\left(\frac{\log n}{\delta^2}\right)$ for $k=1$	$\Omega\left(\frac{n^{(k-1)/2}}{\delta}\right)$ for $k>2$	$\Omega\left(\frac{n}{\delta^2}\right)$ for $k=2$

▶ **Theorem 6.** *For any constant k, for testing whether a distribution is fully uniform or δ-far from every k-wise uniform distribution, there exists an algorithm with high probability ($>\frac{2}{3}$) with sample complexity $O(k)^k\cdot n^{k/2}\cdot\frac{1}{\delta^2}\cdot\left(\log\frac{n}{\delta}\right)^{k/2}$.*

In fact, for testing whether a distribution is αk-wise uniform or δ-far from k-wise uniform with $\alpha>4$ (assuming αk is an even integer), there exists an algorithm with high probability ($>\frac{2}{3}$) with sample complexity $O(\alpha)^{k/2}\cdot n^{k/2}\cdot\frac{1}{\delta^2}\cdot\left(\frac{n^k}{\delta^4}\right)^{1/(\alpha-2)}$.

We remark that testing fully uniformity can be treated as a special case of testing αk-wise uniformity approximately, by setting $\alpha=\log\frac{n}{\delta}$.

Testing full uniformity has been studied in [16, 9]. Paninski [24] showed that testing whether an unknown distribution on $\{-1,1\}^n$ is $\Theta(1)$-close to fully uniform requires $2^{n/2}$ samples. Rubinfeld and Servedio [26] studied testing whether an unknown monotone distribution is fully uniform or not.

The fully uniform distribution has the nice property that every pair of samples is different in $\frac{n}{2}\pm O(\sqrt{n})$ bits with high probability when the sample size is small. Our algorithm first rejects those distributions that disobey this property. We show that the remaining distributions have small Fourier weight up to level $2k$. Hence by following a similar analysis as the tester in Theorem 4, we can get an improved upper bound when these lower Fourier weights are small.

The lower bound remains the same as testing k-wise vs. far from k-wise. Our tester is tight up to a logarithm factor for the pairwise case, and is tight up to a factor of $\tilde{O}(\sqrt{n})$ when $k>2$.

We compare our result and previous best known bounds from [3] in Table 1. (We omit factors depending on k.)

1.4 Organization

We keep our notations the same as in [23], or see the detailed preliminaries in the full version. We will discuss upper and lower bounds for the Closeness Problem in Section 2. We will discuss the sample complexity of testing k-wise uniformity in Section 3. We present a tester for distinguishing between αk-wise uniformity (or fully uniformity) and far-from k-wise uniformity in Section 4.

2 The Closeness Problem

In this section, we prove the upper bound in Theorem 1 and the lower bound in Theorem 2. One interesting fact is that we use duality of linear programming (LP) in both the upper and lower bound. We think this is the proper perspective for analyzing these questions.

2.1 Upper bound

The key idea for proving the upper bound is mixture distributions. Given an (ϵ, k)-wise uniform density φ, we try to mix it with some other distribution ψ using mixture weight w, such that the mixture distribution $\frac{1}{1+w}(\varphi + w\psi)$ is k-wise uniform and is close to the original distribution. The following lemma shows that the distance between the original distribution and the mixture distribution is bounded by the weight w.

▶ **Lemma 7.** *If $\varphi' = \frac{1}{1+w}(\varphi + w\psi)$ for some $0 \le w \le 1$ and density functions φ, ψ, then $d_{\mathrm{TV}}(\varphi, \varphi') \le w$.*

Proof. $d_{\mathrm{TV}}(\varphi, \varphi') = \frac{1}{2}\|\varphi' - \varphi\|_1 = \frac{1}{2}\|\varphi' - ((1+w)\varphi' - w\psi))\|_1 = \frac{1}{2}w\|\varphi' - \psi\|_1 \le w.$ ◀

Therefore we only need to show the existence of an appropriate ψ for some small w. The constraints on ψ can be written as an LP feasibility problem. Therefore by Farkas' Lemma we only need to show that its dual is not feasible. The variables in the dual LP can be seen as a density function of degree at most k.

Proof of Theorem 1 (general k case). Given density function φ, we try to find another density function ψ with constraints

$$\widehat{\psi}(S) = -\frac{1}{w}\widehat{\varphi}(S)$$

for all $1 \le |S| \le k$. Suppose such a density function ψ exists. Then it is trivial that $\frac{\varphi + w\psi}{1+w}$ is also a density function and is k-wise uniform. By Lemma 7, we know that $d_{\mathrm{TV}}(\varphi, \mathrm{kWISE}) \le w$.

The rest of proof is to show that such a ψ exists when $w = e^k\sqrt{\mathbf{W}^{1\dots k}[\varphi]}$. We can write it as an LP with variables $\psi(x)$ for $x \in \{-1,1\}^n$ and constraints:

$$\begin{aligned}
&\widehat{\psi}(\emptyset) = 1, \\
&\widehat{\psi}(S) = -\frac{1}{w}\widehat{\varphi}(S), && \forall 1 \le |S| \le k, \\
&\psi(x) \ge 0, && \forall x \in \{-1,1\}^n,
\end{aligned}$$

where $\widehat{\psi}(S) = \mathbf{E}[\psi(\boldsymbol{x})\boldsymbol{x}^S]$ is a linear combination of variables $\psi(x)$.

