Testing k-wise and Almost k-wise Independence

Noga Alon[*]
Tel Aviv University
nogaa@tau.ac.il

Alexandr Andoni
MIT
andoni@mit.edu

Tali Kaufman
MIT
kaufmant@mit.edu

Kevin Matulef[†]
MIT
matulef@mit.edu

Ronitt Rubinfeld[‡]
MIT
ronitt@theory.csail.mit.edu

Ning Xie[§]
State Univ. of New York at Buffalo
ningxie@gmail.com

November 20, 2006

Abstract

In this work, we consider the problems of testing whether a distribution over $\{0,1\}^n$ is k-wise or (ϵ, k)-wise independent using samples drawn from that distribution.

To distinguish k-wise independent distributions from those that are δ-far in statistical distance, we upper bound the number of required samples by $\tilde{O}(n^k/\delta^2)$ and lower bound it by $\Omega(n^{\frac{k-1}{2}}/\delta)$ (these bounds hold for constant k, and essentially the same bounds hold for general k). To achieve these bounds, we use Fourier analysis to relate a distribution's distance from k-wise independence to its *biases* (a measure of the parity imbalance it induces on a set of variables). The relationships we derive are tighter than previously known, and are of independent interest.

To distinguish (ϵ, k)-wise independent distributions from those that are δ-far in statistical distance, we upper bound the number of required samples by $O\left(\frac{k \log n}{\delta^2 \epsilon^2}\right)$ and lower bound it by $\Omega\left(\frac{\sqrt{k \log n}}{(\epsilon+\delta)\sqrt{\log 1/(\epsilon+\delta)}}\right)$. Although these bounds are an exponential improvement (in terms of n and k) over the corresponding bounds for testing k-wise independence, we show that the *time* complexity of testing (ϵ, k)-wise independence is unlikely to be $\mathrm{poly}(n, 1/\epsilon, 1/\delta)$ for $k = \Theta(\log n)$, since this would disprove a plausible conjecture about the hardness of finding hidden cliques in random graphs. Under the conjecture, our result implies that for, say, $k = \log n$ and $\epsilon = 1/n^{0.99}$, there is a set of (ϵ, k)-wise independent distributions, and a set of distributions at distance $\delta = 1/n^{0.51}$ from (ϵ, k)-wise independence, which are indistinguishable by polynomial time algorithms.

[*] Research supported in part by the Israel Science Foundation, by a USA-Israeli BSF grant, and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

[†] Supported by an NSF graduate fellowship.

[‡] Supported by NSF grant 012702-001.

[§] Work done while visiting CSAIL, MIT.

1 Introduction

A probability distribution over $\{0,1\}^n$ is *k-wise independent* if its restriction to any k coordinates is uniform. Similarly a distribution is *(ϵ, k)-wise independent* if, roughly, its restriction to any k coordinates is almost uniform. Such distributions look random "locally," to an observer of only k coordinates, even though they may be far from random "globally." Because of this key feature, k-wise and (ϵ, k)-wise independent distributions are important concepts in probability, complexity, and algorithm design [19, 21, 24, 25].

Given samples drawn from a distribution over $\{0,1\}^n$, it is natural to wonder whether the distribution generating those samples is k-wise independent. An experimenter, for example, who receives data in the form of a vector of n bits might like to know whether every setting of k of those bits is equally likely to occur, or whether some settings of k bits are more or less likely.

In this work, we seek new ways of elucidating the structure of k-wise independent distributions, and of analyzing a distribution's statistical distance to k-wise independence. We use our new understanding to develop efficient algorithms for *testing* k-wise and (ϵ, k)-wise independence – that is, algorithms that with high probability accept distributions which are k-wise independent, and reject distributions which are δ-far in statistical distance from any k-wise independent distribution.

Previous work addressed the problem of testing related properties of distributions, including uniformity [17, 8] and independence [7, 26, 9]. Although we are unaware of any previous work on testing k-wise and (ϵ, k)-wise independence, the theorems in [4] combined with a generalization of the algorithm in [17] yield natural testing algorithms which we improve upon.

1.1 Our Results and Techniques

The formal definition of a testing algorithm for k-wise or (ϵ, k)-wise independent distributions is given below. The complexity of a testing algorithm is measured both in terms of the number of samples required (sample complexity), and the computational time needed to process those samples (time complexity).

Definition 1.1 (Testing k-wise ((ϵ, k)-wise) independence)**.** *Let* $0 < \epsilon, \delta < 1$*, and let D be a distribution over* $\{0,1\}^n$*. We say that an algorithm* tests k-wise ((ϵ, k)-wise) independence *if, given access to a set* $Q \subset \{0,1\}^n$ *of samples drawn independently from D, it outputs: 1) "Yes" if D is a k-wise ((ϵ, k)-wise) independent distribution, 2) "No" if the statistical distance of D to any k-wise ((ϵ, k)-wise) independent distribution is at least δ. The tester may fail to give the right answer with probability at most* $1/3$*. We call* $|Q|$ *the* query complexity *of the algorithm.*

In Table 1, we summarize the sample and time bounds that our algorithms achieve, along with the lower bounds that we prove for the associated testing problems. In interpreting these results, it is useful to think of δ and ϵ as constants, so that the complexity measures are functions of only n and k. The O^* and Ω^* notation is defined as follows: $O^*(f) = O(f^{1+o(1)})$ and $\Omega^*(f) = \Omega(f^{1-o(1)})$. For constant k, one can replace the O^* and Ω^* in the statement of our results with $\tilde{O}$ and Ω respectively.

1.1.1 Testing k-wise independence

In Section 3, we present an algorithm for testing k-wise independence. We use the notion of a *bias over a set T* which is a measure of the parity imbalance of the distribution over the set T of variables:

Definition 1.2. *For a distribution D over* $\{0,1\}^n$*, the* bias *of D over a non-empty set $T \subseteq [n]$ is defined as* $bias_D(T) \triangleq \Pr_{x \leftarrow D}[\oplus_{i \in T} x_i = 0] - \Pr_{x \leftarrow D}[\oplus_{i \in T} x_i = 1]$*. We say $bias_D(T)$ is an l-th level bias if $|T| = l$.*

A well-known fact says that a distribution is k-wise independent iff its biases $bias_D(T)$ are zero for all nonempty sets $T \subset [n]$ of size at most k.

This suggests the following simple algorithm: estimate all the *biases* of the distribution over sets of size up to k and output "Yes" iff all of those biases are small enough. We show that this algorithm has $O^*(n^k/\delta^2)$

Table 1: Summary of Testing Results

	Reference	Sample Complexity		Time Complexity	
		Upper	Lower	Upper	Lower
Testing k-wise independence	this paper	$O^*(\frac{n^k}{\delta^2})$	$\Omega^*(\frac{n^{\frac{k-1}{2}}}{\delta})$	$O^*(\frac{n^{2k}}{\delta^2})$	-
	[4]†	$O^*(\frac{n^{2k}}{\delta^2})$	$\Omega^*(n^{\frac{k}{4}})$	$O^*(\frac{n^{3k}}{\delta^2})$	-
Testing (ϵ, k)-wise independence	this paper	$O(\frac{k \log n}{\delta^2 \epsilon^2})$	$\Omega\left(\frac{\sqrt{k \log n}}{(\epsilon+\delta)\sqrt{\log \frac{1}{(\epsilon+\delta)}}}\right)$	$\frac{n^{O(k)}}{\text{poly}(\epsilon,\delta)}$ ‡	$n^{\omega(1)}$ §

†These bounds can be derived from theorems in [4], though they did not explicitly consider the testing problem.

‡This can be achieved trivially.

§This lower bound applies when $k = \Theta(\log n)$ and $\epsilon\delta = n^{-\Theta(1)}$. It is contingent upon a conjecture discussed below.

sample complexity and $O^*(n^{2k}/\delta^2)$ time complexity. We also prove a sample complexity lower bound of $\Omega^*(n^{\frac{k-1}{2}}/\delta)$, showing our upper bound is at most a quadratic factor from optimal.

The analysis of our testing algorithm is based on Theorem 3.1. Let $\Delta(D, \mathcal{D}_{\text{kwi}})$ denote the statistical distance between distribution D and the set of k-wise independent distributions $\mathcal{D}_{\text{kwi}}$. Theorem 3.1 shows that $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq O\left(\sqrt{\sum_{|T|\leq k}(bias(T))^2} \log^{k/2} n\right)$. Previously, the only non-trivial bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ is the one implicit in [4]: $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \sum_{|T|\leq k} |bias(T)|$. In most of the interesting cases, our new bound improves upon their result. For example, the main upper bound result in [4] is: if all the biases of a distribution D over non-empty subsets up to size k are at most ϵ, then $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq n^k \cdot \epsilon$. Using Theorem 3.1, this can be improved to $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq O((\sqrt{n \log n})^k) \cdot \epsilon$.

Our sample lower bound is based on a Random Distribution Lemma (Lemma 3.6), which shows that a uniform distribution over a random set of size $O\left(\frac{(n/k)^{k-1}}{\delta^2}\right)$ is almost surely δ-far from any k-wise independent distribution. In contrast, the lower bound result in [4] shows that any distribution with support size $O\left(\frac{n^{k/2}}{k^k}\right)$ is always 1/2-far from any k-wise independent distribution. Our result applies to random uniform distributions over a large range of support sizes, and shows a tradeoff between a distribution's support size and its distance to k-wise independent distributions.

Fourier-analytic interpretation of our bounds on $\Delta(D, \mathcal{D}_{\text{kwi}})$.

Our upper and lower bounds on $\Delta(D, \mathcal{D}_{\text{kwi}})$, together with the proof techniques, may be of independent interest when interpreted as Fourier-analytic inequalities for bounded functions on the hypercube. The harmonic analysis of such functions has been considered in the Computer Science literature, e.g., in [14]. The connection to Fourier analysis comes from the basic fact that the biases of a distribution D are equal to D's Fourier coefficients (up to a normalization factor).

Bounds on $\Delta(D, \mathcal{D}_{\text{kwi}})$ may be viewed as part of the following general question: fix a family F of functions on the hypercube and a subfamily $H \subset F$ of functions defined via a restriction on their Fourier coefficients. Then, for function $f \in F$, what is the ℓ_1 distance from f to its projection in H, i.e., $\ell_1(f, H)$?[1] In our case F is the set of all bounded functions that sum up to 1 (i.e., distributions), and H further requires that the functions have no non-zero Fourier coefficients over non-empty subsets of size at most k. Then, for example, Parseval's equality gives the following bound on the ℓ_2-norm: $\ell_2(f, H) \geq \|f_{\leq k}\|_2$ where $f_{\leq k}(x) \triangleq \sum_{0<|S|\leq k} \hat{f}_S \chi_S(x)$ is the truncation of f to the low-level Fourier spectrum (the inequality would be an equality if the functions were not bounded). Unfortunately, such a bound implies only very weak bounds for the ℓ_1-norm.

[1] The distance of a function to a set, $\ell_p(f, H)$, is defined to be $\min_{h \in H} \|f - h\|_p$.

In contrast, our upper bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ says that $\ell_1(f, H) \leq \|f_{\leq k}\|_2 \cdot O(\log^{k/2} n)$. To prove such an inequality, we proceed as follows. Given a distribution $D = f$, we approximate D using a function D_1, obtained by forcing all of D's first k-level Fourier coefficients to zero while keeping all others unchanged. Although D_1 is not necessarily a probability distribution (it may map some inputs to negative values), we show how to turn it back into a k-wise independent distribution by "mending" it with a series of carefully chosen small weight k-wise independent distributions. By a deep result in Fourier analysis, the Bonami-Beckner inequality, we bound the distance incurred by the "mending" process. Thus, we are able to bound the total ℓ_1 distance of D to k-wise independence by the distance from D to D_1 plus the "mending" cost.

Furthermore, our lower bound technique (employed by the Random Distribution Lemma) implies that $\ell_1(f, H) \geq \frac{\|f_{\leq k}\|_2}{\|f_{\leq k}\|_\infty}$, which is already useful when we take f to be a uniform function on a randomly chosen support. This inequality follows by taking the convolution of $D = f$ with an auxiliary function and then applying Young's convolution inequality to lower bound the ℓ_1-norm of $D - D'$, where D' is the k-wise independent distribution closest to D.

1.1.2 Testing (ϵ, k)-wise independence

In Section 4, we give an algorithm for testing (ϵ, k)-wise independence that uses $O(k \log n / \delta^2 \epsilon^2)$ samples, and we show that $\Omega\left(\frac{\sqrt{k \log n}}{(\epsilon+\delta)\sqrt{\log 1/(\epsilon+\delta)}}\right)$ samples are required. The lower bound on the sample complexity is achieved by obtaining an $\Omega\left(\frac{k \log n}{\epsilon^2 \log(1/\epsilon)}\right)$ lower bound on the support size of a (ϵ, k)-wise independent distribution. The proof of the lower bound uses significantly different ideas from the lower bound for testing k-wise independence.

In terms of n and k, the sample complexity of testing (ϵ, k)-wise independence is exponentially better than that of testing k-wise independence. However, the time complexity of testing (ϵ, k)-wise independence presents another story. Since the number of samples required by our testing algorithm is only $\text{poly}(n/\epsilon\delta)$, one would hope that the time complexity is polynomial as well. However, we show that for some k this is not likely to be the case. Specifically, in Theorem 4.4 we show that for $k = \Theta(\log n)$ and $\epsilon\delta = n^{-O(1)}$, no polynomial time tester exists for this testing problem, under a plausible conjecture on the hardness of finding a hidden clique in random graphs. Finding hidden cliques in random graphs has been studied since [18, 23]. We discuss our conjecture in detail in Section 4.

Computational indistinguishability of (ϵ, k)-wise independent distributions.

The initial motivation of [4] was to show that a randomized algorithm requiring only k-wise independent distributions (i.e., $O(k \log n)$ random bits) can be further derandomized using (ϵ, k)-wise independent distributions (requiring only $O(k + \log(n/\epsilon))$ random bits), by showing that any (ϵ, k)-wise independent distribution is close in statistical distance to some k-wise independent distribution for $\epsilon = 1/poly(n, 2^k)$. They instead proved that an (ϵ, k)-wise independent distribution can be at distance $\geq 1/2$ from k-wise independence even for ϵ as small as $\epsilon = n^{-k/5}$. One can view their results as showing that k-wise (i.e., $(0, k)$-wise) and $(n^{-k/5}, k)$-wise independent distributions are far apart information-theoretically.

