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
email: matulef@mit.edu

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

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

ABSTRACT

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

For the problem of distinguishing k-wise independent distributions from those that are δ-far from k-wise independence 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 may be of independent interest.

To distinguish (ϵ, k)-wise independent distributions from those that are δ-far from (ϵ, k)-wise independence 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 $\Omega\left(\frac{\sqrt{k \log n}}{2^k(\epsilon+\delta)\sqrt{\log 1/2^k(\epsilon+\delta)}}\right)$. Although these bounds are an exponential improvement (in terms of n and k) over the corresponding bounds for testing k-wise independence, we give evidence that the *time* complexity of testing (ϵ, k)-wise independence is unlikely to be $\text{poly}(n, 1/\epsilon, 1/\delta)$ for $k = \Theta(\log n)$, since this would disprove a plausible conjecture concerning 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.

Categories and Subject Descriptors: F.0 [General]; G.3 [Probability and Statistics]: Distribution functions.

General Terms: Algorithms, Theory.

Keywords: k-wise independence, almost k-wise independence, property testing, Fourier analysis, hidden-clique.

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 its restriction to any k coordinates is ϵ-close to uniform, in max-norm[1]. 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, 2, 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 an efficient algorithm for *testing* k-wise independence – that is, an algorithm that with high probability accepts distributions that are k-wise independent and rejects distributions that are δ-far in statistical distance from any k-wise independent distribution. We similarly study the problem of testing (ϵ, k)-wise independence.

Previous work addressed the problem of testing related properties of distributions, including uniformity [17, 8] and independence [7, 26, 9]. To the best of our knowledge, no previous work has addressed the problem of testing k-wise and (ϵ, k)-wise independence, however the theorems in [4] combined with a generalization of the algorithm in [17] yield natural testing algorithms which we improve upon.

*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 and supported in part by NSF grant 012702-001.

STOC'07, June 11–13, 2007, San Diego, California, USA.

[1]For formal definitions, see the Preliminaries.

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 approach yields an algorithm with $O^*(n^k/\delta^2)$ 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 nontrivial 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., $\ell_1(f, H)$?[2] In our case F is the set of all functions mapping to $[0,1]$ and sum up to 1 (i.e., distributions), and H (i.e., k-wise independent distributions) further requires that the functions have all Fourier coefficients over non-empty subsets of size at most k to be zero. 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 (if the functions were not restricted to mapping to $[0,1]$, then the lower bound is attainable thus making the inequality an equality. However, the constraint that the functions under consideration are distributions makes the problem much harder). Unfortunately, such a bound implies only very weak bounds for the ℓ_1-norm.

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

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

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	$O(\frac{k \log n}{\delta^2 \epsilon^2})$	$\Omega\left(\frac{\sqrt{k \log n}}{2^k(\epsilon+\delta)\sqrt{\log \frac{1}{2^k(\epsilon+\delta)}}}\right)$	$\frac{n^{O(k)}}{\text{poly}(\epsilon,\delta)}$ ‡	$n^{\omega(1)}$ §

†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.

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 $\Omega\left(\frac{\sqrt{k \log n}}{2^k(\epsilon+\delta)\sqrt{\log 1/2^k(\epsilon+\delta)}}\right)$ samples are required[3]. The lower bound on the sample complexity is achieved by obtaining an $\Omega\left(\frac{k \log n}{2^{2k}\epsilon^2 \log(1/2^k\epsilon)}\right)$ lower bound on the support size of any (ϵ, k)-wise independent distribution which is uniform over its support. 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 $\text{poly}(\log n/\epsilon\delta)$, 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 $\epsilon = 1/\text{poly}(n, 2^k)$. 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 $(1/\text{poly}(n, 2^k), k)$-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 member of the first family from a random member of the second cannot be done in polynomial time with a polynomial number of samples.

[3]A more careful analysis can improve the sample upper bound to $O\left(\frac{k \log n}{2^k\epsilon^2\delta^2}\right)$ for $\epsilon < 1/2^k$. We defer these details to the full version of the paper.