The dual LP has variables $\psi'(x)$ for $x \in \{-1,1\}^n$ with constraints:

$$\begin{aligned}
&\widehat{\psi'}(\emptyset) = 1, \\
&\widehat{\psi'}(S) = 0, && \forall |S| > k, \\
&\psi'(x) \ge 0, && \forall x \in \{-1,1\}^n, \\
&\frac{1}{w}\sum_{1 \le |S| \le k} \widehat{\varphi}(S)\widehat{\psi'}(S) > 1.
\end{aligned}$$

The original LP is feasible if and only if its dual LP is infeasible, by Farkas' Lemma. This completes the proof, since when $w = e^k\sqrt{\mathbf{W}^{1\dots k}[\varphi]}$, for any density function ψ' with degree k we have

$$\frac{1}{w}\sum_{1 \le |S| \le k} \widehat{\varphi}(S)\widehat{\psi'}(S) \le \frac{1}{e^k\sqrt{\mathbf{W}^{1\dots k}[\varphi]}} \sum_{1 \le |S| \le k} |\widehat{\varphi}(S)||\widehat{\psi'}(S)| \le \frac{1}{e^k}\|\widehat{\psi'}\|_2 \le 1,$$

where the second inequality holds by Cauchy-Schwarz, and the last inequality holds by hypercontractivity. ◀

The proof of special case $k = 1$ for Theorem 1 is included in the appendix.

2.2 Lower bound

Interestingly, our proof of the lower bound also utilizes LP duality. We can write the closeness problem in the form of linear programming with variables $\varphi'(x)$ for $x \in \{-1,1\}^n$, as follows:

$$\begin{aligned}
\text{minimize} \quad & d_{\mathrm{TV}}(\varphi,\varphi') = \frac{1}{2}\|\varphi-\varphi'\|_1 \\
\text{subject to:} \quad & \widehat{\varphi'}(\emptyset) = 1, \\
& \widehat{\varphi'}(S) = 0, && \forall 1 \le |S| \le k, \\
& \varphi'(x) \ge 0, && \forall x \in \{-1,1\}^n.
\end{aligned}$$

We ignore the factor of $1/2$ in the minimization for convenience in the following analysis. The dual LP, which has variables $p(x), q(x)$ for $x \in \{-1,1\}^n$, is the following:

$$\begin{aligned}
\text{maximize} \quad & \langle \varphi, q\rangle - \widehat{p}(\emptyset) \\
\text{subject to:} \quad & p(x) - q(x) \ge 0, && \forall x \in \{-1,1\}^n, \\
& q(x) \le 1, && \forall x \in \{-1,1\}^n, \\
& p(x) \ge -1, && \forall x \in \{-1,1\}^n, \\
& \deg(p) \le k.
\end{aligned}$$

Thus given a pair of Boolean functions p, q satisfying the constraints, the quantity $(\langle \varphi, q\rangle - \widehat{p}(\emptyset))$ is a lower bound for our closeness problem. Our distribution φ achieving the lower bound is a symmetric polynomial, homogeneous of degree k (except that it has a constant term of 1, as is necessary for every density function). We define

$$\varphi(x) = 1 + \mu\binom{n}{k}^{-1/2} \sum_{|S|=k} x^S, \quad p(x) = \mu\binom{n}{k}^{-1/2} \sum_{|S|=k} x^S, \quad q(x) = \min(p(x), 1),$$

where $\mu = \frac{\sqrt{k!}}{2(Ck)^k}$ with some sufficiently large constant C. We have $\epsilon = \max_{1\le|S|\le k} |\widehat{\varphi}(S)| = \mu\binom{n}{k}^{-1/2}$.

We use properties of Krawtchouk and Hermite polynomials to get a lower bound of $\Omega\left(\frac{1}{k}\right)^k$ for the optimization.

This completes the proof, because φ is at least δ-far from k-wise uniform with $\delta = \Omega\left(\frac{1}{k}\right)^k$, and we have $\epsilon = \mu\binom{n}{k}^{-1/2} \le \frac{n^{-k/2}}{2^{\Omega(k)}}$. Therefore we have $\delta \ge \Omega\left(\frac{1}{k}\right)^k n^{k/2}\epsilon$.

The detailed proof of Theorem 2 is included in the full version.

3 The Testing Problem

In this section, we study the problem of testing whether a distribution is k-wise uniform or δ-far from k-wise uniform. These bounds are based on new bounds for the Closeness Problem. We present a new testing algorithm for general k in Section 3.1. We give a lower bound for the pairwise case in Section 3.2.

3.1 Upper bound

Given m samples from φ, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$, we will first show that

$$\Delta(\boldsymbol{X}) = \operatorname*{avg}_{1\le s<t\le m} \left(\sum_{1\le|S|\le k} \boldsymbol{x}_s^S \boldsymbol{x}_t^S \right)$$

is a natural estimator of $\mathbf{W}^{1\ldots k}[\varphi]$.

▶ **Lemma 8.**

$$\mu = \mathbf{E}[\Delta(\boldsymbol{X})] = \mathbf{W}^{1\dots k}[\varphi];$$
$$\mathbf{Var}[\Delta(\boldsymbol{X})] \le \frac{4}{m^2} L_k(\varphi) + \frac{4}{m}\sqrt{L_k(\varphi)\mu}, \tag{1}$$

where $L_k(\varphi) = \sum_{1\le|S_1|,|S_2|\le k} \widehat{\varphi}(S_1 \oplus S_2)^2$.

Hence we can bound the samples we need for estimating $\mathbf{W}^{1\dots k}$.