Despite the large statistical distance, one can ask whether there are $(1/poly(n, 2^k), k)$-wise independent distributions that are poly-time indistinguishable from $(0, k)$-wise independence, under some computational hardness assumption (such (ϵ, k)-wise independent distributions should still require $O(k + \log(n/\epsilon))$ random bits to be useful for derandomization). Although we do not answer the above question or give a result useful for derandomization, our above hardness of testing result yields some evidence for an affirmative answer. Specifically, we show that for, say, $k = \log n$, there is a family of $(n^{-0.99}, k)$-wise independent distributions, and a family of $(n^{-0.51}, k)$-wise independent distributions that are poly-time indistinguishable under the aforementioned hidden clique conjecture. Even though any distribution from the first family is at distance $\delta \geq n^{-0.52}$ from any distribution from the second family (as we show), the conjecture implies that distinguishing a random distribution from the first family from a random member of the second cannot be done in polynomial time with a polynomial number of samples.

2 Preliminaries

We use $[n]$ to denote the set $\{1, \ldots, n\}$. For an integer $k = o(n)$, define $M_{n,k} = \sum_{i=1}^{k} \binom{n}{i}$ to be the number of non-empty subsets of $[n]$ of size at most k. Then $M_{n,k} \leq n^k$ and $M_{n,k} = \Omega^*(n^k)$.

We will restrict our attention to probability distributions over $\{0,1\}^n$ which are specified by distribution functions $D : \{0,1\}^n \to [0,1]$ such that $\sum_{x \in \{0,1\}^n} D(x) = 1$. The *support* of D, $\mathrm{Supp}(D)$, is the set of points x at which $D(x) \neq 0$. Let $A = \{a_1, \ldots, a_m\}$ be a multiset of cardinality m, where $a_i \in \{0,1\}^n$. The uniform distribution over A, denoted U_A, is defined to be $U_A(x) = \frac{|\{i \in [m] : a_i = x\}|}{m}$. We use U_n to denote the uniform distribution over $\{0,1\}^n$.

2.1 k-wise and (ϵ, k)-wise Independent Distributions, and Distances

Definition 2.1. *A distribution D is (ϵ, k)*-wise independent *if for any k indexes $i_1 < i_2 < \ldots < i_k$, and any vector $\overrightarrow{v} \in \{0,1\}^k$ of k bits, $\left|\Pr_{x \leftarrow D}\left[x_{i_1} x_{i_2} \ldots x_{i_k} = \overrightarrow{v}\right] - 2^{-k}\right| \leq \epsilon$. When $\epsilon = 0$, we say that D is just k*-wise independent*. The set of all k-wise independent distributions and (ϵ, k)-wise independent distributions are denoted by $\mathcal{D}_{kwi}$ and $\mathcal{D}_{(\epsilon,k)}$ respectively.*

For two distributions D_1, D_2, we denote their statistical distance by $\Delta(D_1, D_2) \triangleq \max_{S \subseteq \{0,1\}^n} |\Pr[D_1(S)] - \Pr[D_2(S)]|$. It is immediate to verify that $\Delta(D_1, D_2) = \frac{1}{2} \sum_x |D_1(x) - D_2(x)|$ and $0 \leq \Delta(D_1, D_2) \leq 1$.

The *distance of a distribution D to k-wise independence*, denoted $\Delta(D, \mathcal{D}_{\mathrm{kwi}})$, is defined to be the minimum statistical distance of D to any k-wise independent distribution, i.e. $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \triangleq \min_{D' \in \mathcal{D}_{\mathrm{kwi}}} \Delta(D, D')$. If $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq \delta$, we say D is *δ-close* to k-wise independence. Otherwise, we say D is *δ-far*. These concepts are defined identically for (ϵ, k)-wise independence, with $\mathcal{D}_{(\epsilon,k)}$ in place of $\mathcal{D}_{\mathrm{kwi}}$.

2.2 The Fourier Transform and the Bonami-Beckner Inequality

The set of functions $f : \{0,1\}^n \to \mathbb{R}$ is a vector space of dimension 2^n in which the inner product between two elements f and g is defined as $\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) g(x)$. For each $S \subseteq [n]$, define the character $\chi_S : \{0,1\}^n \to \{-1, 1\}$ as $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. The set of 2^n functions, $\{\chi_S : S \subseteq [n]\}$, forms an orthonormal basis for the vector space. This implies that any function $f : \{0,1\}^n \to \mathbb{R}$ can be expanded uniquely as $f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x)$, where $\hat{f}(S) = \langle f, \chi_S(x) \rangle$ is the Fourier coefficient of f over set S. The p-norm of f is $\|f\|_p = \left(\frac{1}{2^n} \sum_{x \in \{0,1\}^n} |f(x)|^p\right)^{1/p}$. Parseval's equality, $\|f\|_2^2 = \sum_{S \subseteq [n]} \hat{f}(S)^2$, follows directly from the orthonormality of the basis.

For two functions $f, g : \{0,1\}^n \to \mathbb{R}$, their *convolution* is defined as $(f * g)(x) \triangleq \frac{1}{2^n} \sum_{y \in \{0,1\}^n} f(y) g(x - y)$. It is easy to show that $\widehat{fg} = \hat{f} \hat{*} \hat{g}$ and $\widehat{f * g} = \hat{f} \hat{g}$ for any $f, g : \{0,1\}^n \to \mathbb{R}$. It is also easy to show that $\|f * g\|_\infty \leq \|f\|_\infty \|g\|_1$, which is a simple special case of Young's convolution inequality.

A powerful tool in Fourier analysis over $\{0,1\}^n$ is the hyper-contractive estimate due independently to Beckner [10] and Bonami [12]. The following is one form, proven in [12]:

Theorem 2.2. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function that is a linear combination of $\{\chi_T : |T| \leq k\}$. Then, for any even $p > 2$, $\|f\|_p \leq \left(\sqrt{p-1}\right)^k \|f\|_2$.*

2.3 Characterizing k-wise Independence Using Biases

Up to a normalization factor, the biases are equal to the Fourier coefficients of the distribution function D. More precisely, $\hat{D}(T) = \frac{1}{2^n} \mathit{bias}_D(T)$, for $T \neq \emptyset$. Thus, we sometimes use the terms biases and Fourier coefficients interchangeably. The following well-known facts relate biases to k-wise independence:

Fact 2.3. *A distribution is k-wise independent iff all the biases over sets $T \subset [n]$, $0 < |T| \leq k$, are zero. In particular, for the uniform distribution U_n, $bias_{U_n}(T) = 0$ for all T.*

Fact 2.4. $\Delta(D, \mathcal{D}_{kwi}) \geq \frac{1}{2} \max_{T \subseteq [n], 0<|T| \leq k} bias_D(T)$.

3 Testing k-wise independence

In this section, we study the problem of testing whether a distribution is k-wise independent or δ-far from from k-wise independence. Our upper bound and lower bound results for testing are based on new upper and lower bounds on $\Delta(D, \mathcal{D}_{\text{kwi}})$ in term of D's first k-level biases. We present our upper bounds in Section 3.1 and lower bounds in Section 3.2.

3.1 Upper bounds

In this section, we first prove an upper bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$, then present our testing algorithm as well as the sample and time complexity of our algorithm. For brevity, let $b_1 \triangleq \sum_{|S| \leq k} |bias_D(S)|$ and $b_2 \triangleq \sqrt{\sum_{|S| \leq k} bias_D(S)^2}$. Note that $b_2 \leq b_1 \leq \sqrt{M_{n,k}} b_2 < n^{k/2} b_2$.

The only previously known upper bound for $\Delta(D, \mathcal{D}_{\text{kwi}})$ is given in [4], where it is implicitly shown that $\Delta(D, \mathcal{D}_{\text{kwi}}) \leq b_1$. Our new bound is the following.

Theorem 3.1 (Upper bound on distance)**.** *$\Delta(D, \mathcal{D}_{kwi}) \leq O\left((\log n)^{k/2} \sqrt{\sum_{|S| \leq k} bias_D(S)^2}\right)$. Consequently, $\Delta(D, \mathcal{D}_{kwi}) \leq O\left((n \log n)^{k/2}\right) \max_{|S| \leq k} |bias_D(S)|$.*

Since b_2 is always smaller than or equal to b_1, our upper bound is no weaker than that of [4] up to a polylogarithmic factor. However, for many distributions of interest, b_2 is much smaller than b_1 (e.g., when all the biases are roughly of the same magnitude, as in the case of random uniform distributions, then $b_2 = O^*(b_1/n^{k/2})$).

The basic ideas of our proof are the following. We first operate in the Fourier space to construct a "pseudo-distribution" D_1 by forcing all the first k-level Fourier coefficients to be zero. D_1 is not a distribution because it may assume negative values at some points. We then correct all these negative points by a series of convex combinations of D_1 with k-wise independent distributions. In this way we maintain that all the first k-level Fourier coefficients are still zero; on the other hand, we increase the weights at negative points so that they now assume non-negative values. During the correction, we distinguish between two kinds of points which have negative weights: Light points whose magnitudes are small and heavy points whose magnitudes are large. We use two different types of k-wise independent distributions to handle these two kinds of points. Using Bonami-Beckner's inequality, we show that only a small number of points are heavy, thus obtaining a better bound for $\Delta(D, \mathcal{D}_{\text{kwi}})$.

Proof of Theorem 3.1. The following lemma bounds the ℓ_1-distance between a function and its convex combination with other distributions.

Lemma 3.2. *Let f be a real function defined on a domain $\mathcal{D}$ such that $\sum_{x \in \mathcal{D}} f(x) = 1$. Let $D_1, \ldots, D_\ell$ be distributions over the same domain $\mathcal{D}$. Suppose there exist positive real numbers $w_1, \ldots, w_\ell$ such that $D' \triangleq \frac{1}{1+\sum_{i=1}^{\ell} w_i}(f + \sum_{i=1}^{\ell} w_i D_i)$ is non-negative for all $x \in \mathcal{D}$. Then $\|f(x) - D'(x)\|_1 \leq 2^{-n+1} \sum_{i=1}^{\ell} w_i$.*

Proof. $\|f(x) - D'(x)\|_1 = \|\sum_{i=i}^{\ell} w_i(D' - D_i)\|_1 \leq \sum_{i=i}^{\ell} w_i \|D' - D_i\|_1 \leq 2^{-n+1} \sum_{i=i}^{\ell} w_i$. □

We first construct a real function $D_1 : \{0,1\}^n \to \mathbb{R}$ based on D but forcing all its first k-level biases to be zero. D_1 is defined by explicitly specifying all of its Fourier coefficients:

$$\hat{D}_1(S) = \begin{cases} 0, & \text{if } S \neq \emptyset \text{ and } |S| \leq k \\ \hat{D}(S), & \text{otherwise.} \end{cases}$$

Since $\hat{D}_1(\emptyset) = \hat{D}(\emptyset) = \frac{1}{2^n}$, we have $\sum_x D_1(x) = 1$. Note that in general D_1 is not a distribution because it is possible that for some x, $D_1(x) < 0$. By Parseval's equality, $\|D - D_1\|_2 = \frac{1}{2^n}\sqrt{\sum_{|T|\leq k} bias_D(T)^2} = \frac{1}{2^n} b_2$. Hence by the Cauchy-Schwartz inequality, we can upper bound the ℓ_1-norm of $D - D_1$ as $\|D - D_1\|_1 \leq 2^{-n} \cdot b_2$. Now we define another function $D_2 : \{0,1\}^n \to \mathbb{R}$ as

$$\hat{D}_2(S) = \begin{cases} \hat{D}(S), & \text{if } S \neq \emptyset \text{ and } |S| \leq k \\ 0, & \text{otherwise.} \end{cases}$$

By the linearity of the Fourier transform, $D_1(x) + D_2(x) = D(x)$. Since $D(x) \geq 0$ for all $x \in \{0,1\}^n$, we have $D_1(x) \geq -D_2(x)$. By the Fourier transform, $|D_2(x)| = \left|\frac{1}{2^n}\sum_{1\leq|S|\leq k} bias_D(S)\chi_S(x)\right| \leq \frac{1}{2^n}\sum_{1\leq|S|\leq k} |bias_D(S)| = \frac{1}{2^n} b_1$. Hence the magnitudes of $D_1(x)$'s negative points are upper bounded by $\frac{1}{2^n} b_1$, i.e. $D_2(x) \geq -\frac{1}{2^n} b_1$.

By the linearity of the Fourier transform, if we define a function D' as the convex combination of D_1 with some k-wise independent distributions so that D' is non-negative, then D' will be a k-wise independent distribution, since all the Fourier coefficients of D' on the first k levels are zero.

If we use a uniform distribution to correct all the negative weights of D_1, then we will get an upper bound almost the same (up to a factor of 3/2) as that of [4]. To improve on this, we distinguish between two kinds of points where D_1 may assume negative weights: heavy points and light points. Let $\lambda = (2\sqrt{\log n})^k$. We call a point x *heavy* if $D_2(x) \geq \lambda b_2/2^n$, and *light* if $0 < D_2(x) < \lambda b_2/2^n$. For light points, we still use a uniform distribution to correct them; but for *each* heavy point, say z, we will use a special k-wise independent distribution $U_{\text{BCH-z}}(x)$, constructed by [2]:

Theorem 3.3. *([2]) For any $z \in \{0,1\}^n$, there is a k-wise independent distribution $U_{BCH\text{-}z}(x)$ over $\{0,1\}^n$ such that $U_{BCH\text{-}z}(z) = \frac{1}{|\operatorname{Supp}(U_{BCH\text{-}z})|} = \Omega(n^{-\lfloor k/2\rfloor})$.* [2]

Thus, we define D' by

$$D'(x) = \frac{D_1(x) + \lambda b_2 U_n(x) + \sum_{z \text{ is heavy}} w U_{\text{BCH-z}}(x)}{1 + \lambda b_2 + \sum_{z \text{ is heavy}} w}.$$

We select $w = \frac{|\operatorname{Supp}(U_{\text{BCH-z}})|}{2^n} b_1$. Since $D_1(x) \geq -\frac{b_1}{2^n}$, one can check that $D'(x)$ is non-negative for both heavy and light points. Hence D' is a k-wise independent distribution.

Next we bound the number of heavy points. Note that $D_2(x)$ has only the first k-level Fourier coefficients, hence we can use Bonami-Beckner's inequality to bound the probability of $|D_2(x)|$ assuming large values, and thus the total number of heavy points.

First we scale $D_2(x)$ to make it of unit ℓ_2-norm. Define $f(x) = \frac{2^n}{b_2} D_2(x)$. Then $\|f\|_2 = \frac{2^n}{b_2}\|D_2\|_2 = \frac{2^n}{b_2}\sqrt{\frac{1}{2^n}\sum_{x\in\{0,1\}^n} D_2(x)^2} = \frac{2^n}{b_2}\sqrt{\frac{1}{2^{2n}}\sum_{1\leq|S|\leq k} bias_D(S)^2} = 1$, where the second to last step follows from Parseval's equality. Now using the higher moment inequality method, we have, for even p,

$$\Pr[|f(x)| \geq \lambda] \leq \frac{\mathbf{E}_x\left[|f(x)|^p\right]}{\lambda^p} = \frac{\|f\|_p^p}{\lambda^p}.$$

[2] Note that, as shown by [13, 2], the support sizes of such constructions are essentially optimal.