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, $\text{Supp}(D)$, is the set of points x at which $D(x) \neq 0$. Let $A = \{a_1, \ldots, a_m\}$ be a multi-set 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 ([3]). *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, $\left|\Pr_{x\leftarrow D}[x_{i_1}x_{i_2}\ldots x_{i_k} = \overrightarrow{v}] - 2^{-k}\right| \leq \epsilon$. When $\epsilon = 0$, we say that D is* k-wise independent. *The set of all k-wise independent distributions and (ϵ, k)-wise in-*

dependent 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 $0 \le \Delta(D_1, D_2) \le 1$.

The *distance of a distribution D to k-wise independence*, denoted $\Delta(D, \mathcal{D}_{\text{kwi}})$, is defined to be the minimum statistical distance of D to any k-wise independent distribution, i.e. $\Delta(D, \mathcal{D}_{\text{kwi}}) \triangleq \min_{D'\in\mathcal{D}_{\text{kwi}}} \Delta(D, D')$. If $\Delta(D, \mathcal{D}_{\text{kwi}}) \le \delta$, 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 $\mathcal{D}_{\text{kwi}}$.

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[4] 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 $\widehat{fg} = \hat{f}*\hat{g}$ 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 $\|f * g\|_\infty \le \|f\|_\infty\|g\|_1$, 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]. Following is the form proved in [12]:

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

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, $\hat{D}(T) = \frac{1}{2^n} bias_D(T)$, 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 $T \subset [n]$, $0 < |T| \le k$, are zero. In particular, for the uniform distribution U_n, $bias_{U_n}(T) = 0$ for all T.*

[4]If $f = D$ is a distribution, this definition differs from the commonly used distance metrics by a normalization factor. For example, for $p = 1$, $\|D\|_1 = \frac{1}{2^n}|D|_1$, where $|D|_1 = \sum_{x\in\{0,1\}^n} |D(x)|$; and for $p = 2$, $\|D\|_2 = \frac{1}{\sqrt{2^n}}|D|_2$, where $|D|_2 = \sqrt{\sum_{x\in\{0,1\}^n} |D(x)|^2}$.

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

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 $b_1 \triangleq \sum_{|S|\le k} |bias_D(S)|$ and $b_2 \triangleq \sqrt{\sum_{|S|\le k} bias_D(S)^2}$. Note that $b_2 \le b_1 \le \sqrt{M_{n,k}}b_2 < n^{k/2}b_2$.

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

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

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. This insures that all the first k-level Fourier coefficients remain zero, while increasing the weights at negative points so that they 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 over domain $\{0,1\}^n$ such that $\sum_x f(x) = 1$. Let $D_1, \ldots, D_\ell$ be distributions over $\{0,1\}^n$. Suppose there exist positive real numbers $w_1, \ldots, w_\ell$ such that $D'(x) \triangleq \frac{1}{1+\sum_{i=1}^{\ell} w_i}(f(x) + \sum_{i=1}^{\ell} w_i D_i(x))$ is non-negative for all $x \in \{0,1\}^n$. Then $\frac{2^n}{2}\|f(x) - D'(x)\|_1 \le \sum_{i=1}^{\ell} w_i$.*

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

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-Schwarz 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 $D_1(x) \leq -\lambda b_2/2^n$, and *light* if $-\lambda b_2/2^n < D_1(x) < 0$. 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 in [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 $U_{BCH\text{-}z}(z) = \frac{1}{|\operatorname{Supp}(U_{BCH\text{-}z})|} = \Omega(n^{-\lfloor k/2 \rfloor})$.* [5]

Thus, we define D' by

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

We set $w_z = \frac{|\operatorname{Supp}(U_{\text{BCH-}z})|}{2^n}b_1$. 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 this number is at most the number of points at which $D_2(x) \geq \lambda b_2/2^n$. Observe 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,

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

By Theorem 2.2, $\|f\|_p \leq \left(\sqrt{p-1}\right)^k \|f\|_2 = \left(\sqrt{p-1}\right)^k$. Plug in $\lambda = (2\sqrt{\log n})^k$ and $p = \log n$, and without loss of generality, assume that p is even, then we have

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

Therefore,

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

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

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

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

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.