▶ **Theorem 9** ($\mathbf{W}^{1\dots k}$ Estimation Test)**.** *Let* $\varphi : \{-1,1\}^n \to \mathbb{R}^{\ge 0}$ *be a density function, promised to satisfy* $\mathbf{W}^i[\varphi] \le An^{i/2}$ *for any* $i = 0, 1, \dots, 2k$. *There is an algorithm that, given*

$$m \ge 1000\frac{2^k\sqrt{A}n^{k/2}}{\theta} \tag{2}$$

samples, distinguishing with probability at least $3/4$ *whether* $\mathbf{W}^{1\dots k}[\varphi] \le \frac{1}{2}\theta$ *or* $\mathbf{W}^{1\dots k}[\varphi] > \theta$.

This $\mathbf{W}^{1\dots k}$ Estimation Test is just what we need for testing k-wise uniformity with the upper bound of the Closeness Problem.

Proof of Theorem 4. From Theorem 1 we know that if density φ is δ-far from k-wise uniform, then $\mathbf{W}^{1\dots k}[\varphi] > \left(\frac{\delta}{e^k}\right)^2$; On the other hand if φ is k-wise uniform, by definition we have $\mathbf{W}^{1\dots k}[\varphi] = 0$. Therefore distinguishing between k-wise uniform and δ-far from k-wise uniform can be reduced to distinguishing between $\mathbf{W}^{1\dots k}[\varphi] > \left(\frac{\delta}{e^k}\right)^2$ and $\mathbf{W}^{1\dots k}[\varphi] = 0$.

For any density function φ, $|\widehat{\varphi}(S)| = \left|\mathbf{E}[\varphi(\boldsymbol{x})\boldsymbol{x}^S]\right| \le 1$ for any $S \subseteq [n]$. Therefore assigning $A = n^k$, we have

$$\mathbf{W}^i[\varphi] = \sum_{|S|=i} \widehat{\varphi}(S)^2 \le n^i \le An^{i/2}$$

for every $i = 0, 1, \dots, 2k$.

Hence we can run the $\mathbf{W}^{1\dots k}$ Estimator Test in Theorem 9 with parameter $\theta = \left(\frac{\delta}{e^k}\right)^2$ and $A = n^k$, the we solve then Testing Problem with sample complexity $2^{O(k)}n^k/\delta^2$.

In fact by precise calculation we can further improve the factor only related to k to $O\left(\frac{1}{k}\right)^{k/2}$, but we will omit the proof here. ◀

The proofs of Lemma 8 and Theorem 9 are included in the appendix.

3.2 Lower bound for the pairwise case

An upper bound for the Closeness Problem implies an upper bound for the Testing Problem. But a lower bound for Closeness does not obviously yield a lower bound for the testing problem. The function used to show the lower bound for the Closeness Problem is far from k-wise uniform, but it is not sufficient to say that it is hard to distinguish between it and some k-wise uniform distribution. In [3], they show that it is hard to distinguish between the fully uniform distribution and the uniform distribution on a random set of size around $O(n^{k-1}/\delta^2)$; this latter distribution is far from k-wise uniform with high probability for $k > 2$.

We show that the density function φ we used for the lower bound for the Closeness Problem is a useful density to use for a testing lower bound in the pairwise case. However it is not hard to distinguish between the fully uniform distribution and φ. Our trick is

giving φ a random "center". We remind the reader that we denote by $\varphi^{+t}(x) = \varphi(x \circ t)$ the distribution φ shifted by vector t. We claim that with $m = o(n/\delta^2)$ samples, it is hard to distinguish the fully uniform distribution from φ^{+t} with a uniformly randomly chosen t.

▶ **Lemma 10.** *Let φ be the density function defined by $\varphi(x) = 1 + \frac{\delta}{n}\sum_{i<j} x_i x_j$. Assume $m < n/\delta^2$. Let $\Phi : (\{-1,1\}^n)^m \to \mathbb{R}^{\geq 0}$ be the density associated to the distribution on m-tuples of strings defined as follows: First, choose t in $\{-1,1\}^n$ uniformly; then choose m strings independently from φ^{+t}. Let $\mathbf{1}$ denote the constantly 1 function on $(\{-1,1\}^n)^m$, the density associated to the uniform distribution. Then the χ^2-divergence between Φ and $\mathbf{1}$, $\|\Phi - \mathbf{1}\|_2^2$, is bounded by:*

$$\|\Phi - \mathbf{1}\|_2^2 \leq O\left(\frac{m\delta^2}{n}\right).$$

The proof of lemma 10 is included in the full paper.

Now we are ready to give the lower bound for sample complexity of testing fully uniform vs. far-from-pairwise uniform.

Proof of Theorem 5. If $m = o(n/\delta^2)$, by Lemma 10 we have $\|\Phi - \mathbf{1}\|_2^2 \leq o(1)$. Then any tester cannot distinguish, with more than $o(1)$ advantage, whether those m samples are fully uniform or they are drawn from φ^{+t} for some random t.

On the other hand, the proof of Theorem 2 shows that φ is $\Omega(\delta)$-far from pairwise uniform, and from the Fourier characterization, we have that φ^{+t} is pairwise uniform whenever φ is. We can conclude that testing fully uniform versus δ-far from pairwise-uniform needs sample complexity at least $\Omega(n/\delta^2)$. ◀

Unfortunately, we do not see an obvious way to generalize this lower bound to $k > 2$.