By Lemma 2.2, $\|f\|_p \le \left(\sqrt{p-1}\right)^k \|f\|_2 = \left(\sqrt{p-1}\right)^k$. Plug in $\lambda = (2\sqrt{\log n})^k$ and $p = \log n$, and w.l.o.g. assume that p is even, then we have

$$\Pr[|f(x)| \ge 2^k \log^{k/2} n] \le \frac{(p-1)^{pk/2}}{\lambda^p} < \frac{p^{pk/2}}{\left(2\sqrt{\log n}\right)^{pk}} = (\frac{1}{2})^{k\log n} = \frac{1}{n^k}.$$

Therefore,

$$\Pr\left[D_1(x) \le -2^k(\log n)^{k/2}\frac{b_2}{2^n}\right] \le \Pr\left[D_2(x) \ge 2^k(\log n)^{k/2}b_2/2^n\right] \le \Pr\left[|D_2(x)| \ge 2^k(\log n)^{k/2}b_2/2^n\right]$$
$$= \Pr\left[|f(x)| \ge 2^k(\log n)^{k/2}\right] < 1/n^k.$$

In other words, there are at most $2^n/n^k$ heavy points. By Lemma 3.2 we get (recall that $|\operatorname{Supp}(U_{\text{BCH-}z})| = O\left(n^{\lfloor k/2\rfloor}\right)$ and $b_1 \le n^{k/2}b_2$)

$$\frac{1}{2}|D'-D_1|_1 \le \lambda b_2 + \sum_{z \text{ heavy}} w(z) \le (2\sqrt{\log n})^k b_2 + \frac{2^n}{n^k}\frac{|\operatorname{Supp}(U_{\text{BCH-}z})|}{2^n} b_1 = (2\sqrt{\log n})^k b_2 + O(b_2) = O\left((\log n)^{k/2}b_2\right).$$

Finally, by the triangle inequality, $\Delta(D,D') = \frac{2^n}{2}\|D-D'\|_1 \le \frac{2^n}{2}(\|D-D_1\|_1 + \|D_1-D'\|_1) = O\left((\log n)^{k/2}b_2\right)$.

□

Armed with Theorem 3.1, we are ready to describe our algorithm for testing k-wise independence. The algorithm is simple in nature: it estimates all the first k-level biases of the distribution and returns "Yes" if they are all small. Let C_k be the hidden constant in $O(\cdot)$ in the second part of Theorem 3.1.

```
Algorithm Test-KWI-Closeness($D$,$k$,$\delta$)
From $D$, draw a set $Q$ of samples of size $|Q| = O\left(k\log n/\delta'^2\right)$, where $\delta' = \frac{\delta}{3C_k(n\log n)^{k/2}}$.
For each non-empty subset $S \subseteq [n], |S| \le k$, use $Q$ to estimate $bias_D(S)$ to within an additive term of $\delta'$.
If $\max_S |bias_D(S)| \le 2\delta'$ return "Yes"; else return "No".
```

The analysis of `Test-KWI-Closeness` establishes the following theorem (full proof appears in Appendix A.1).

Theorem 3.4 (Testing k-wise independence upper bounds)**.** *Testing k-wise independence can be solved using $O(k(\log n)^{k+1}n^k/\delta^2) = O^*(\frac{n^k}{\delta^2})$ samples from the distribution and in time $O^*(\frac{n^{2k}}{\delta^2})$.*

3.2 Lower bounds

In this section, we give a lower bound on the sample complexity of our testing algorithm. However, we first motivate our study from the perspective of real functions defined over the boolean cube.

The upper bound given in Theorem 3.1 naturally raises the following question: Can we give a lower bound on $\Delta(D,\mathcal{D}_{\text{kwi}})$ in term of the first k-level biases of D? The only known answer to this question we are aware of is the folklore lower bound in Fact 2.4: $\Delta(D,\mathcal{D}_{\text{kwi}}) \ge \frac{1}{2}\max_{1\le|S|\le k}|bias_D(S)|$. This bound is too weak for many distributions, as demonstrated in [4], who gave a family of distributions that have all the first k-level biases at most $O\left(\frac{1}{n^{1/5}}\right)$, but are at least 1/2-away from any k-wise independent distribution. Their proof is based on a min-entropy argument, which seems to work only for distributions with small support size.

In fact, this statistical distance lower bound problem can be put into a more general framework. Given a function $f : \{0,1\}^n \to \mathbb{R}$, can we give a lower bound on $\|f\|_1$ if only the first k-level Fourier coefficients of f are known? Hausdorff-Young's inequality gives $\|f\|_1 \ge \|\hat{f}\|_\infty$, which leads directly to the bound we just discussed (Fact 2.4). We develop a new approach to lower bound $||f||_1$ in terms of f's first k-level Fourier coefficients (details appear in Appendix A.2). Our method works for general k and is based on convolving f with an auxiliary function and then applying Young's convolution inequality. Applying our lower bound result to $\Delta(D, \mathcal{D}_{\mathrm{kwi}})$, we get:

Theorem 3.5 (Lower bound on distance)**.** *Given a distribution D over $\{0,1\}^n$, define a family of functions $\mathcal{D}_g \subseteq \mathbb{R}^{\{0,1\}^n}$ such that for all $g \in \mathcal{D}_g$, the Fourier coefficients of g satisfy:*

$$\hat{g}(S) = \begin{cases} 0, & \text{if } S = \emptyset \text{ or } |S| > k \\ sign(bias_D(S)) & \text{if } |S| \leq k \text{ and } bias_D(S) \neq 0 \\ \pm 1, & \text{if } |S| \leq k \text{ and } bias_D(S) = 0, \end{cases}$$

where $sign(x) = 1$ if $x > 0$, $sign(x) = -1$ if $x < 0$ and $sign(x) = 0$ otherwise. Then for all $g \in \mathcal{D}_g$, $\Delta(D, \mathcal{D}_{kwi}) \geq \frac{\frac{1}{2}\sum_{|S|\leq k}|bias_D(S)|}{||g||_\infty}$.

Under this framework, we prove the following lower bound on distances between random uniform distributions and k-wise independence, which is the basis of our sample lower bound result, Theorem 3.7 (The proof is deferred to Appendix A.3). Note that by Theorem 3.1, this bound is almost tight (see Proposition A.16 for details).

Lemma 3.6 (Random Distribution Lemma)**.** *Let $k > 2$. Let $Q = \frac{M_{n,k}}{n\delta^2} < 2^{n^{1/3}}$, with $\delta \leq 1$. If we sample uniformly at random Q strings from $\{0,1\}^n$ to form a random multiset $\mathcal{Q}$ and let $U_{\mathcal{Q}}(x)$ be the uniform distribution over $\mathcal{Q}$, then for all large enough n,* $\Pr_{\mathcal{Q}}[\Delta(U_{\mathcal{Q}}, \mathcal{D}_{kwi}) > 0.228\delta] = 1 - o(1)$.

Theorem 3.7 (Sample lower bound)**.** *For $k > 2$ and $\delta \leq 0.228$, Testing k-wise independence requires at least $|Q| = \Omega\left(\frac{1}{\delta} \cdot (\frac{n}{k})^{\frac{k-1}{2}}\right)$ samples from the distribution.*

Our lower bound result rules out the possibility of polynomial time testing algorithms for $k = \omega(1)$. To give an idea of how Theorem 3.7 follows from Lemma 3.6, note that U_n is k-wise independent, and by Lemma 3.6, $U_{\mathcal{Q}}$ is far from k-wise independent. But any algorithm making $o(\sqrt{Q})$ will not see any collisions and thus will fail to distinguish between these two distributions.

4 Testing (ϵ, k)-wise independence

In this section, we study the sample and time complexity of distinguishing whether a distribution is (ϵ, k)-wise independent or is at distance at least δ from any (ϵ, k)-wise independent distribution (as defined in 1.1). We call this testing problem Test(ϵ, k)-independence *to within distance δ* (we drop the reference to δ whenever it is clear from the context). On one hand, compared to testing k-wise independence, we prove that exponentially fewer samples suffice for Test(ϵ, k)-independence. On the other hand, this exponential improvement does not carry over to the time complexity; we show that it is unlikely that there is a poly(n) time algorithm for Test(ϵ, k)-independence.

We begin by describing our sample complexity results: while testing k-wise independence requires $\Omega(n^{\frac{k-1}{2}})$ samples, we show that $O\left(\frac{k \lg n}{\epsilon^2\delta^2}\right)$ samples suffice for testing (ϵ, k)-wise independence. In particular, the sample complexity of Test(ϵ, k)-independence is only poly($n/\epsilon\delta$), even for the case when $k = \omega(1)$ and $\epsilon, \delta = n^{-O(1)}$. Specifically, we show that:

Theorem 4.1 (Sample upper bound)**.** *For any* $0 < \epsilon, \delta < 1$, Test(ϵ, k)-independence *to within distance δ can be solved using $|Q| = O\left(\frac{k \log n}{\epsilon^2\delta^2}\right)$ samples from the distribution D.*

Theorem 4.2 (Sample lower bound)**.** *For $\epsilon > \frac{1}{n^{k/4}}$, $0 < \delta < \frac{1}{2} - \epsilon$, any tester solving* Test(ϵ, k)-independence *to within distance δ requires at least $|Q| = \Omega\left(\frac{\sqrt{k \log n}}{(\epsilon+\delta)\sqrt{\log 1/(\epsilon+\delta)}}\right)$ samples from the distribution.*

To prove the sample upper bound Theorem 4.1, we prove a relationship between Test(ϵ, k)-independence and the problem of distinguishing an (ϵ, k)-wise independent distribution from one that is not even (ϵ', k)-wise independent, for some $\epsilon' > \epsilon > 0$ (see definition 4.5 for a formal statement). For the latter problem, we simply compute the minimum $\tilde{\epsilon}$ such that D is $(\tilde{\epsilon}, k)$-wise independent, and compare $\tilde{\epsilon}$ to ϵ and ϵ'.

To obtain the lower bound, we study the minimum support of a distribution D which is (ϵ, k)-wise independent, and show it is $\Omega\left(\frac{k \log n}{\epsilon^2 \log(1/\epsilon)}\right)$; the rest of the proof is similar to the proof of Theorem 3.7 (full proofs appear in Appendix B).

We now turn to the time complexity result. In contrast to the positive result for sample complexity, we show that the time complexity *cannot be* poly$(n/\epsilon\delta)$ for $k = \Theta(\log n)$, under the following conjecture regarding the hardness of finding a hidden clique in a random graph. In the following, let $t = t(n)$ be a nondecreasing function of n so that $t(n) > \lg^3 n$ (the bigger $t(n)$, the stronger the conjecture and our result).

Conjecture 4.3 (HC-Find$[t]$)**.** *For $n > 0$, let G be a random graph on n vertices generated by the following process, $\mathcal{G}_{n,1/2,t}$: connect each pair of vertices with probability $1/2$, then choose a random set of t vertices, and interconnect these vertices to form a clique (called the* hidden clique*). Then there is no randomized* poly(n) *time algorithm that, for all n, given G, outputs a clique of size t, with success probability at least $1 - 1/n$.*

We discuss this conjecture in more detail in Section 4.1. Assuming the conjecture, we prove the following theorem on time complexity of Test(ϵ, k)-independence.

Theorem 4.4 (Time lower bound)**.** *Assume conjecture* HC-Find$[t(n)]$ *holds for some $t(n) \geq \lg^3 n$. Let $k = \alpha \lg n$ for a constant $\alpha \leq 1$, $\epsilon = \frac{2\alpha \lg^2 n}{n^\alpha}$, and $\delta = \frac{t(n^\alpha/6)}{2n^\alpha}$. Then there is no* poly$(n)$ *time algorithm that solves* Test(ϵ, k)-independence *to within distance δ, even given access to any polynomial number of samples from the distribution.*

The proof of the theorem appears in Section 4.2. Note that for the above settings, Test(ϵ, k)-independence can be solved in $n^{O(k)} = 2^{O(\log^2 n)}$ time, and thus it is not a priori clear whether one can prove such hardness result under a more standard assumption, such as $\mathbf{P} \neq \mathbf{NP}$.

To prove our results on the sample and time complexity of Test(ϵ, k)-independence, we study a closely related problem. Specifically, we consider the problem of distinguishing between a distribution D that is (ϵ, k)-wise independent and a distribution that is not even (ϵ', k)-wise independent for $\epsilon' > \epsilon > 0$. It is somewhat easier to obtain upper and lower bounds for the latter problem from which we can deduce the bounds on the original Test(ϵ, k)-independence problem. We define the new problem below and describe its relation to Test(ϵ, k)-independence; the proof of the relation is deferred to Appendix B.1. As mentioned in the preliminaries, $\mathcal{D}_{(\epsilon,k)}$ denotes the set of all (ϵ, k)-wise independent distributions.

Definition 4.5 (Test(ϵ, k)-vs-(ϵ', k)-independence)**.** *Let $0 < \epsilon < \epsilon' < 1$, and D be a distribution over $\{0,1\}^n$. We call a* tester for Test(ϵ, k)-vs-(ϵ', k)-independence *an algorithm that, given a set $Q \subset \{0,1\}^n$ drawn i.i.d. from D, outputs: 1) "Yes", if $D \in \mathcal{D}_{(\epsilon,k)}$; and 2) "No", if $D \notin \mathcal{D}_{(\epsilon',k)}$. The tester may fail with probability at most* $1/3$.

Lemma 4.6. *Let $0 < \epsilon, \delta < 1$. If there exists a tester for* Test(ϵ, k)-vs-$(\epsilon + \epsilon\delta, k)$-independence *using $Q = Q(n, k, \epsilon, \delta)$ samples and $T = T(n, k, \epsilon, \delta)$ time, then there exists a tester for* Test(ϵ, k)-independence *to within distance δ using Q samples and T time.*

Conversely, if there exists a tester for Test(ϵ, k)-independence *to within distance δ using Q samples and T time, then there exists a tester for* Test(ϵ, k)-vs-$(\epsilon + \delta, k)$-independence *using Q samples and T time.*

In the rest of the section, we discuss the plausibility of the hidden clique conjecture 4.3, and present the proof of the Theorem 4.4 based on the conjecture.

4.1 The Hidden Clique Conjecture

The problem of finding a hidden clique in a random graph has been open since the works of [18, 23]. For $t = o(\sqrt{n})$, there is no known polynomial time algorithm that finds even a $(1+\epsilon)\log_2 n$ clique, for any constant $\epsilon > 0$. When $t \geq \Omega(\sqrt{n})$, [5] and [15] exhibit polynomial time algorithms that do find the hidden clique of size t.