The analysis of `Test-KWI-Closeness` establishes the following theorem (the full proof can be found in the full version of this paper).

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 $O^*(\frac{n^{2k}}{\delta^2})$.*

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

Algorithm `Test-KWI-Closeness`(D,k,δ)

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

Figure 1: Algorithm for testing if a distribution is k-wise independent.

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 $\Delta(D,\mathcal{D}_{\rm kwi})$ 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: $\Delta(D,\mathcal{D}_{\rm kwi}) \geq \frac{1}{2}\max_{1\leq|S|\leq k}|bias_D(S)|$. 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 $\|f\|_1 \geq \|\hat{f}\|_\infty$, which is equivalent to the bound stated in 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 the full version). 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 $\Delta(D,\mathcal{D}_{\rm kwi})$, 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 $x>0$ and $sign(x)=-1$ if $x<0$. 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 3.9. Note that by Theorem 3.1, this bound is almost tight as implied by our upper bound result.

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 multi-set $\mathcal{Q}$ and let $U_\mathcal{Q}(x)$ be the uniform distribution over $\mathcal{Q}$, then for all large enough n, $\Pr_\mathcal{Q}[\Delta(U_\mathcal{Q},\mathcal{D}_{kwi})>0.228\delta] = 1-o(1)$.*

Proof sketch. We will follow the lower bound techniques developed in Theorem 3.5 to prove this lemma. However, for ease of analysis, we will use different auxiliary functions. Let $D'(x)$ be the k-wise independent distribution with minimum statistical distance to $U_\mathcal{Q}$. Define $f_\mathcal{Q}(x) = U_\mathcal{Q}(x) - D'(x)$. Then we have $\hat{f}_\mathcal{Q}(S) = \hat{U}_\mathcal{Q}(S)$ for all $S\subseteq[n]$, $S\neq\emptyset$ and $|S|\leq k$, and $\Delta(U_\mathcal{Q},\mathcal{D}_{\rm kwi}) = 2^{n-1}\|f_\mathcal{Q}\|_1$. Define $g_\mathcal{Q}(x): \{0,1\}^n\to\mathbb{R}$ as

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

Also define the convolution $h_\mathcal{Q}(x) \triangleq (f_\mathcal{Q} * g_\mathcal{Q})(x)$. Then

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

by the convolution theorem. Applying Young's inequality, we have that $\|f_\mathcal{Q}\|_1 \geq \frac{\|h_\mathcal{Q}\|_\infty}{\|g_\mathcal{Q}\|_\infty}$. We will prove the Lemma 3.6 by proving the following two lemmas bounding $\|h_\mathcal{Q}\|_\infty$ and $\|g_\mathcal{Q}\|_\infty$, respectively.

Lemma 3.7. *For all large enough n,*

$$\Pr_\mathcal{Q}\left[\|h_\mathcal{Q}\|_\infty \geq 0.999\frac{M_{n,k}}{2^{2n}Q}\right] = 1-o(1).$$

Lemma 3.8. *Let $\frac{M_{n,k}}{n} \leq Q < 2^{n^{1/3}}$. For all $k>2$ and large enough n, $\Pr_\mathcal{Q}\left[\|g_\mathcal{Q}\|_\infty \leq \frac{2.19}{2^n}\sqrt{\frac{nM_{n,k}}{Q}}\right] = 1-o(1)$.*

Now it is easy to see that the Lemma follows from Lemma 3.7 and Lemma 3.8 by a simple union bound argument.