4 Testing αk-wise/fully uniform vs. far from k-wise uniform

4.1 The algorithm

In this section we show a sample-efficient algorithm for testing whether a distribution is αk-wise/fully uniform or δ-far from k-wise uniform. As a reminder, Theorem 9 indicates that the sample complexity of estimating $\mathbf{W}^{1\dots k}[\varphi]$ is bounded by the Fourier weight up to level $2k$. This suggests using a filter test to try to "kick out" those distributions with noticeable Fourier weight up to degree $2k$.

Filter Test. Draw m_1 samples from φ. If there exists a pair of samples $\boldsymbol{x}, \boldsymbol{y}$ such that $|\sum_{i=1}^n \boldsymbol{x}_i \boldsymbol{y}_i| > t\sqrt{n}$, output "Reject"; Otherwise, output "Accept".

The full algorithm is combining the Filter Test and $\mathbf{W}^{1\dots k}$ Estimation Test.

Full Algorithm. Do Filter Test with m_1 samples and parameter t. If it rejects, say "No". Otherwise, do $\mathbf{W}^{1\dots k}$ Estimation Test with m_2 samples and $\theta = (\delta/e^k)^2$. Say "No" if it outputs "$\mathbf{W}^{1\dots k}[\varphi] > \theta$" and say "Yes" otherwise.

Here "Yes" means φ is αk/fully uniform, and "No" means φ is δ-far from k-wise uniform. We will decide the parameters m_1, t, m_2 in the Filter Test and the full algorithm later.

For simplicity, we define $\overline{k} = \alpha k$. We will focus on testing αk-wise uniform vs. far from k-wise uniform in the analysis. For fully uniformity, the analysis is almost the same, and we will discuss it at the end of this subsection.

First of all, we show that if φ is $\overline{k}$-wise uniform, it will pass the Filter Test with high probability, when we choose m_1 and t properly.

▶ **Lemma 11.** *If φ is $\overline{k}$-wise uniform (assuming $\overline{k}$ is even), the Filter Test will accept with probability at least .9 when $m_1^2 \leq \frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}$.*

Secondly, we claim that for any distribution φ that does not get rejected by the Filter Test, it is close to a distribution φ' with upper bounds on Fourier weights of each of its levels.

▶ **Lemma 12.** *Any distribution φ either gets rejected by the Filter Test with probability at least .9, or there exists some distribution φ' such that:*

1. *φ' and φ are $\frac{8}{m_1}$-close in total variation distance;*
2. *$\mathbf{W}^i[\varphi'] \leq \frac{10^7}{m_1^2} n^i + t^i n^{i/2}$ for all $i = 1, \ldots, n$.*

If φ is not rejected by the Filter Test, Lemma 12 tells us that it is close to some distribution φ' with bounded Fourier weights on each of its levels. Even though we are drawing samples from φ, we can "pretend" that we are drawing samples from φ' since the are close.

▶ **Claim 13.** *Let $m_2 \leq \frac{m_1}{200}$, and let $A(X^{(m_2)})$ be any event related to m_2 variables in $\{-1, 1\}^n$, $X^{(m_2)} = \{x_1, \ldots, x_{m_2}\}$. Then we have*

$$\left| \Pr_{\boldsymbol{X}^{(m_2)} \sim \varphi} [A(\boldsymbol{X}^{(m_2)})] - \Pr_{\boldsymbol{X}^{(m_2)} \sim \varphi'} [A(\boldsymbol{X}^{(m_2)})] \right| \leq .08,$$

when φ and φ' are $\frac{8}{m_1}$-close.

Now we are ready to analyze the full algorithm.

We set the parameter $t = \left(10^{11}(4e^4)^k \overline{k}^{\overline{k}/2} \frac{n^k}{\delta^4}\right)^{\frac{1}{\overline{k}-2k}}$ and $m_1 = \sqrt{\frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}}$ in the Filter Test; and set $m_2 = \frac{1}{200} m_1$ and $\theta = \left(\frac{\delta}{e^k}\right)^2$ in the $\mathbf{W}^{1\ldots k}$ Estimation test.

In total we use $m_1 + m_2 = O\left(\sqrt{\frac{t^{\overline{k}}}{\overline{k}^{\overline{k}/2}}}\right)$ samples in the full algorithm. By plugging in the definition of t and $\overline{k} = \alpha k$, we can simplify the sample complexity to $O(\alpha)^{k/2} \cdot n^{k/2} \cdot \frac{1}{\delta^2} \cdot \left(\frac{n^k}{\delta^4}\right)^{1/(\alpha-2)}$.

By careful calculation, we derive that the full algorithm outputs the correct answer with probability at least 2/3.

For distinguishing between a distribution of being fully uniform and a distribution of being δ-far from k-wise uniform, the modification we need is that, in Lemma 11, we use Hoeffding's inequality and get

$$\Pr_{\boldsymbol{x}, \boldsymbol{y} \sim \varphi} \left[\left| \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i \right| > t\sqrt{n} \right] \leq 2e^{-t^2/2}$$

and then have the constraint $m_1^2 \leq \frac{1}{10} e^{t^2/2}$. Following exactly the same analysis, we get the same algorithm with sample complexity $O(k)^k \cdot n^{k/2} \cdot \frac{1}{\delta^2} \cdot \left(\log \frac{n}{\delta}\right)^{k/2}$.

The proofs of Lemma 11, Lemma 12, Claim 13 and Theorem 6 are in the appendix. The proof of Lemma 12 is included in the full version.

References

1 Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In *Advances in Neural Information Processing Systems*, pages 3591–3599, 2015.

2 Sarah R. Allen, Ryan O'Donnell, and David Witmer. How to refute a random CSP. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, pages 689–708, 2015.