Conjecture 4.3 is a generalization of the conjecture of the hardness of the problem of finding a $(1+\epsilon)\log_2 n$ clique in a random graph from $\mathcal{G}_{n,1/2} = \mathcal{G}_{n,1/2,0}$ (i.e., *without* inserting any hidden clique) [20, 18]. This problem is a long-standing open question raised by [22] (see also the survey of [16] and the references therein). Although a random graph $\mathcal{G}_{n,1/2}$ has a clique of size $(2-o(1))\log_2 n$ with high probability [6], there is no known polynomial time algorithm that finds even a clique of size $(1+\epsilon)\log_2 n$ for constant $\epsilon > 0$ (a simple greedy algorithm finds a $(1-\epsilon)\log_2 n$ clique, w.h.p.). The failure to exhibit such a polynomial time algorithm led to the conjecture that there is no algorithm able to find a $(1+\epsilon)\log_2 n$ clique in polynomial time [18, 20]. Furthermore, the problem of finding a clique of size $\frac{3}{2}\log_2 n$ in a random graph has been proposed as a "hard problem" for cryptographic purposes [20].

4.2 Time complexity lower bound: proof of Theorem 4.4

Below we show that if conjecture HC-Find$[t]$ holds, then the running time of any tester for Test(ϵ, k)-vs-(ϵ', k)-independence is super-polynomial in n for $k = \alpha \lg n$, for any constant $\alpha \leq 1$, and $\epsilon = \frac{2\alpha \lg^2 n}{n^\alpha} = n^{-O(1)}$, $\epsilon' = \frac{t(n^\alpha/6)}{n^\alpha} = n^{-O(1)}$. The theorem then follows by applying Lemma 4.6.

To prove the theorem, we first prove that the conjecture HC-Find$[t]$ implies the following conjecture on the hardness of *deciding* whether a hidden clique is present or not in a random graph. The conjecture is also parametrized by the minimum size of the hidden clique, $t = t(n)$, a non-decreasing function of n.

Conjecture 4.7 (HC-Decide$[t]$)**.** *For $n > 0$, let G be a random graph on n vertices that is generated via either $\mathcal{G}_{n,1/2}$ or $\mathcal{G}_{n,1/2,t'}$, where $t' \geq t(n)$ may be chosen adversarialy. Then there is no polynomial time algorithm that for any n, given G, can output whether G came from $\mathcal{G}_{n,1/2}$ or $\mathcal{G}_{n,1/2,t'}$, with success probability at least $1 - 1/n^3$.*

We show in Appendix B.4 that if HC-Find$[t]$ holds, then HC-Decide$[t/3]$ also holds.

Now, to prove the theorem, it is sufficient to give a reduction from the problem of distinguishing between $\mathcal{G}_{m,1/2}$ and $\mathcal{G}_{m,1/2,t'}$ to the problem Test(ϵ, k)-vs-(ϵ', k)-independence, where $t' \geq t$, $m = 2^{k-1} = n^{\Omega(1)}$, $\epsilon = \frac{2\alpha \lg^2 n}{n^\alpha}$, $\epsilon' = \frac{t(n^\alpha/6)}{n^\alpha}$. Let $\mathcal{T}$ be a tester that decides whether $D \in \mathcal{D}_{\epsilon,k}$ or $D \notin \mathcal{D}_{\epsilon',k}$ with error probability $\leq n^{-4}$ (we can amplify the success probability by running the tester $\mathcal{T}$ for $O(\log n)$ times, each with a new set of samples Q).

Suppose we are given a graph G on $m = 2^{k-1}$ vertices, generated either via $\mathcal{G}_{m,1/2}$ or $\mathcal{G}_{m,1/2,t'(m)}$. Let A be the adjacency matrix of G with the diagonal entries set randomly to 0 or 1. From the matrix $A \in M_{m,m}$, we construct a new matrix $B \in M_{m,n}$ by appending $n - m$ columns to the right, where each new entry is randomly chosen from $\{0, 1\}$. We view matrix B as describing a distribution $D_B : \{0,1\}^n \to [0,1]$ defined to be uniform on the set of the m rows of B: $D_B(x) = \frac{|\{i | B_i = x\}|}{m}$, where B_i is the i^{th} row of B.

We claim that, with high probability, if $G \in \mathcal{G}_{m,1/2}$, then $D_B \in \mathcal{D}_{(\epsilon,k)}$, and, conversely, if $G \in \mathcal{G}_{m,1/2,t'}$, then $D_B \notin \mathcal{D}_{(\epsilon',k)}$. These properties immediately imply the reduction to the tester for Test(ϵ, k)-vs-(ϵ', k)-independence: generate the sample set Q by drawing samples according the distribution D_B and feed it to the tester. If the tester returns "Yes" (i.e., $D_B \in \mathcal{D}_{(\epsilon,k)}$), return $G \in \mathcal{G}_{m,1/2}$. Otherwise (i.e., $D_B \notin \mathcal{D}_{(\epsilon',k)}$), return $G \in \mathcal{G}_{m,1/2,t(m)}$.

Next we prove that if $G \in \mathcal{G}_{m,1/2}$ then w.h.p. $D_B \in \mathcal{D}_{(\epsilon,k)}$, and if $G \in \mathcal{G}_{m,1/2,t(m)}$ then $D_B \notin \mathcal{D}_{(\epsilon',k)}$. To simplify the argument, for a matrix B, we define a parameter $g_k(B)$ that roughly corresponds to the minimum $\tilde{\epsilon}$ such that D_B is $(\tilde{\epsilon}, k)$-wise independent:

Definition 4.8. *Let k be such that* $1 \leq k \leq n$*. For a matrix* $B \in M_{m,n}(\{0,1\})$ *and* $\overrightarrow{v} \in \{0,1\}^k$*, we define a* $(k, \overrightarrow{v})$-repetition *to be a set of distinct columns* $C = \{i_1, i_2, \ldots, i_k\}$ *and a set of distinct rows R, such that* $R = \{r \in [m] \mid B_{ri_1} B_{ri_2} \ldots B_{ri_k} = \overrightarrow{v}\}$*. We define* $g_k(B)$ *to be the maximum value of* $|R|/m$*, over all* $(k, \overrightarrow{v})$*-repetitions for all choices of* $\overrightarrow{v} \in \{0,1\}^k$*.*

Note that when $g_k(B) \geq 2 \cdot 2^{-k}$*, the minimum* $\tilde{\epsilon}$ *for which* $D_B \in \mathcal{D}_{(\tilde{\epsilon},k)}$ *is* $\tilde{\epsilon} = g_k(B) - 2^{-k}$*.*

Now, on one hand, if $G \in \mathcal{G}_{m,1/2}$, then B is a random 0/1 matrix, and by an easy union bound calculation, $g_k(B) \leq \frac{k \lg n}{(k - \lg m)m}$ with probability at least $1 - O\left((2e/k)^k\right) \geq 1 - n^{-4}$. Thus, since $g_k(B) \geq 1/m = 2 \cdot 2^{-k}$, we conclude that $D_B \in \mathcal{D}_{(\epsilon,k)}$, where $\epsilon \leq \frac{k \lg n}{(k - \lg m)m} - 2^{-k} \leq \frac{2\alpha \lg^2 n}{n^\alpha}$. This is the only part where the reduction can fail.

On the other hand, if $G \in \mathcal{G}_{m,1/2,t'}$, then B contains a clique of size $t' \geq t(m)$ and thus a $(k, 1^k)$-repetition with $|R| \geq \frac{t(m)-1}{2}$, implying that $g_k(B) \geq \frac{t(m)-1}{2m}$. Thus $D_B \notin \mathcal{D}_{(\epsilon',k)}$, where $\epsilon' = \frac{t(m)-1}{2m} - 2^{-k} = \frac{t(n^\alpha/2)-2}{n^\alpha} \geq \frac{t(n^\alpha/2)}{n^\alpha}$.

The total error probability is at most n^{-4} from the tester, plus n^{-4} from the above reduction. This finishes the proof of the Theorem 4.4.

Acknowledgements. We would like to thank Piotr Indyk for introducing us to the hidden clique problem, and Kenneth Regan for helpful discussions and careful reading of an early draft of this paper.

References

[1] N. Alon. Problems and results in extremal combinatorics (Part I). *Discrete Math.*, 273:31–53, 2003.

[2] N. Alon, L. Babai, and A. Itai. A fast and simple randomized algorithm for the maximal independent set problem. *J. of Algorithms*, 7:567–583, 1986.

[3] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992. Earlier version in FOCS'90.

[4] N. Alon, O. Goldreich, and Y. Mansour. Almost k-wise independence versus k-wise independence. *Inform. Process. Lett.*, 88:107–110, 2003.

[5] N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures and Algorithms*, 13:457–466, 1998.

[6] N. Alon and J. Spencer. *The Probabilistic Method.* John Wiley and Sons, second edition, 2000.

[7] T. Batu, E. Fisher, L. Fortnow, R. Kumar, R. Rubinfeld, and P. White. Testing random variables for independence and identity. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 442–451, 2001.

[8] T. Batu, L. Fortnow, R. Rubinfeld, W. D. Smith, and P. White. Testing that distributions are close. In *Proc. 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 189–197, 2000.

[9] T. Batu, R. Kumar, and R. Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proc. 36th Annual ACM Symposium on the Theory of Computing*, pages 381–390, New York, NY, USA, 2004. ACM Press.

[10] W. Beckner. Inequalities in fourier analysis. *Annals of Mathematics*, 102:159–182, 1975.

[11] S. Bernstein. *The Theory of Probabilities*. Gostehizdat Publishing House, Moscow, 1946.

[12] A. Bonami. Etude des coefficients fourier des fonctiones de $l^p(g)$. *Ann. Inst. Fourier (Grenoble)*, 20(2):335–402, 1970.

[13] B. Chor, J. Friedman, O. Goldreich, J. Håstad, S. Rudich, and R. Smolensky. The bit extraction problem and t-resilient functions. In *Proc. 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 396–407, 1985.

[14] I. Dinur, E. Friedgut, G. Kindler, and R. O'Donnell. On the Fourier tails of bounded functions over the discrete cube. In *Proc. 38th Annual ACM Symposium on the Theory of Computing*, pages 437–446, New York, NY, USA, 2006. ACM Press.

[15] U. Feige and R. Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures and Algorithms*, 16:195–208, 2000.

[16] A. Frieze and C. McDiarmid. Algorithmic theory of random graphs. *Random Structures and Algorithms*, 10(1-2):5–42, 1997.

[17] O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. Technical Report TR00-020, Electronic Colloquium on Computational Complexity, 2000.

[18] M. Jerrum. Large cliques elude the metropolis process. *Random Structures and Algorithms*, 3(4):347–359, 1992.

[19] A. Joffe. On a set of almost deterministic k-independent random variables. *Annals of Probability*, 2:161–162, 1974.

[20] A. Juels and M. Peinado. Hiding cliques for cryptographic security. *Des. Codes Cryptography*, 20(3):269–280, 2000.

[21] R. Karp and A. Wigderson. A fast parallel algorithm for the maximal independent set problem. *Journal of the ACM*, 32(4):762–773, 1985.

[22] R. M. Karp. The probabilistic analysis of some combinatorial search algorithms. In J. F. Traub, editor, *Algorithms and Complexity: New directions and Recent Results*, pages 1–19, New York, 1976. Academic Press.

[23] L. Kučera. Expected complexity of graph partitioning problems. *Disc. Applied Math.*, 57(2-3):193–212, 1995.

[24] M. Luby. A simple parallel algorithm for the maximal independent set problem. *SIAM J. on Comput.*, 15(4):1036–1053, 1986. Earlier version in STOC'85.

[25] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. on Comput.*, 22(4):838–856, 1993. Earlier version in STOC'90.

[26] R. Rubinfeld and R. A. Servedio. Testing monotone high-dimensional distributions. In *Proc. 37th Annual ACM Symposium on the Theory of Computing*, pages 147–156, New York, NY, USA, 2005. ACM Press.

A Testing k-wise independent distributions

In this section we provide the omitted details from Section 3.

A.1 Testing algorithm and its analysis

In this section we will present our testing algorithm in details. Recall that C_k denotes the hidden constant in Theorem 3.1, i.e. $\Delta(D, \mathcal{D}_{\text{kwi}}) \le C_k (n \log n)^{k/2} \max_{S \neq \emptyset, |S| \le k} |bias_D(S)|$.

We will use the following algorithm to estimate the bias of a distribution D over any non-empty subset S with error parameter δ.

```
**Algorithm** Estimate-Bias($D$,$S$,$k$,$\delta$)
Set $m = O\left((k \log n)/\delta^2\right)$.
Set $n_{odd} = 0$.
(Assume the sample set is $Q = \{X_1, \ldots, X_m\}$)
For $i = 1$ to $m$
    If $\oplus_{j \in S} X^i_j = 1$
        $n_{odd} = n_{odd} + 1$.
Output $bias_D(S) = \frac{2n_{odd}}{m} - 1$.
```

Lemma A.1. *Let $bias_D(S)$ be the bias computed by* Estimate-Bias(D,S,k,δ)*, and $\overline{bias}_D(S)$ be the expected value of $bias_D(S)$ (i.e., the bias of distribution D over S). Then with probability at least $1 - \frac{1}{3n^k}$, $|bias_D(S) - \overline{bias}_D(S)| \le \delta$.*

Proof. Let n_{odd} and n_{even} be the number of strings of odd parity and even parity, respectively, over S. Without loss of generality, assume that $\overline{bias}_D(S) \ge 0$ (otherwise replace n_{odd} with n_{even} in the following argument). Define the indicator random variables χ_i for $i = 1, \ldots, m$, such that $\chi_i = \oplus_{j \in S} X^i_j$. It is clear that χ_i are $0/1$ random variables and $\mathbf{E}[\chi_i] = n_{odd}/m \ge 1/2$. Now applying Chernoff bound to χ_i gives the desired result, since $\overline{bias}_D(S) = 2\mathbf{E}[\chi_i] - 1$. □

Now we are ready to describe the algorithm of testing closeness to k-wise independence, which (implicitly) uses Estimate-Bias as a subroutine.

```
**Algorithm** Test-KWI-Closeness($D$,$k$,$\delta$)
Set $\delta' = \frac{\delta}{3C_k(n \log n)^{k/2}}$.
Set $m = O\left(k \log n/\delta'^2\right)$.
For each non-empty subset $S \subseteq [n]$ of size at most $k$
    Set $n_{odd}(S) = 0$.
(Assume the sample set is $Q = \{X_1, \ldots, X_m\}$)
For $i = 1$ to $m$
    For each non-empty subset $S \subseteq [n]$ of size at most $k$
        If $\oplus_{j \in S} X^i_j = 1$
            $n_{odd}(S) = n_{odd}(S) + 1$.
For each non-empty subset $S \subseteq [n]$ of size at most $k$
    $bias_D(S) = \frac{2n_{odd}(S)}{m} - 1$.
Set $\Delta = C_k(n \log n)^{k/2} \max_S |bias_D(S)|$.
If $\Delta \le \frac{2}{3}\delta$.
  **accept**;
Else
  **reject**.
```

Next we prove the correctness of Test-KWI-Closeness(D,k,δ).