3 Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 496–505, 2007.

4 Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *Journal of Algorithms*, 7(4):567–583, 1986.

5 Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

6 Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Information Processing Letters*, 88(3):107–110, 2003. doi:10.1016/S0020-0190(03)00359-4.

7 Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009. doi:10.1007/s00037-009-0272-6.

8 Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*, pages 442–451, 2001. doi:10.1109/SFCS.2001.959920.

9 Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing that distributions are close. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 259–269, 2000. doi:10.1109/SFCS.2000.892113.

10 Tuğkan Batu, Ravi Kumar, and Ronitt Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 381–390, 2004. doi:10.1145/1007352.1007414.

11 Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009. doi:10.1137/070691954.

12 Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *Journal of the ACM*, 57(5):28:1–28:10, 2010. doi:10.1145/1754399.1754401.

13 Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, pages 670–683, 2016. doi:10.1145/2897518.2897528.

14 Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem of t-resilient functions. In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, pages 396–407, 1985. doi:10.1109/SFCS.1985.55.

15 Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, pages 685–694. IEEE, 2016.

16 Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 68–75. Springer, 2011.

17 Richard M. Karp and Avi Wigderson. A fast parallel algorithm for the maximal independent set problem. *Journal of the ACM*, 32(4):762–773, 1985. doi:10.1145/4221.4226.

18 Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 132–145, 2017. doi:10.1145/3055399.3055485.

19 Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, pages 168–177. IEEE, 2016.

20 Michael Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM Journal on Computing*, 15(4):1036–1053, 1986. doi:10.1137/0215074.

21 Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error-correcting codes*. Elsevier, 1977.

22 Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993. doi:10.1137/0222053.

23 Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014.

24 Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008. doi:10.1109/TIT.2008.928987.

25 Calyampudi Radhakrishna Rao. Factorial experiments derivable from combinatorial arrangements of arrays. *Journal of the Royal Statistical Society*, 9(1):128–139, 1947.

26 Ronitt Rubinfeld and Rocco A. Servedio. Testing monotone high-dimensional distributions. *Random Structures & Algorithms*, 34(1):24–44, 2009. doi:10.1002/rsa.20247.

27 Ronitt Rubinfeld and Ning Xie. Robust characterizations of k-wise independence over product spaces and related testing results. *Random Structures & Algorithms*, 43(3):265–312, 2013.

28 Ning Xie. *Testing k-wise independent distributions*. PhD thesis, Massachusetts Institute of Technology, 2012.

A Proof omitted from Section 2

Proof of Theorem 1 (case $k = 1$). By identifying each x_i with $-x_i$ if necessary, we may assume without loss of generality that $\widehat{\varphi}(\{i\}) \geq 0$ for all i. In addition, by reordering the coordinates, we may assume without loss of generality that $0 \leq \widehat{\varphi}(\{1\}) \leq \cdots \leq \widehat{\varphi}(\{n\}) = \epsilon$. Define ψ_j to be the density of the distribution over $\{-1,1\}^n$ which is uniform on coordinates $x_1, \ldots, x_{j-1}$, and has x_i constantly fixed to be -1 for $j \leq i \leq n$. It is easy to check $\widehat{\psi_j}(\{i\}) = 0$ for $i < j$ and $\widehat{\psi_j}(\{i\}) = -1$ for $i \geq j$.

We define φ' as

$$\varphi' = \frac{1}{1+\epsilon}\left(\varphi + \sum_{j=1}^{n} w_j \psi_j\right),$$

where

$$w_1 = \widehat{\varphi}(\{1\}), \qquad w_j = \widehat{\varphi}(\{j\}) - \widehat{\varphi}(\{j-1\}) \quad \forall 1 < j \leq n.$$

It is easy to check that φ' is a density function and

$$\widehat{\varphi'}(\{i\}) = \frac{1}{1+\epsilon}\left(\widehat{\varphi}(\{i\}) + \left(\sum_{j=1}^{i} w_j\right)(-1)\right) = 0.$$

Therefore φ' is 1-wise uniform. Then by Lemma 7,

$$d_{TV}(\varphi, \text{1WISE}) \le \frac{1}{2}\|\varphi - \varphi'\|_1 \le \sum_{j=1}^{n} w_j = \epsilon. \qquad \blacktriangleleft$$

B Proof omitted from Section 3

Proof of Lemma 8. We denote $F(x,y) = \sum_{1\le|S|\le k} x^S y^S$. We know that

$$\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}[\boldsymbol{x}^S\boldsymbol{y}^S] = \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\varphi}[\boldsymbol{x}^S]\mathop{\mathbf{E}}_{\boldsymbol{y}\sim\varphi}[\boldsymbol{y}^S] = \widehat{\varphi}(S)^2,$$

when $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent samples drawn from φ. Therefore by linearity of expectation, $\mathbf{E}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}[F(\boldsymbol{x},\boldsymbol{y})] = \mathbf{W}^{1\dots k}[\varphi]$, and clearly by taking the average,

$$\mu = \mathbf{E}[\Delta(\boldsymbol{X})] = \mathbf{E}[\text{avg}_{s<t}F(\boldsymbol{x}_s,\boldsymbol{x}_t)] = \text{avg}_{s<t}\,\mathbf{E}[F(\boldsymbol{x}_s,\boldsymbol{x}_t)] = \mathbf{W}^{1\dots k}[\varphi].$$

We need to expand the variance:

$$\mathbf{Var}\left[\mathop{\text{avg}}_{s<t}(F(\boldsymbol{x}_s,\boldsymbol{x}_t))\right] = \frac{1}{\binom{m}{2}^2}\sum_{\substack{s<t\\ s'<t'}}\mathbf{Cov}[F(\boldsymbol{x}_s,\boldsymbol{x}_t),F(\boldsymbol{x}_{s'},\boldsymbol{x}_{t'})]. \tag{3}$$

We will discuss these covariances in three cases.