Theorem 3.4. *Let D be a distribution over $\{0,1\}^n$. If $\Delta(D, \mathcal{D}_{kwi}) \le \frac{2\delta}{3C_k(n\log n)^{k/2}}$, then* `Test-KWI-Closeness` *accepts with probability at least $2/3$; If $\Delta(D, \mathcal{D}_{kwi}) > \delta$, then* `Test-KWI-Closeness` *accepts with probability at most $1/3$. Furthermore, the sample complexity of* `Test-KWI-Closeness` *is $O(kC_k(\log n)^{k+1}n^k/\delta^2) = O^*(\frac{n^k}{\delta^2})$, and running time of* `Test-KWI-Closeness` *is $O^*(\frac{n^{2k}}{\delta^2})$.*

Proof of Theorem 3.4. The running time and sample complexity analysis is straightforward. If $\Delta(D, \mathcal{D}_{\text{kwi}}) \le \frac{2\delta}{3C_k(n\log n)^{k/2}}$, then by Fact 2.4, $\overline{bias}_D(S) \le \frac{\delta}{3C_k(n\log n)^{k/2}}$ for every $1 \le |S| \le k$. By Lemma A.1, $|bias_D(S) - \overline{bias}_D(S)| \le \frac{\delta}{3C_k(n\log n)^{k/2}}$ with probability at least $1 - \frac{1}{3n^k}$. Thus union bound gives, with probability at least $1 - M_{n,k}\frac{1}{3n^k} \ge 2/3$ (since $M_{n,k} \le n^k$), $|bias_D(S) - \overline{bias}_D(S)| \le \frac{\delta}{3C_k(n\log n)^{k/2}}$ holds for each S. This implies that, for every non-empty S of size at most k, $C_k(n\log n)^{k/2}|bias_D(S)| \le \frac{2}{3}\delta$. Therefore, the algorithm accepts.

If $\Delta(D, \mathcal{D}_{\text{kwi}}) > \delta$, by Theorem 3.1, $C_k(n\log n)^{k/2} \max_{S\neq\emptyset,|S|\le k} |\overline{bias}_D(S)| > \delta$. A similar analysis shows that with probability at least $2/3$, $C_k(n\log n)^{k/2} \max_{S\neq\emptyset,|S|\le k} |bias_D(S)| > \frac{2}{3}\delta$ and hence the algorithm rejects.

□

Note that for constant k, `Test-KWI-Closeness` gives an algorithm testing k-wise independence running in time sublinear (in fact, polylogarithmic) in the size of the support ($N = 2^n$) of the distribution.

A.2 New lower bounds for $\Delta(D, \mathcal{D}_{\text{kwi}})$

In this section, we will develop a new framework to prove lower bound on the distance between a distribution and k-wise independent distributions and apply this method to prove Theorem 3.5. In fact, our techniques developed here may be of independent interest: We give a new lower bound on the ℓ_1-norm of a function $f : \{0,1\}^n \to \mathbb{R}$ in terms of f's first k-level Fourier coefficients. Our method is based on convolving f with an auxiliary function and applying Young's convolution inequality:

Theorem A.2 (Young's convolution inequality)**.** *Let $1 \le p, q, r \le \infty$, such that $\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1$. Then for any $f, g : \{0,1\}^n \to \mathbb{R}$, $\|f * g\|_r \le \|f\|_p\|g\|_q$.*

Given a distribution D over $\{0,1\}^n$. Let D' be the k-wise independent distribution which is closest to D, i.e., $\Delta(D, \mathcal{D}_{\text{kwi}}) = \Delta(D, D') = \frac{1}{2}\|D - D'\|_1$. Define $f(x) = D(x) - D'(x)$. Then we have

$$\hat{f}(S) = \frac{1}{2^n} bias_D(S), \quad \text{for all non-empty subsets } S \text{ with } |S| \le k,$$

and

$$\Delta(D, \mathcal{D}_{\text{kwi}}) = \frac{1}{2}\sum_{x\in\{0,1\}^n} |f(x)| = 2^{n-1}\|f\|_1.$$

We will try to get a lower bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ by bounding the ℓ_1-norm of $f(x)$ from below.

Theorem A.3. *Let $f : \{0,1\}^n \to \mathbb{R}$. Define a family of functions $\mathcal{F}_g \subseteq \mathbb{R}^{\{0,1\}^n}$ such that for all $g \in \mathcal{F}_g$, the Fourier coefficients of g satisfy*

$$\hat{g}(S) = \begin{cases} 0, & \text{if } S = \emptyset \text{ or } |S| > k \\ sign(\hat{f}(S)) & \text{if } |S| \le k \text{ and } \hat{f}(S) \neq 0 \\ \pm 1, & \text{if } |S| \le k \text{ and } \hat{f}(S) = 0. \end{cases}$$

Then for all $g \in \mathcal{F}_g$,

$$\|f\|_1 \ge \frac{\sum_{|S|\le k} |\hat{f}(S)|}{\|g\|_\infty}.$$

In particular,

$$||f||_1 \geq \frac{\sum_{|S|\leq k} |\hat{f}(S)|}{\min_{g\in\mathcal{F}_g} ||g||_\infty}.$$

Note that for all S such that $\hat{f}(S) = 0$, we have the freedom of choosing either $+1$ or -1 to minimize $||g||_\infty$ and get better lower bound.

Proof. Setting $p = 1$, then Young's convolution inequality (Theorem A.2) gives, for any $1 \leq r \leq \infty$, and any $f, g : \{0,1\}^n \to \mathbb{R}$,

$$||f||_1 \geq \frac{||f * g||_r}{||g||_r}.$$

Now we define function g as in the Theorem and define $h(x) \triangleq (f * g)(x)$. Then by the convolution theorem,

$$\hat{h}(S) = \begin{cases} |\hat{f}(S)|, & \text{if } S \text{ is non-empty and } |S| \leq k \\ 0, & \text{otherwise.} \end{cases}$$

By the definition of the Fourier transform,

$$|h(x)| = |\sum_S \hat{h}(S)\chi_S(x)| = \left|\sum_{|S|\leq k} |\hat{f}(S)|\chi_S(x)\right| \leq \sum_{|S|\leq k} |\hat{f}(S)| = h(0),$$

since for all $S \subseteq [n]$, $\chi_S(0) = 1$ and the evaluation of any function at 0 is simply the sum of all its Fourier coefficients. Thus, $||h||_\infty = h(0) = \sum_{|S|\leq k} |\hat{f}(S)|$. Now take r tending to infinity, we get

$$||f||_1 \geq \frac{\sum_{|S|\leq k} |\hat{f}(S)|}{||g||_\infty}.$$

□

Thus we get a lower bound for $\Delta(D, \mathcal{D}_{\text{kwi}})$:

Theorem 3.5. *Let D be a distribution over $\{0,1\}^n$, and let $\mathcal{F}_g$ be defined as in Theorem A.3 but replacing $\hat{f}(S)$ with $bias_D(S)$. Then for all $g \in \mathcal{F}_g$, $\Delta(D, \mathcal{D}_{kwi}) \geq \frac{\frac{1}{2}\sum_{|S|\leq k} |bias_D(S)|}{||g||_\infty}$.*

If all the low level Fourier coefficients of f are non-zero, then there is a unique $g \in \mathcal{F}_g$ that corresponds to f. Otherwise, there may be many g's in $\mathcal{F}_g$ all correspond to f. If this is the case, for the purpose of proving lower bound, we may pick the one with the smallest infinity norm. On the other hand, there are many different f's that correspond to the same g. A nice property of function g is that only the first k-level Fourier coefficients are non-zero and all these coefficients are in $\{-1, 1\}$. By the monotonicity of norms and Parseval's equality, we have $||g||_\infty \geq ||g||_2 = \sqrt{\sum_{1 \leq |S|\leq k} 1} = \sqrt{M_{n,k}}$. And a trivial upper bound is $||g||_\infty \leq M_{n,k}$. Note that if $||g||_\infty \ll M_{n,k}$, then our new lower bound on $\Delta(D, \mathcal{D}_{\text{kwi}})$ probably will give a much better bound than the trivial lower bound $\Delta(D, \mathcal{D}_{\text{kwi}}) \geq \frac{1}{2}\max_S |bias_D(S)|$. Next we will provide some evidence showing the strength of our new lower bound: among $2^{M_{n,k}} = 2^{O(n^k)}$ possible g's, at most an exponentially small portion of them may have $||g||_\infty = \Omega(\sqrt{nM_{n,k}})$. Thus most g's will give good lower bound.

Theorem A.4. *Let $\vec{g}$ be an $M_{n,k}$-dimensional vector with its $M_{n,k}$ components being $g(x)$'s non-zero Fourier coefficients, then for all $c > 0$ and for all sufficiently large n,*

$$\Pr_{\vec{g}\in_R\{-1,1\}^{M_{n,k}}}\left[||g||_\infty > 1.18\sqrt{c+1}\sqrt{nM_{n,k}}\right] < 2^{-cn}.$$

Proof. We will need the following simple Chernoff-type tail bound (see Corollary A.1.2 of [6])

Lemma A.5. *Let x_i, $1 \leq i \leq m$, be mutually independent random variables with* $\Pr[x_i = 1] = \Pr[x_i = -1] = \frac{1}{2}$ *and set $S_m = x_1 + \cdots + x_m$. Let $a > 0$. Then*

$$\Pr[|S_m| > a] < 2e^{-\frac{a^2}{2m}}.$$

Let x be an arbitrary element in $\{0, 1\}^n$. Then

$$g(x) = \sum_{i=1}^{M_{n,k}} \hat{g}(S_i)\chi_{S_i}(x) = \sum_{i=1}^{M_{n,k}} Y_i,$$

where we define $Y_i = \hat{g}(S_i)\chi_{S_i}(x)$. Now if $\hat{g}(S_i)$'s are independent random variables uniformly distributed in $\{-1, 1\}^{M_{n,k}}$, so are Y_i's. Hence we can apply Lemma A.5 to bound the probability of $|g(x)|$ assuming large values. Set $a = 1.18\sqrt{(c+1)M_{n,k}n} > \sqrt{\frac{2.005}{\log_2 e}M_{n,k}(cn+n)}$, then $a > \sqrt{\frac{2}{\log_2 e}M_{n,k}(cn+n+1)}$ and $a^2 > \frac{2}{\log_2 e}M_{n,k}(cn+n+1)$ for all sufficiently large n. Now Lemma A.5 gives

$$\Pr_{\vec{g}}[|g(x)| > a] = \Pr\left[\left|\sum\nolimits_{i=1}^{M_{n,k}} Y_i\right| > a\right] < 2e^{-\frac{a^2}{2M_{n,k}}} < 2^{-cn} \cdot 2^{-n}$$

Applying the union bound argument to all 2^n strings gives

$$\begin{aligned}\Pr_{\vec{g}}[||g||_\infty > a] &= \Pr_{\vec{g}}[\exists x \in \{0,1\}^n \text{ s.t. } |g(x)| > a] \\ &< 2^{-cn}.\Box\end{aligned}$$

A.3 Proof of the Random Distribution Lemma

For completeness, we restate the Lemma here.

Lemma 3.6 (Random Distribution Lemma)**.** *Assume that $k > 2$. Let $Q = \frac{M_{n,k}}{n\delta^2} < 2^{n^{1/3}}$, with $\delta \leq 1$. If we sample uniformly at random Q strings from $\{0, 1\}^n$ to form a random multiset $\mathcal{Q}$ and let $U_Q(x)$ be the uniform distribution over Q, then for all large enough n,* $\Pr_Q[\Delta(U_Q, \mathcal{D}_{kwi}) > 0.228\delta] = 1 - o(1)$.

Proof. We will follow the lower bound techniques developed in the previous section to prove this lemma. However, for ease of analysis, we will use functions different from those used in the previously. Let $D'(x)$ be the k-wise independent distribution with minimum statistical distance to U_Q. Define

$$f_Q(x) = U_Q(x) - D'(x).$$

Then we have

$$\hat{f}_Q(S) = \hat{U}_Q(S), \quad \text{for all } S \subseteq [n],\ S \neq \emptyset \text{ and } |S| \leq k,$$

and

$$\Delta(U_Q, \mathcal{D}_{\text{kwi}}) = 2^{n-1}||f_Q||_1.$$

Define $g_Q(x) : \{0, 1\}^n \to \mathbb{R}$ as

$$\hat{g}_Q(S) = \begin{cases} \hat{f}_Q(S), & \text{if } S \neq \emptyset \text{ and } |S| \leq k, \\ 0, & \text{otherwise.} \end{cases}$$

Also define the convolution $h_Q(x) \triangleq (f_Q * g_Q)(x)$, then

$$\hat{h}_Q(S) = \begin{cases} \hat{f}_Q(S)^2, & \text{if } S \neq \emptyset \text{ and } |S| \leq k, \\ 0, & \text{otherwise.} \end{cases}$$

by the convolution theorem. Applying Young's inequality gives

$$\| f_Q\|_1 \geq \frac{\|h_Q\|_\infty}{\|g_Q\|_\infty}.$$

We will prove the Lemma 3.6 by proving the following two lemmas bounding $\|h_Q\|_\infty$ and $\|g_Q\|_\infty$, respectively.

Lemma A.6. *For all large enough n,* $\Pr_Q\left[\|h_Q\|_\infty \geq 0.999\frac{M_{n,k}}{2^{2n}Q}\right] = 1 - o(1)$.

Lemma A.7. *Let $\frac{M_{n,k}}{n} \leq Q < 2^{n^{1/3}}$. Then for all $k > 2$ and large enough n,* $\Pr_Q\left[\|g_Q\|_\infty \leq \frac{2.19}{2^n}\sqrt{\frac{nM_{n,k}}{Q}}\right] = 1 - o(1)$.

Now we prove the Lemma assuming Lemma A.6 and Lemma A.7: By the union bound, with probability $1 - o(1)$, both the lower bound of $\|h_Q\|_\infty$ and the upper bound of $\|g_Q\|_\infty$ hold. Then we have

$$\Delta(U_Q, \mathcal{D}_{\text{kwi}}) = \frac{1}{2}2^n\| f_Q\|_1 \geq \frac{1}{2} \cdot \frac{0.999\frac{M_{n,k}}{Q}}{2.19\sqrt{\frac{M_{n,k}n}{Q}}} > 0.228\sqrt{\frac{M_{n,k}}{nQ}},$$

as desired. □

In the following proofs of Lemma A.6 and Lemma A.7, we will assume that all the elements in multiset Q are distinct. This will not affect our results, since by the Birthday paradox, the probability of seeing a collision in Q is $o(1)$.