Case 1: $|\{s,t\}\cap\{s',t'\}| = 2$. Let $\boldsymbol{x},\boldsymbol{y}\sim\varphi$ be independent random variables.

$$\mathbf{Cov}[F(\boldsymbol{x},\boldsymbol{y}),F(\boldsymbol{x},\boldsymbol{y})] = \mathop{\mathbf{Var}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}[F(\boldsymbol{x},\boldsymbol{y})] \le \mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}[F(\boldsymbol{x},\boldsymbol{y})^2] = \mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{1\le|S|\le k}\boldsymbol{x}^S\boldsymbol{y}^S\right)^2\right].$$

Notice here all $\boldsymbol{x}_i$'s are Rademacher variables with $\boldsymbol{x}_i^2 = 1$, and similarly for the $\boldsymbol{y}_i$'s. Therefore

$$\begin{aligned}\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{1\le|S|\le k}\boldsymbol{x}^S\boldsymbol{y}^S\right)^2\right] &= \sum_{1\le|S_1|,|S_2|\le k}\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\boldsymbol{x}^{S_1\oplus S_2}\boldsymbol{y}^{S_1\oplus S_2}\right]\\ &= \sum_{1\le|S_1|,|S_2|\le k}\widehat{\varphi}(S_1\oplus S_2)^2 = L_k(\varphi).\end{aligned}$$

Case 2: $|\{s,t\} \cap \{s',t'\}| = 1$. Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \sim \varphi$ be independent random variables. Similar to Case 1, we have:

$$\begin{aligned}
\mathbf{Cov}[F(\boldsymbol{x},\boldsymbol{y}), F(\boldsymbol{x},\boldsymbol{z})] &\le \mathbf{E}[F(\boldsymbol{x},\boldsymbol{y})F(\boldsymbol{x},\boldsymbol{z})] \\
&= \mathbf{E}\left[\left(\sum_{1\le|S_1|\le k} \boldsymbol{x}^{S_1}\boldsymbol{y}^{S_1}\right)\left(\sum_{1\le|S_2|\le k} \boldsymbol{x}^{S_2}\boldsymbol{z}^{S_2}\right)\right] \\
&= \mathbf{E}\left[\sum_{1\le|S_1|,|S_2|\le k} \boldsymbol{x}^{S_1\oplus S_2}\boldsymbol{y}^{S_1}\boldsymbol{z}^{S_2}\right] \\
&= \sum_{1\le|S_1|,|S_2|\le k} \widehat{\varphi}(S_1\oplus S_2)\widehat{\varphi}(S_1)\widehat{\varphi}(S_2) \\
&\le \sqrt{\sum_{1\le|S_1|,|S_2|\le k} \widehat{\varphi}(S_1\oplus S_2)^2}\sqrt{\sum_{1\le|S_1|,|S_2|\le k} \widehat{\varphi}(S_1)^2\widehat{\varphi}(S_2)^2} \\
&= \sqrt{L_k(\varphi)}\mu,
\end{aligned}$$

where the inequality comes from Cauchy–Schwarz.

Case 3: $|\{s,t\} \cap \{s',t'\}| = 0$. Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w} \sim \varphi$ be independent random variables. Clearly $F(\boldsymbol{x},\boldsymbol{y})$ and $F(\boldsymbol{z},\boldsymbol{w})$ are independent and therefore $\mathbf{Cov}[F(\boldsymbol{x},\boldsymbol{y}), F(\boldsymbol{z},\boldsymbol{w})] = 0$.

Plugging all these cases into eq. (3), we get

$$\begin{aligned}
\mathbf{Var}[\Delta(\boldsymbol{X})] &= \mathbf{Var}\left[\operatorname*{avg}_{s<t}(F(\boldsymbol{x}_s,\boldsymbol{x}_t)\right] \\
&= \frac{1}{\binom{m}{2}^2}\left(\binom{m}{2}L_k(\varphi) + m(m-1)(m-2)\sqrt{L_k(\varphi)}\mu\right) \\
&\le \frac{4}{m^2}L_k(\varphi) + \frac{4}{m}\sqrt{L_k(\varphi)}\mu.
\end{aligned}$$

◀

Proof of Theorem 9. The algorithm is simple: we report "$\mu \le \frac{1}{2}\theta$" if $\Delta(\boldsymbol{X}) \le \frac{3}{4}\theta$ and report "$\mu > \theta$" if $\Delta(\boldsymbol{X}) > \frac{3}{4}\theta$.

Now we need to bound $L_k(\varphi)$ to bound the variance of $\Delta(\boldsymbol{X})$. For a fixed subset $|S| \le 2k$, how many pairs of $1 \le |S_1|, |S_2| \le k$ are there satisfying $S = S_1 \oplus S_2$? We denote $S_1 = S_1' \cup T$, $S_2 = S_2' \cup T$, where S_1', S_2', T are disjoint. Then $S = S_1' \cup S_2'$. For a fixed set S, there are at most $2^{|S|}$ different ways to split it into two sets S_1', S_2'. Because $\max\{S_1', S_2'\} \ge \lceil |S|/2 \rceil$ and $|S_1|, |S_2| \le k$, we have $|T| \le k - \lceil |S|/2 \rceil$. Therefore there are at most

$$\sum_{j=0}^{k-\lceil|S|/2\rceil} \binom{n-|S|}{j} \le \frac{2n^{k-\lceil|S|/2\rceil}}{(k-\lceil|S|/2\rceil)!}$$

ways to choose the set T for any fixed S_1', S_2'. Hence,

$$\begin{aligned}
L_k(\varphi) &= \sum_{1\le|S_1|,|S_2|\le k} \widehat{\varphi}(S_1 \oplus S_2)^2 \\
&= \sum_{|S|\le 2k} \sum_{\substack{S_1'\cap S_2'=\emptyset \\ S_1'\cup S_2'=S}} \sum_{\substack{T\cap S_1'=\emptyset, T\cap S_2'=\emptyset \\ |T|+\max\{|S_1'|,|S_2'|\}\le k}} \widehat{\varphi}(S)^2 \\
&\le \sum_{|S|\le 2k} 2^{|S|} \frac{2n^{k-\lceil |S|/2\rceil}}{(k-\lceil |S|/2\rceil)!}\widehat{\varphi}(S)^2 \\
&= \sum_{i=0}^{2k} 2^i \frac{2n^{k-\lceil i/2\rceil}}{(k-\lceil i/2\rceil)!}\mathbf{W}^i[\varphi].
\end{aligned}$$

Plugging in $\mathbf{W}^i[\varphi] \le An^{i/2}$, we get

$$L_k(\varphi) \le \sum_{i=0}^{2k} 2^i \frac{2n^{k-\lceil i/2\rceil}}{(k-\lceil i/2\rceil)!}\mathbf{W}^i[\varphi] \le 2^{2k+2}An^k. \tag{4}$$

By substituting eq. (4) and eq. (2) into eq. (1), we have

$$\mathbf{Var}[\Delta(\boldsymbol{X})] \le \frac{4}{500^2}\theta^2 + \frac{4}{500}\theta\mu \le \frac{1}{64}\max\{\theta^2, \mu^2\}.$$

Then we conclude our proof by Chebyshev's inequality:

$$\begin{aligned}
\mathbf{Pr}\left[|\Delta(\boldsymbol{X}) - \mu| \le \frac{1}{4}\max\{\theta,\mu\}\right] &\ge \mathbf{Pr}\left[|\Delta(\boldsymbol{X}) - \mu| \le 2\sqrt{\mathbf{Var}[\Delta(\boldsymbol{X})]}\right] \\
&\ge 1 - \left(\frac{1}{2}\right)^2 = \frac{3}{4}.
\end{aligned}$$

◀

C Proofs omitted in Section 4

Proof of Lemma 11. If φ is $\overline{k}$-wise uniform with $\overline{k}$ even, then by Markov's inequality on the $\overline{k}$-th moment, we have

$$\mathop{\mathbf{Pr}}_{\substack{\boldsymbol{x},\boldsymbol{y}\sim\varphi \\ \text{independent}}}\left[\left|\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right| > t\sqrt{n}\right] = \mathop{\mathbf{Pr}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right)^{\overline{k}} > (t\sqrt{n})^{\overline{k}}\right] \le \frac{\mathbf{E}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right)^{\overline{k}}\right]}{t^{\overline{k}}n^{\overline{k}/2}}.$$

When we expand $\left(\sum_{i=1}^n x_iy_i\right)^{\overline{k}}$, each term is at most degree $\overline{k}$ in x or y. Because $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent random variables chosen from $\overline{k}$-wise uniform distribution φ, the whole polynomial behaves the same as if $\boldsymbol{x}$ and $\boldsymbol{y}$ were chosen from the fully uniform distribution:

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left(\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{y}_i\right)^{\overline{k}}\right] &= \mathop{\mathbf{E}}_{\boldsymbol{z}\sim\{-1,1\}^n}\left[\left(\sum_{i=1}^n \boldsymbol{z}_i\right)^{\overline{k}}\right] \\
&\le \overline{k}^{\overline{k}/2}\left(\mathop{\mathbf{E}}_{\boldsymbol{z}\sim\{-1,1\}^n}\left[\left(\sum_{i=1}^n \boldsymbol{z}_i\right)^2\right]\right)^{\overline{k}/2} \\
&= \overline{k}^{\overline{k}/2}n^{\overline{k}/2}.
\end{aligned}$$

The inequality uses hypercontractivity; see Theorem 9.21 in [23]. Hence we have

$$\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left|\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{y}_i\right| > t\sqrt{n}\right] \leq \frac{\overline{k}^{\overline{k}/2}}{t^{\overline{k}}}.$$

When drawing m_1 examples, there are at most $\binom{m_1}{2} \leq \frac{1}{2}m_1^2$ pairs. Hence by the union bound, the probability of φ getting rejected is at most $\frac{m_1^2\overline{k}^{\overline{k}/2}}{2t^{\overline{k}}} \leq \frac{1}{10}$. ◀