A.3.1 Proof of Lemma A.6

We prove the lower bound of $\|h_Q\|_\infty$ by computing the expectation and variance of $\|h_Q\|_\infty$. Then a simple application of Chebyshev's inequality gives the desired bound. The calculations are straightforward but rather tedious.

Proof of Lemma A.6. By the definition of Fourier transform

$$|h_Q(x)| = \left|\sum_{1\leq|S|\leq k} \hat{h}_Q(S)\chi_S(x)\right| \leq \sum_{1\leq|S|\leq k} \left|\hat{h}_Q(S)\right| = \sum_{1\leq|S|\leq k} \hat{h}_Q(S) = h_Q(0).$$

Therefore

$$\|h_Q\|_\infty = h_Q(0) = \sum_{1\leq|S|\leq k} \hat{f}_Q(S)^2.$$

Then for all non-empty subset S with $|S| \leq k$,

$$\begin{aligned}\hat{f}_Q(S) &= \frac{1}{2^n}\sum_{x\in\{0,1\}^n} U_Q(x)\chi_S(x) \\ &= \frac{1}{2^nQ}\sum_{x\in Q}\chi_S(x);\end{aligned}$$

and

$$\begin{aligned}\hat{f_Q}(S)^2 &= \frac{1}{2^{2n}} \sum_{x,y\in\{0,1\}^n} U_Q(x)\chi_S(x)U_Q(y)\chi_S(y)\\ &= \frac{1}{2^{2n}Q^2} \sum_{x,y\in Q} \chi_S(x)\chi_S(y);\end{aligned}$$

To facilitate the calculation of the expectation and variance of $||h_Q||_\infty$, we first state two simple technical claims.

Claim A.8. *Let x and y be two* distinct *strings chosen uniformly at random from* $\{0,1\}^n$*, then for all* $n > 1$*,* $x + y$ *is equal to every element in* $\{0,1\}^n \setminus \{0^n\}$ *with equal probability.*

Proof. First we fix an x, then the map $y \to x + y$ is a one-to-one correspondence between $\{0,1\}^n \setminus \{x\}$ and $\{0,1\}^n \setminus \{0^n\}$. Then notice that y equals every element in $\{0,1\}^n \setminus \{x\}$ with equal probability. □

Claim A.9. *Let x, y, x' and y' be four* distinct *strings chosen uniformly at random from* $\{0,1\}^n$*. Then for all* $n > 2$*,* $x + y + x' + y'$ *is equal to every element in* $\{0,1\}^n$ *with equal probability.*

Proof. Let $z_1 = x+y$. By claim A.8, z_1 equals all strings in $\{0,1\}^n \setminus \{0^n\}$ with equal probability. Then $z_1 + x'$ equals all strings in $\{0,1\}^n \setminus \{x'\}$ with equal probability. But x' takes all values in $\{0,1\}^n$ equally often, so is $z_1 + x' = x + y + x'$. Therefore $x + y + x' + y'$ is uniformly distributed over $\{0,1\}^n$. □

Proposition A.10.

$$\mathbf{E}_Q[||h_Q||_\infty] = \frac{M_{n,k}}{2^{2n}Q}(1 - \frac{Q-1}{2^n - 1}).$$

Proof.

$$\begin{aligned}\mathbf{E}_Q[||h_Q||_\infty] &= \mathbf{E}_Q\left[\sum_{1\le|S|\le k} \hat{f_Q}(S)^2\right]\\ &= \frac{1}{2^{2n}Q^2}\mathbf{E}_Q\left[\sum_{1\le|S|\le k}\sum_{x,y\in Q}\chi_S(x)\chi_S(y)\right]\\ &= \frac{M_{n,k}}{2^{2n}Q} + \frac{1}{2^{2n}Q^2}\mathbf{E}_Q\left[\sum_{1\le|S|\le k}\sum_{x,y\in Q,x\neq y}\chi_S(x)\chi_S(y)\right]\\ &= \frac{M_{n,k}}{2^{2n}Q} + \frac{1}{2^{2n}Q^2}\mathbf{E}_Q\left[\sum_{1\le|S|\le k}\sum_{x\in Q}\sum_{z\neq 0^n, z-x\in Q}\chi_S(z)\right]\\ &= \frac{M_{n,k}}{2^{2n}Q} + \frac{M_{n,k}Q(Q-1)}{2^{2n}Q^2}\mathbf{E}_{z\neq\{0^n\}}[\chi_S(z)].\end{aligned}$$

By Claim A.8, z is uniformly distributed over $\{0,1\}^n \setminus \{0^n\}$. Since for any $S \neq \emptyset$, $\sum_{z\in\{0,1\}^n}\chi_S(z) = 0$, hence $\sum_{z\in\{0,1\}^n\setminus\{0^n\}}\chi_S(z) = -1$, and $\mathbf{E}_{z\in\{0,1\}^n\setminus\{0^n\}}[\chi_S(z)] = -\frac{1}{2^n-1}$. Then we have

$$\mathbf{E}_Q[||h_Q||_\infty] = \frac{M_{n,k}}{2^{2n}Q}(1 - \frac{Q-1}{2^n - 1}).$$

This completes the proof. □

Proposition A.11.

$$\mathbf{E}_Q\left[\|h_Q\|_\infty^2\right] = \frac{M_{n,k}^2}{2^{4n}Q^2}(1-\frac{Q-1}{2^n-1})^2 + \frac{2M_{n,k}Q(Q-1)}{2^{4n}Q^4}(1-\frac{2(Q-2)}{2^n-1}) - \frac{M_{n,k}(Q-1)^2}{2^{4n}(2^n-1)^2Q^2}$$
$$= \frac{M_{n,k}^2}{2^{4n}Q^2}(1-\frac{Q-1}{2^n-1})^2 + \frac{2M_{n,k}}{2^{4n}Q^2}(1-o(1)).$$

Proof.

$$\begin{aligned}\mathbf{E}_Q\left[\|h_Q\|_\infty^2\right] &= \mathbf{E}_Q\left[(\sum_{1\le|S|\le k}\hat{f}_Q(S)^2)^2\right]\\ &= \mathbf{E}_Q\left[\sum_{1\le|S|\le k}\sum_{1\le|T|\le k}\hat{f}_Q(S)^2\hat{f}_Q(T)^2\right]\\ &= \frac{1}{2^{4n}Q^4}\mathbf{E}_Q\left[\sum_{1\le|S|\le k}\sum_{1\le|T|\le k}\sum_{x,y\in Q}\sum_{x',y'\in Q}\chi_S(x+y)\chi_T(x'+y')\right].\end{aligned}$$

Then one can distinguish between 12 different cases and calculate their expectations respectively. We omit the details here. □

Therefore we have

$$\mathrm{Var}(\|h_Q\|_\infty) = \frac{1}{2^{4n}}\frac{2M_{n,k}}{Q^2}(1-o(1)),$$

and

$$\sigma(\|h_Q\|_\infty) = \frac{1}{2^{2n}}\frac{\sqrt{2M_{n,k}}}{Q}(1-o(1)).$$

Finally we apply Chebyshev's inequality, which states that for any $t>0$ $\Pr[|X-\mathbf{E}[X]|>t\sigma(X)]<\frac{1}{t^2}$, to $\|h_Q\|_\infty$ to finish the proof of Lemma A.6. □

A.3.2 Proof of Lemma A.7

The proof of Lemma A.7 is more involved: A simple calculation shows that $g_Q(x)$ equals a summation of Q independent random variables $Y_1,\ldots,Y_Q$ determined by the random subset $\mathcal{Q}$, where $-M_{n,k}\le Y_i\le M_{n,k}$. However, a direct application of Hoeffding's bound to the sum can only gives $\|g_Q\|_\infty = O(M_{n,k})$, thus $\Delta(U_Q,\mathcal{D}_{\mathrm{kwi}}) = \Omega(\frac{1}{Q})$, which is too weak. We improve on this is by noticing that the variance of Y_i is small, thus Bernstein's inequality [11] gives a better bound. This approach gives us the desired result but also imposes a restriction that $\delta = O(1/n)$. We overcome this difficulty by the observation that for most of the random variables, $|Y_i|$ is much smaller than $M_{n,k}$, as implied by Bonami-Beckner's inequality. This enables us to distinguish between two kinds of Y_i's: Those $|Y_i|$ are small and those $|Y_i|$ are large, and sum them separately. Followings are the details.

Proof of Lemma A.7. Fix an arbitrary $x \in \{0,1\}^n$. Then

$$\begin{aligned} g_Q(x) &= \sum_{1\le|S|\le k} \hat{f}_Q(S)\chi_S(x) \\ &= \frac{1}{2^n}\sum_{1\le|S|\le k}\sum_{y\in\{0,1\}^n} U_Q(y)\chi_S(x)\chi_S(y) \\ &= \frac{1}{2^n Q}\sum_{1\le|S|\le k}\sum_{y\in Q}\chi_S(x+y) \\ &= \frac{1}{2^n Q}\sum_{y\in Q}\sum_{1\le|S|\le k}\chi_S(x+y) \\ &= \frac{1}{2^n Q}\sum_{y\in Q} Y_x(y), \end{aligned}$$

where $Y_x(y) \triangleq \sum_{1\le|S|\le k}\chi_S(x+y)$. Note that the summation is over *independent* random variables $Y_x(y)$ in Q.

We will distinguish between two kinds of strings: We call a string y is x*-bad* if $|Y_x(y)| \ge M_{n,k}/n^{1.1}$ and x*-good* otherwise. Then we can do the summation over strings that are x-good and strings that are x-bad separately:

$$\begin{aligned} 2^n Q|g_Q(x)| &= \left|\sum_{y\in Q,\, y \text{ is } x\text{-good}} Y_x(y) + \sum_{y\in Q,\, y \text{ is } x\text{-bad}} Y_x(y)\right| \\ &\le \left|\sum_{y\in Q,\, y \text{ is } x\text{-good}} Y_x(y)\right| + \left|\sum_{y\in Q,\, y \text{ is } x\text{-bad}} Y_x(y)\right|. \end{aligned}$$

Next we define a set $\mathcal{B} = \{w \in \{0,1\}^n : |\sum_{1\le|S|\le k}\chi_S(w)| \ge \frac{M_{n,k}}{n^{1.1}}\}$. This definition gives us the following simple characterization of strings those are x-bad by observing that $\mathcal{B}$ is just the set of all strings that are 0^n-bad.

Claim A.12. *A string y is x-bad if and only if $x+y \in \mathcal{B}$, therefore y is* bad *for exactly $|\mathcal{B}|$ many of strings.*

Proof. The first part follow directly from the definitions of x-bad and set $\mathcal{B}$. For the second part, note that for each element w in $\mathcal{B}$, y is bad for $y+w$. □

Our next claim shows that in fact only an exponentially small portion of all the strings in $\{0,1\}^n$ are in $\mathcal{B}$, and hence each string y is bad for only an exponentially small portion of all strings in $\{0,1\}^n$.

Claim A.13. *Let $k > 2$ be a constant natural number. Then for all large enough n,*

$$|\mathcal{B}| < 2^n/2^{n^{1-\frac{2.3}{k}}}.$$

Proof. Consider a function $F(x) = \sum_{1\le|S|\le k}\chi_S(x)$. By Parseval's equality, $\|F\|_2 = \sqrt{\sum_{1\le|S|\le k} 1} = \sqrt{M_{n,k}}$. Since F has only the first k levels Fourier coefficients, Bonami-Beckner's inequality applies. Hence we have, for all even number $p > 2$

$$\begin{aligned}
\Pr_x[|F(x)| \geq \frac{M_{n,k}}{n^{1.1}}] &= \Pr_x[|F(x)| \geq \frac{\sqrt{M_{n,k}}}{n^{1.1}} \|F\|_2] \\
&\leq \frac{\mathbf{E}_x\left[|F(x)|^p\right]}{(\frac{\sqrt{M_{n,k}}}{n^{1.1}} \|F\|_2)^p} \\
&= \frac{\|F\|_p^p}{(\frac{\sqrt{M_{n,k}}}{n^{1.1}} \|F\|_2)^p} \\
&\leq \frac{(p-1)^{pk/2}}{(\frac{\sqrt{M_{n,k}}}{n^{1.1}})^p} \\
&< \left(\frac{p^{k/2}}{\frac{\sqrt{M_{n,k}}}{n^{1.1}}} \right)^p .
\end{aligned}$$

Now take $p = (\frac{\sqrt{M_{n,k}}}{2n^{1.1}})^{2/k}$, and w.l.o.g. assume that p is even, we have $\Pr_x[|F(x)| \geq \frac{M_{n,k}}{n^{1.1}}] < (\frac{1}{2})^{n^{1-\frac{2.3}{k}}}$, for all sufficiently large n. This completes the proof. □

Since each string y is only bad for a small number of strings, if we choose uniformly at random $Q < 2^{n/3}$ strings to form a set $\mathcal{Q}$, then almost surely there is no string x in $\{0,1\}^n$ witnessing more than one string in $\mathcal{Q}$ that is x-bad.

Claim A.14. *Let $\mathcal{Q}$ be a random subset constructed by uniformly at random choosing $Q < 2^{n/3}$ distinct elements from $\{0,1\}^n$. Then with probability $1 - o(1)$, for each $x \in \{0,1\}^n$, there is at most one string in $\mathcal{Q}$ which is x-bad.*

Proof. We will bound the probability that there is an x which has at least two strings in $\mathcal{Q}$ that are x-bad. Fix an arbitrary $x \in \{0,1\}^n$. The probability that there are two strings $y_1, y_2 \in \mathcal{Q}$ which are x-bad is $\binom{Q}{2}\frac{|\mathcal{B}|}{2^n}\frac{|\mathcal{B}|-1}{2^n-1} < \frac{Q^2|\mathcal{B}|^2}{2^{2n}} = o(\frac{1}{2^n})$. We finish the proof by applying a union bound argument over all $x \in \{0,1\}^n$. □

If we apply the Hoeffding bound directly to the sum of strings that are x-good, we would not get the desired result. Instead, we will employ the following Bernstein's inequality [11], which gives a better bound on the sum of independent random variables when we have a good bound on the variance of the random variables being summed.

Theorem A.15 (Bernstein's inequality)**.** *Let $X_1, \ldots, X_Q$ be independent real-valued random variables such that $|X_i| \leq C$ for all $1 \leq i \leq Q$. Let $\sigma^2 = \frac{1}{Q}\sum_{i=1}^{Q} \mathrm{Var}(X_i)$. Then for any $t > 0$*

$$\Pr[|\sum_{i=1}^{Q} X_i - \mathbf{E}\left[X\right]| > Qt] \leq e^{-\frac{Qt^2}{2\sigma^2 + \frac{2Ct}{3}}}$$

We will first compute the expectation and variance of $Y_x(y)$ over the universe (namely $\{0,1\}^n$). Then due to the fact that the number of x-bad strings is exponentially smaller than the cardinality of the universe, the expectation as well as the variance of the set of x-good strings are almost identical to those of the universe.