Proof of Claim 13. We denote by Φ(respectively, Φ') the joint distribution of m_2 samples from φ(respectively, φ'). Then by a union bound we know that Φ and Φ' are .04-close, since $m_2\frac{8}{m_1} \leq .04$. We denote $\mathbf{1}[A(\boldsymbol{X}^{(m_2)})]$ as the indicator function of event A happening on $\boldsymbol{X}^{(m_2)}$. Then we have

$$\begin{aligned}\left|\Pr_{\boldsymbol{X}^{(m_2)}\sim\varphi}[A(\boldsymbol{X}^{(m_2)})] - \Pr_{\boldsymbol{X}^{(m_2)}\sim\varphi'}[A(\boldsymbol{X}^{(m_2)})]\right| &= \left|\sum_{X^{(m_2)}}\mathbf{1}[A(X^{(m_2)})]\left(\Phi(X^{(m_2)}) - \Phi'(X^{(m_2)})\right)\right| \\ &\leq \sum_{X^{(m_2)}}\left|\Phi(X^{(m_2)}) - \Phi'(X^{(m_2)})\right| \\ &= 2d_{\mathrm{TV}}(\Phi,\Phi') \leq .08\end{aligned}$$

which completes the proof. ◀

Proof of Theorem 6. We discuss distinguishing between $\overline{k}$-wise uniform and δ-far from k-wise uniform first. In the Overall Algorithm, we set the parameters $t = \left(10^{11}(4e^4)^k\overline{k}^{\overline{k}/2}\frac{n^k}{\delta^4}\right)^{\frac{1}{\overline{k}-2k}}$ and $m_1 = \sqrt{\frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}}$ in the Filter Test; and, we set $m_2 = \frac{1}{200}m_1$ and $\theta = \left(\frac{\delta}{e^k}\right)^2$ in the $\mathbf{W}^{1\ldots k}$ Estimation test.

In total we use $m_1 + m_2 = O\left(\sqrt{\frac{t^{\overline{k}}}{\overline{k}^{\overline{k}/2}}}\right)$ samples in the Overall Algorithm. By plugging in the definition of t and $\overline{k} = \alpha k$, we can simplify the sample complexity to $O(\alpha)^{k/2}\cdot n^{k/2}\cdot \frac{1}{\delta^2}\cdot\left(\frac{n^k}{\delta^4}\right)^{1/(\alpha-2)}$.

The rest of the proof is to show the correctness of this algorithm. We discuss the two cases.

"Yes" case: Suppose φ is $\overline{k}$-wise uniform. By Lemma 11 we know that φ will pass the Filter Test with probability at least .9 since $m_1^2 = \frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}$.

Now φ is $\overline{k}$-wise uniform with $\overline{k} > 2k$, which means $\widehat{\varphi}(S) = 0$ for any $1 \leq |S| \leq 2k$. Therefore by setting $\delta = \left(\frac{\theta}{e^k}\right)^2$ and $A = 1$, Theorem 9 tells us that m_2 samples are large enough for $\mathbf{W}^{1\ldots k}$ Estimation Test to output "$\mathbf{W}^{1\ldots k}[\varphi] \leq \frac{1}{2}\theta$" with probability 3/4.

The overall probability of the Overall Algorithm saying "Yes" is therefore at least $.9 \times \frac{3}{4} > \frac{2}{3}$.

"No" case: Suppose φ is δ-far from k-wise uniform. Either φ gets rejected by the Filter Test with probability .9, or according to Lemma 12, we know that there exists some distribution φ' which is $\frac{8}{m_1}$-close to φ and $\mathbf{W}^i[\varphi'] \leq \frac{10^7}{m_1^2}n^i + t^i n^{i/2}$ for all $i = 1,\ldots,n$.

The second stage is slightly tricky. As described in Claim 13, at the expense of losing .08 probability, we may pretend we are drawing samples from φ' rather than φ. Notice that $m_1^2 = \frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}} = \omega(n^k)$. We have

$$\mathbf{W}^i[\varphi'] \le \frac{10^7}{m_1^2} n^i + t^i n^{i/2} = (1 + o(1))t^i n^{i/2} \le A n^{i/2}$$

for $i = 0, \ldots, 2k$ with parameter $A = 1.01t^{2k}$. Then plugging $A = 1.01t^{2k}$ and $\theta = \left(\frac{\delta}{e^k}\right)^2$ into Theorem 9, we know that the $\mathbf{W}^{1\cdots k}$ Estimation Test will say "$\mathbf{W}^{1\cdots k}[\varphi] > \theta$" with probability at least $\frac{3}{4}$ when φ' is δ-far from k-wise uniform, provided we have at least $1005\frac{(2e^2)^k t^k n^{k/2}}{\delta^2}$ samples. It is easy to check $m_2 = \frac{1}{200}\sqrt{\frac{t^{\overline{k}}}{5\overline{k}^{\overline{k}/2}}}$ is sufficient.

However, in the real algorithm we are drawing samples from φ rather than φ'. From Claim 13, we know that the estimator will accept with probability at least $\frac{3}{4} - .08 > \frac{2}{3}$ when φ' is δ-far from k-wise uniform. Notice that φ and φ' are $\frac{8}{m_1}$-close, where $\frac{8}{m_1} = o\left(\frac{\delta^4}{n^k}\right)$. Hence if φ is δ-far from k-wise uniform, φ' is also δ-far from k-wise uniform, which completes the proof.

Finally, for distinguishing between a distribution being fully uniform and a distribution being δ-far from k-wise uniform, the modification we need is that in Lemma 11 we use Hoeffding's inequality to get

$$\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\varphi}\left[\left|\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{y}_i\right| > t\sqrt{n}\right] \le 2e^{-t^2/2},$$

and then we have the constraint $m_1^2 \le \frac{1}{10}e^{t^2/2}$. Following exactly the same analysis, we get the same algorithm with sample complexity $O(k)^k \cdot n^{k/2} \cdot \frac{1}{\delta^2} \cdot \left(\log \frac{n}{\delta}\right)^{k/2}$. ◀