Indeed, by direct calculation

$$\mathbf{E}_y\left[Y_x(y)\right] = \mathbf{E}_y\left[\sum_{1\le|S|\le k}\chi_S(y)\right] = \sum_{1\le|S|\le k}\mathbf{E}_y\left[\chi_S(y)\right] = \sum_{1\le|S|\le k} 0 = 0,$$

and

$$\begin{aligned}\mathbf{E}_y\left[Y_x(y)^2\right] &= \mathbf{E}_y\left[\left(\sum_{1\le|S|\le k}\chi_S(y)\right)^2\right]\\ &= \mathbf{E}_y\left[\sum_{1\le|S|,|T|\le k}\chi_S(y)\chi_T(y)\right]\\ &= \mathbf{E}_y\left[\sum_{1\le|S|\le k}\chi_S(y)^2\right] + \mathbf{E}_y\left[\sum_{1\le|S|\le k}\sum_{S'\neq\emptyset}\chi_{S'}(y)\right] \quad (S' \triangleq S\Delta T)\\ &= M_{n,k} + 0\\ &= M_{n,k}.\end{aligned}$$

Since for all $x, y \in \{0,1\}^n$, $0 \le |Y_x(y)| \le M_{n,k}$, we have

$$|\mathbf{E}_{y\text{ is }x\text{-good}}\left[Y_x(y)\right]| \le \frac{M_{n,k}2^n}{2^{n^{1-\frac{2.3}{k}}}(2^n - 2^{n^{1-\frac{2.3}{k}}})} = o(1),$$

and

$$\mathrm{Var}_{y\text{ is }x\text{-good}}(Y_x(y)) \le \frac{M_{n,k}2^n}{2^n - 2^{n^{1-\frac{2.3}{k}}}} - o(1)^2 = (1+o(1))M_{n,k}.$$

Now we are ready to put it all together. With probability $1 - o(1)$, for all $x \in \{0,1\}^n$ there is at most one x-bad string in Q. We will call such random sets *good*. Conditioned on this, the contribution of sum over x-bad strings in (A.3.2) is at most $M_{n,k}$. We then apply Bernstein's inequality to the sum over x-good strings with $\sigma^2 = M_{n,k}(1+o(1)), C = \frac{M_{n,k}}{n^{1.1}}$. By setting $t = \sqrt{2.02\ln 2\frac{M_{n,k}n}{Q}} < 1.19\sqrt{\frac{M_{n,k}n}{Q}}$, and note that for all $Q \ge M_{n,k}/n$, $Ct = o(\sigma^2)$, we have

$$\Pr_{Q\text{ is good}}\left[\left|\sum_{y\in Q,\, y\text{ is }x\text{-good}} Y_x(y)\right| \ge Qt\right] \le 2^{-n}2^{-0.01n} = o(\frac{1}{2^n}).$$

The union bound over all $x \in \{0,1\}^n$ implies that, with probability $1 - o(1)$, for all x

$$2^n|g_Q(x)| \le \frac{1.19\sqrt{M_{n,k}nQ} + M_{n,k}}{Q} \le 2.19\sqrt{\frac{M_{n,k}n}{Q}}.$$

i.e. with probability $1 - o(1)$,

$$\|g_Q\|_\infty \le \frac{2.19}{2^n}\sqrt{\frac{M_{n,k}n}{Q}}.$$

This completes the proof of Lemma A.7. □

A.3.3 Tightness of the Lemma 3.6

Our lower bound on the statistical distance between a random distribution and k-wise independent distributions is almost tight due to the following proposition

Proposition A.16. *Let S be a random multiset formed by uniformly sampling $\Omega(k(\log n)^{k+1}n^k/\delta^2)$ times from $\{0,1\}^n$. Then with high probability, U_S is δ-close to k-wise independent.*

Proof. By Chernoff bound, for every $S \subseteq [n], |S| \leq k, S \neq \emptyset$, with probability at least $(1-\frac{1}{3n^k})$, $|bias_{U_S}(S)| \leq O(\delta/(n\log n)^{k/2})$. By a union bound argument, this holds for all S with probability at least $2/3$. Applying Theorem 3.1 gives the desired result. □

A.4 Sample lower bound

For completeness, we give a detailed proof of Theorem 3.7.

Proof of Theorem 3.7. We will show that if the algorithm makes too few queries, then it cannot successfully distinguish between two distributions far apart with high probability. Consider the following two distributions. The first one is the uniform distribution U_n over $\{0,1\}^n$. Obviously, U_n is k-wise independent for all $1 \leq k \leq n$. The second distribution U_Q is a uniform distribution over a multiset Q, where Q is constructed by uniformly and randomly sampling $Z = \left(\frac{0.228}{\delta}(\frac{n}{k})^{\frac{k-1}{2}}\right)^2 \leq 0.228^2\frac{M_{n,k}}{n\delta^2}$ times from $\{0,1\}^n$. By Lemma 3.6, with probability $1-o(1)$, U_Q is at least δ-far from any k-wise independent distribution. Now let $\mathcal{A}$ be any algorithm that makes $Q = o(\sqrt{Z}) = o\left(\frac{1}{\delta}(\frac{n}{k})^{\frac{k-1}{2}}\right)$ queries. Let D_{U_n} and D_{U_Q} be distributions over sample sets of size Q that algorithm $\mathcal{A}$ obtains from U_n and U_Q respectively. By the Birthday Paradox, with probability $1-o(1)$, all the strings queried from U_n are distinct and all the strings queried from U_Q are distinct. Conditioned on this, the statistical distance between D_{U_n} and D_{U_Q} is zero, since both of the distributions are uniform distributions over m distinct strings randomly selected from $\{0,1\}^n$. Therefore, $\mathcal{A}$ cannot distinguish these two distributions with success probability bounded away from 1/2 by a constant. By the union bound, the total probability that $\mathcal{A}$ succeeds is at most $\frac{1}{2}+o(1)$. This completes the proof. □

B Testing (ϵ, k)-wise independence

In this section we provide the omitted details from Section 4.

B.1 Relation of Test(ϵ, k)-independence and Test(ϵ, k)-vs-(ϵ', k)-independence

Here we prove the Lemma 4.6. We break down the lemma into two propositions and prove each separately.

Proposition B.1. *Let $0 < \epsilon, \delta < 1$. If $\Delta(D, \mathcal{D}_{(\epsilon,k)}) > \delta$, then $D \notin \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$. Hence any algorithm for solving* Test(ϵ, k)-vs-($\epsilon+\epsilon\delta, k$)-independence *can be used to solve* Test(ϵ, k)-independence *to within distance δ with the same sample and time complexity.*

Proof. We prove the contrapositive: that if $D \in \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$, then $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \leq \delta$. Suppose D is $(\epsilon+\epsilon\delta, k)$-wise independent. Then construct a new distribution D' that is (ϵ, k)-wise independent and such that $\Delta(D, D') \leq \delta$ as follows:

$$D' = \begin{cases} D, & \text{with probability } 1-\delta \\ U_n, & \text{with probability } \delta \end{cases}$$

Clearly, $\Delta(D, D') \leq \delta$. Also, for any k indices $i_1 < i_2 < \ldots i_k$, and any vector $\overrightarrow{v} \in \{0,1\}^k$, we have that

$$\left|\Pr_{x \leftarrow D'}\left[x_{i_1}x_{i_2}\ldots x_{i_k} = \overrightarrow{v}\right] - 2^{-k}\right| \leq (\epsilon+\epsilon\delta)(1-\delta) + 0\cdot\delta \leq \epsilon$$

Now, to solve the problem $\textsc{Test}(\epsilon,k)$-INDEPENDENCE of distribution D, simply invoke $\textsc{Test}(\epsilon,k)$-VS-$(\epsilon+\epsilon\delta,k)$-INDEPENDENCE on D. If D is such that $\Delta(D,\mathcal{D}_{(\epsilon,k)}) > \delta$, then $D \notin \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$ and the tester for the latter problem will report "No". Otherwise, if $D \in \mathcal{D}_{(\epsilon,k)}$, then the tester for the latter problem with report "Yes". □

Next we show how to reduce solving $\textsc{Test}(\epsilon,k)$-VS-$(\epsilon',k)$-INDEPENDENCE to solving $\textsc{Test}(\epsilon,k)$-INDEPENDENCE to within distance $\delta = \epsilon' - \epsilon$.

Proposition B.2. *Let $0 \le \epsilon < \epsilon' < 1$. If $\Delta(D,\mathcal{D}_{(\epsilon,k)}) \le \epsilon' - \epsilon$ then $D \in \mathcal{D}_{(\epsilon',k)}$. Hence, any algorithm for solving* $\textsc{Test}(\epsilon,k)$-INDEPENDENCE *to within distance $\delta = \epsilon' - \epsilon$ also solves* $\textsc{Test}(\epsilon,k)$-VS-$(\epsilon',k)$-INDEPENDENCE *(with the same sample and time complexity).*

Proof. Let $D' \in \mathcal{D}_{(\epsilon,k)}$ be such that $\Delta(D,D') \le \delta = \epsilon' - \epsilon$. Then, for any indices $i_1 < i_2 < \ldots i_k$, and any vector $\overrightarrow{v} \in \{0,1\}^k$, we have that

$$\Pr_{x\leftarrow D}\left[x_{i_1}x_{i_2}\ldots x_{i_k} = \overrightarrow{v}\right] - 2^{-k} \le \left(\Pr_{x\leftarrow D'}\left[x_{i_1}x_{i_2}\ldots x_{i_k} = \overrightarrow{v}\right] + \delta\right) - 2^{-k} \le \epsilon + \delta = \epsilon'$$

and

$$\Pr_{x\leftarrow D}\left[x_{i_1}x_{i_2}\ldots x_{i_k} = \overrightarrow{v}\right] - 2^{-k} \ge \left(\Pr_{x\leftarrow D'}\left[x_{i_1}x_{i_2}\ldots x_{i_k} = \overrightarrow{v}\right] - \delta\right) - 2^{-k} \ge -\epsilon - \delta = -\epsilon'$$

Thus, $D \in \mathcal{D}_{(\epsilon',k)}$.

To conclude, we solve $\textsc{Test}(\epsilon,k)$-VS-$(\epsilon',k)$-INDEPENDENCE on D by a simple invocation to $\textsc{Test}(\epsilon,k)$-INDEPENDENCE on D. If D is such that $D \notin \mathcal{D}_{(\epsilon',k)}$ then $\Delta(D,\mathcal{D}_{(\epsilon,k)}) > \epsilon' - \epsilon$ and the tester for the latter problem with report "No". Otherwise, if $D \in \mathcal{D}_{(\epsilon,k)}$, then the tester returns "Yes". □

B.2 Sample complexity upper bound: proof of Theorem 4.1

We give an algorithm for $\textsc{Test}(\epsilon,k)$-VS-$(\epsilon',k)$-INDEPENDENCE, and use the relation of Lemma 4.6 to derive the upper bound for $\textsc{Test}(\epsilon,k)$-INDEPENDENCE. In our algorithm for $\textsc{Test}(\epsilon,k)$-VS-$(\epsilon',k)$-INDEPENDENCE, we do *not* use biases. Note that using biases in the natural way would introduce an approximation error of $2^{\Omega(k)}$ (see [3] for relations between the parameter ϵ and the biases).

Theorem B.3 (Sample upper bound)**.** *Let $0 \le \epsilon < \epsilon' < 1$.* $\textsc{Test}(\epsilon,k)$-VS-$(\epsilon',k)$-INDEPENDENCE *can be solved using $Q = O\left(\frac{k\log n}{(\epsilon'-\epsilon)^2}\right)$ samples from the distribution.*

Proof. The algorithm proceeds in a straight-forward way: first, using the samples Q, compute a distribution $\tilde{D}$ that is an approximation to D, and then check whether $\tilde{D}$ is closer to being (ϵ,k)-wise independent, or is closer to not even being (ϵ',k)-wise independent. Specifically, given the multiset of queries Q, construct a distribution $\tilde{D} : \{0,1\}^d \to [0,1]$ that is uniformly distributed on Q, i.e., $\tilde{D}(x) = \frac{|\{i\in[|Q|] \mid q_i=x\}|}{|Q|}$, where $Q = \{q_1,\ldots q_{|Q|}\}$. Then we can compute the minimum $\tilde{\epsilon}$ such that $\tilde{D}$ is $(\tilde{\epsilon},k)$-wise independent. If $\tilde{\epsilon} \le \frac{\epsilon+\epsilon'}{2}$, then we declare D is (ϵ,k)-wise independent, and, if $\tilde{\epsilon} > \frac{\epsilon+\epsilon'}{2}$, we declare that D is not (ϵ',k)-wise independent.

For this algorithm, we need to prove two properties. The first is that if D is (ϵ,k)-wise independent, then $\tilde{\epsilon} \le \frac{\epsilon+\epsilon'}{2}$. The second is that if D is not (ϵ',k)-wise independent, then $\tilde{\epsilon} > \frac{\epsilon+\epsilon'}{2}$.

We introduce the following notation. For $C \subset [n]$, $|C| = k$, $\overrightarrow{v} \in \{0,1\}^k$, let $\tilde{p}_{C,\overrightarrow{v}} = \Pr_{x\leftarrow\tilde{D}}\left[x|_C = \overrightarrow{v}\right] - 2^{-k}$. The important property of $\tilde{p}_{C,\overrightarrow{v}}$ is that $\tilde{\epsilon} = \max_{C,\overrightarrow{v}}\left|\tilde{p}_{C,\overrightarrow{v}}\right|$.

We prove that, with high probability, $\max_{C,\overrightarrow{v}}\left|\tilde{p}_{C,\overrightarrow{v}}\right|$ is tightly concentrated around its true value. Fix any $C,\overrightarrow{v}$ as above. Then we have that $\tilde{p}_{C,\overrightarrow{v}} = \sum_{i=1}^{|Q|}\frac{X_i}{|Q|} - 2^{-k}$, where X_i is an indicator variable equal to 1 iff

$q_i|_C = \overrightarrow{v}$. Note that $\mathbf{E}\left[X_i\right] = \Pr_{x \leftarrow D}\left[x|_C = \overrightarrow{v}\right]$. By the Chernoff bound, for $\sigma = \frac{\epsilon' - \epsilon}{2}$, we have that

$$\begin{aligned} \Pr_Q\left[|\tilde{p}_{C,\overrightarrow{v}} - (\Pr_{x \leftarrow D}\left[x|_C = \overrightarrow{v}\right] - 2^{-k})| \geq \sigma\right] &= \Pr_Q\left[|\tilde{p}_{C,\overrightarrow{v}} - (\mathbf{E}\left[X_i\right] - 2^{-k})| \geq \sigma\right] \\ &= \Pr_Q\left[\left|\sum_{i=1}^{|Q|} \frac{X_i}{|Q|} - \mathbf{E}\left[X_i\right]\right| \geq \sigma\right] \\ &\leq \exp\left[-\Omega((\epsilon' - \epsilon)^2 \cdot |Q|)\right] \end{aligned}$$

Thus, setting $|Q| = O\left(\frac{k \log n}{(\epsilon' - \epsilon)^2}\right)$, with probability at least $1 - n^{-\Omega(k)}$, we have that $\left|\tilde{p}_{C,\overrightarrow{v}} - (\mathbf{E}\left[X_i\right] - 2^{-k})\right| \leq \sigma$.

Now, to show the first property, just note that $|\Pr_{x \leftarrow D}\left[x|_C = \overrightarrow{v}\right] - 2^{-k}| \leq \epsilon$ when D is (ϵ, k)-wise independent. Thus, $|\tilde{p}_{C,\overrightarrow{v}}| \leq \epsilon + \sigma = \frac{\epsilon + \epsilon'}{2}$. By a union bound, with at least a constant probability, $\max_{C,\overrightarrow{v}} |p_{C,\overrightarrow{v}}| \leq \frac{\epsilon + \epsilon'}{2}$.

For the second property, we show that $\tilde{\epsilon} > \frac{\epsilon + \epsilon'}{2}$ if D is not (ϵ', k)-wise independent. Let $C, \overrightarrow{v}$ be such that $|\Pr_{x \leftarrow D}\left[x|_C = \overrightarrow{v}\right] - 2^{-k}| > \epsilon'$. Then, by the above deviation bound, with high probability, $|\tilde{p}_{C,\overrightarrow{v}} - (\Pr_{x \leftarrow D}\left[x|_C = \overrightarrow{v}\right] - 2^{-k})| \leq \frac{\epsilon' - \epsilon}{2}$. Finally, we deduce that $\left|\tilde{p}_{C,\overrightarrow{v}}\right| > \epsilon' - \sigma = \frac{\epsilon + \epsilon'}{2}$ with high probability. □

B.3 Sample complexity lower bound: proof of Theorem 4.2

In this section we study the lower bound on sample complexity for the problem Test(ϵ, k)-independence to within distance δ.

We first study the minimum support of a distribution D which is (ϵ, k)-wise independent. We show that the minimum support of such a distribution is $\Omega\left(\frac{k}{\epsilon^2 \log \frac{1}{\epsilon}} \log n\right)$. We then apply this argument to distributions that are $(\epsilon + \delta, k)$-wise independent. Specifically, since a distribution D such that $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \leq \delta$ is also $(\epsilon + \delta, k)$-wise independent (by Lemma 4.6), the minimum support size argument implies that D has support size at least

$$|\operatorname{Supp}(D)| > \Omega\left(\frac{k}{(\epsilon + \delta)^2 \log \frac{1}{\epsilon + \delta}} \log n\right).$$

For obtaining the lower bound on sample complexity, we consider two distributions that are impossible to distinguish unless we have $\Omega\left(\sqrt{\frac{k}{(\epsilon+\delta)^2 \log \frac{1}{\epsilon+\delta}} \log n}\right)$ samples. The first one is the uniform distribution U_n over $\{0, 1\}^n$. Obviously, U_n is (ϵ, k)-wise independent. The second distribution D^* is constructed via the following random process F: define D^* to be uniform over the set S of $c \frac{k}{(\epsilon+\delta)^2 \log \frac{1}{\epsilon+\delta}} \log n$ elements chosen from U_n (with replacement) for some constant $c > 0$. Since any distribution D^* generated via F has small support size, by the above bound on the support size, we have that $\Delta(D^*, \mathcal{D}_{(\epsilon,k)}) > \delta$. However, using the birthday paradox, unless $\Omega\left(\sqrt{\frac{k}{(\epsilon+\delta)^2 \log \frac{1}{\epsilon+\delta}} \log n}\right)$ samples are drawn, both distributions U_n and an D^* generated via F look the same, and cannot be distinguished by any algorithm. The actual proof follows. The proof uses the following theorem that appears in [1].

Theorem B.4 ([1])**.** *Let B be an n by n real matrix with $b_{i,i} = 1$ for all i and $|b_{i,j}| \leq \epsilon$ for all $i \neq j$. If the rank of B is d, and $\frac{1}{\sqrt{n}} \leq \epsilon \leq \frac{1}{2}$, then*

$$d > \Omega\left(\frac{1}{\epsilon^2 \log \frac{1}{\epsilon}} \log n\right).$$

Theorem B.5 (minimum support size)**.** *Let $\frac{1}{n^{k/4}} < \epsilon < \frac{1}{2}$. The minimum support of a distribution D which is (ϵ, k)-wise independent is $\Omega\left(\frac{k}{\epsilon^2 \log \frac{1}{\epsilon}} \log n\right)$.*

Proof. Consider a distribution D that is (ϵ, k)-wise independent. Assume that D is given as a binary $s \times n$ matrix M_D where s is the support size. A restriction of M_D to a subset $\emptyset \neq I \subset [1, \cdots, n]$, $|I| \leq k$ is denoted as $M_{D,I}$ and it is an $s \times |I|$ matrix that contains the relevant columns of M_D.

For $\emptyset \neq I \subset [1, \cdots, n]$, $|I| \leq \frac{k}{2}$, consider the sum modulo 2 of the columns of $M_{D,I}$ and obtain a vector $v_{M,I}$ of length s. The weight of $v_{M,I}$ is denoted as $w(v_{M,I})$ and it refers to the number of 1's in $v_{M,I}$. The number of different sets I, $\emptyset \neq I \subset [1, \cdots, n]$, $|I| \leq \frac{k}{2}$ is $\Theta(n^{k/2})$. Consider a matrix C of dimension s by $\Theta(n^{k/2})$ whose columns are all possible vectors $v_{M,I}$. The matrix J is a matrix of all 1's. Let $C' = J - C$.

From the definition of (ϵ, k)-wise independent distribution we know that for every $\emptyset \neq I, J \subset [1, \cdots, n]$, $|I|, |J| \leq \frac{k}{2}$, $I \neq J$

$$|\frac{2w(v_{M,I} \oplus v_{M,J}) - s}{s}| \leq \epsilon.$$

Consider now a matrix B of dimension $\Theta(n^{k/2})$ by $\Theta(n^{k/2})$, where its rows and columns are indexed by different sets I, and $B_{I,J} = \frac{2w(v_{M,I} \oplus v_{M,J}) - s}{s}$. Note that $B = [2(C^t \cdot C + C'^t \cdot C') - sJ]/s$. Since C, C', J have rank at most s, the rank of B is also at most s. From the definition of (ϵ, k)-wise independent distribution we obtain that $B_{I,I} = 1$ and $|B_{I,J}| \leq \epsilon$ for $I \neq J$. Hence by Theorem B.4 we obtain

$$Rank(B) > \Omega\left(\frac{k}{\epsilon^2 \log \frac{1}{\epsilon}} \log n\right).$$

However, as mentioned above, $s \geq Rank(B)$. Hence we obtain the claimed lower bound on s:

$$s > \Omega\left(\frac{k}{\epsilon^2 \log \frac{1}{\epsilon}} \log n\right).$$

□

Corollary B.6. *Let* $\frac{1}{n^{k/4}} < \epsilon < \frac{1}{2}$, $0 < \delta \leq \frac{1}{2} - \epsilon$. *The minimum support of a distribution* D *for which* $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \leq \delta$ *is* $\Omega\left(\frac{k \log n}{(\epsilon+\delta)^2 \log \frac{1}{\epsilon+\delta}}\right)$.

Proof. By Lemma 4.6 we get that a distribution D which is δ-far from (ϵ, k)-wise independent distribution is a $(\epsilon + \delta, k)$-wise independent distribution. The corollary follows from the lower bound on the support size of a $(\epsilon + \delta, k)$-wise independent distribution as obtained in Theorem B.5. □

We are now ready to prove Theorem 4.2.

Theorem 4.2 (Sample lower bound). *For* $\epsilon > n^{-k/4}$, $0 < \delta < 1/2 - \epsilon$, *any algorithm solving* Test(ϵ, k)-independence *to within distance* δ *requires at least* $|Q| = \Omega\left(\frac{\sqrt{k \log n}}{(\epsilon+\delta)\sqrt{\log 1/(\epsilon+\delta)}}\right)$ *samples from the distribution* D.

Proof. We will show that if the algorithm has $O\left(\frac{\sqrt{k \log n}}{(\epsilon+\delta)\sqrt{\log 1/(\epsilon+\delta)}}\right)$ samples, then it can not successfully distinguish the following two distributions with high probability. For the first distribution, consider the uniform distribution U_n over $\{0, 1\}^n$. Obviously, U_n is (ϵ, k)-wise independent for all $1 \leq k \leq n$ and $0 \leq \epsilon \leq \frac{1}{2}$. The second distribution D is constructed via the following random process F: define D to be uniform over the set S of $c\frac{k \log n}{(\epsilon+\delta)^2 \log \frac{1}{\epsilon+\delta}}$ elements chosen from U_n (with replacement) for some sufficiently small constant $c > 0$. Since distribution D generated via F has small support size, by Corollary B.6, we have that $\Delta(D, \mathcal{D}_{(\epsilon,k)}) > \delta$.

Now consider the distribution of the sample set Q of size $\frac{c}{100}\frac{k \log n}{(\epsilon+\delta)^2 \log \frac{1}{\epsilon+\delta}}$. For both distribution U_n and D (chosen from F), Q is a set of l distinct element chosen from $\{0, 1\}^n$, with probability at least $11/12$. This means that for any algorithm $\mathcal{A}$, we can show that $\left|\Pr_{Q\leftarrow U_n}[\mathcal{A}(Q)=1] - \Pr_{D\leftarrow F,Q\leftarrow D}[\mathcal{A}(Q)=1]\right| \leq 1/6$. In other words, $\mathcal{A}$ cannot distinghuish U_n from D based on l samples only, with success probability of $2/3$. This completes the proof. □

Theorem B.7. *Let* $\frac{1}{n^{k/4}} < \epsilon < \epsilon' < \frac{1}{2}$. Test$(\epsilon, k)$-vs-$(\epsilon', k)$-independence *requires at least* $\Omega(\sqrt{\frac{k}{(\epsilon')^2 \log \frac{1}{\epsilon'}} \log n})$ *samples from the distribution.*

Proof. By Lemma 4.6 we get that a distribution D which is not (ϵ', k)-wise independent is at least $(\epsilon' - \epsilon)$-far from a (ϵ, k)-wise independent distribution and from the lower bound for Test(ϵ, k)-vs-(ϵ', k)-independence we obtain the claimed lower bound. □

B.4 Hardness of hidden clique: finding vs deciding

In the following, we prove that the hardness of finding a hidden clique in a random graph (conjecture 4.3) implies the hardness of deciding on the presence of a hidden clique in a random graph (conjecture 4.7).

Theorem B.8. *For* $t(n) > \Omega(\log n)$, HC-Find$[t(n)]$ *implies* HC-Decide$[t(n)/3]$.

Proof. The proof is by contradiction. Suppose, for any $n \geq n_0$, we can distinguish in polynomial time whether a graph G is drawn from $\mathcal{G}_{n,1/2}$ or $\mathcal{G}_{n,1/2,t/3}$, with probability at least $1 - 1/2n^2$. Let M be such a distinguisher.

Then, for $n \geq 3n_0$, given a graph G from $\mathcal{G}_{n,1/2,t}$, we can find a clique of size t in G using the distinguisher M as follows. Our algorithm is somewhat similar to the algorithm **BasicFind** used in [15] to find a hidden clique of size $t = \Omega(\sqrt{n})$.

```
1. Let $C = \emptyset$ (representing the current clique).
2. For each vertex $v$ of the graph $G$,
    3. Let $G_v = G \setminus \{v\} \setminus N_v$ be the graph obtained by removing $v$ together with $v$'s neighbors. Let $n_v$
       be the number of vertices in $G_v$.
    4. If $M(G_v)$ outputs "$\mathcal{G}_{n_v,1/2}$", then put $v$ into the set $C$. Do nothing if $M(G_v)$ outputs "$\mathcal{G}_{n_v,1/2,t/3}$".
5. Output $C$.
```

The intuition behind the algorithm is the following. Let K be the planted clique in G. If v is in K, then after removing v and the neighborhood N_v, we remove the entire clique K, and the remaining graph G_v is a random graph from $\mathcal{G}_{n_v,1/2}$. If $v \notin K$, then after removing v and N_v, we have deleted at most $2t/3$ of the clique with high probability, and thus the graph G_v is a random graph with a hidden clique of size at least $t/3$, i.e., chosen from $\mathcal{G}_{n_v,1/2,t'}$ for some $t' > t/3$.

More formally, consider first any vertex v such that $v \notin K$. Then we can view G_v as being generated via the following random process. Pick integer n_v as the number of vertices in the graph obtained by starting with n vertices, deleting the vertex v, and then deleting each vertex with probability $1/2$. Then pick integer t' as follows: take n_v red vertices and $n - 1 - n_v$ blue vertices, then draw randomly $t(n)$ vertices (without repetitions); set t' to be the number of red vertices that were drawn. Finally generate G_v via the process $\mathcal{G}_{n_v,1/2,t'}$. Note that $\Pr[n_v \leq 0.4n] \leq e^{-\Omega(n)}$, and $\Pr[t' \leq t(n_v)/3] \leq \Pr[t' \leq t(n)/3] \leq e^{-\Omega(t(n))}$. Thus, M, run on G_v, will output "$\mathcal{G}_{n_v,1/2,t}$" with probability $1 - e^{-\Omega(t(n))} - n^{-2}/2$.

Now consider any vertex $v \in K$. Then we can view G_v as being generated as follows. Pick n_v according to the following distribution: start with n vertices, delete vertex v and $t(n)-1$ other vertices (the other vertices of the clique K), and then delete each remaining vertex with probability 1/2; the size of the surviving graph gives n_v. Finally, we generate G_v via the process $\mathcal{G}_{n_v,1/2}$. Note that $\Pr[n_v \leq n/3] \leq e^{-\Omega(n)}$. Thus, M, run on G_v, will output "$\mathcal{G}_{n_v,1/2}$" with probability $1 - e^{-\Omega(n)} - n^{-2}/2$.

By the union bound over all vertices v, with probability at least $1 - 1/n$, the algorithm M gives the right answer for all of the n vertices v. Thus, we output $C = K$ with probability at least $1 - 1/n$. □