To prove the lower bound on $\|h_\mathcal{Q}\|_\infty$ stated in Lemma 3.7, we compute the expectation and variance of $\|h_\mathcal{Q}\|_\infty$. Then a simple application of Chebyshev's inequality gives the desired bound. The calculations are straightforward but rather tedious.

The proof of Lemma 3.8 is more involved: A simple calculation shows that $g_\mathcal{Q}(x)$ equals a summation of Q independent random variables $Y_1,\ldots,Y_Q$ determined by the random subset $\mathcal{Q}$, where $-M_{n,k} \leq Y_i \leq M_{n,k}$. However, a direct application of Hoeffding's bound to the sum can only give $\|g_\mathcal{Q}\|_\infty = O(M_{n,k})$, thus $\Delta(U_\mathcal{Q},\mathcal{D}_{\rm kwi}) = \Omega(\frac{1}{Q})$, which is too weak. We improve on this 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 observing 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's that are small and those Y_i's that are large- and sum them separately. All the details of the proof can be found in the full version of this paper. ☐

Theorem 3.9 (Sample Lower Bound)**.** *For $k > 2$ and $\delta \leq 0.228$, Testing k-wise independence requires at least $|Q| = \Omega\left(\frac{1}{\delta} \cdot \left(\frac{n}{k}\right)^{\frac{k-1}{2}}\right)$ 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 3.9 follows from Lemma 3.6, note that U_n is k-wise independent, and by Lemma 3.6, U_Q is far from k-wise independent. But any algorithm making $o(\sqrt{Q})$ queries 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 omit 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 $k = \omega(1)$. 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 $\epsilon > \frac{1}{n^{k/4}}$, $0 < \delta < \frac{1}{2^k} - \epsilon$, any tester solving* TEST(ϵ, k)-INDEPENDENCE *to within distance δ requires $|Q| = \Omega\left(\frac{\sqrt{k \log n}}{2^k(\epsilon+\delta)\sqrt{\log 1/2^k(\epsilon+\delta)}}\right)$ samples from the distribution.*

We prove these theorems in Section 4.2.

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 $1 - 1/n$. (Probability is over both the choice of G and the random coins of the algorithm.)*

We discuss this conjecture in more detail in Section 4.3.1. Assuming the conjecture, we prove the following theorem on the 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 when given access to a polynomial number of samples from the distribution.*

The proof of the theorem appears in Section 4.3.2. 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 a hardness result under a more standard conjecture, 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 that is (ϵ, k)-wise independent and a distribution that is not even (ϵ', k)-wise independent for $\epsilon' > \epsilon > 0$; we call this problem TEST(ϵ, k)-VS-(ϵ', k)-INDEPENDENCE. 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.

In sections that follow, we first define the new problem and describe its relationship to TEST(ϵ, k)-INDEPENDENCE. We then prove the sample and time complexity bounds.

4.1 Relationship between TEST(ϵ, k)-INDEPENDENCE and TEST(ϵ, k)-VS-(ϵ', k)-INDEPENDENCE

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.*

The relationship between TEST(ϵ, k)-INDEPENDENCE and TEST(ϵ, k)-VS-(ϵ', k)-INDEPENDENCE is described by the following lemma.

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.*

Proof. We break down the lemma into two key propositions and give proof sketches for each separately (complete proofs appear in the full version of the paper).

Proposition 4.7. *Let $0 < \epsilon, \delta < 1$. If $\Delta(D, \mathcal{D}_{(\epsilon,k)}) > \delta$, then $D \notin \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$.*

Proof sketch. 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 $\Delta(D, D') \leq \delta$. D' is a mixture of D and the uniform distribution U_n. □

Now, to solve the problem $\textsc{Test}(\epsilon, k)$-INDEPENDENCE on distribution D, we simply invoke $\textsc{Test}(\epsilon, k)$-VS-$(\epsilon + \epsilon\delta, k)$-INDEPENDENCE on D. The correctness follows from the above proposition.

Proposition 4.8. *Let $0 \le \epsilon < \epsilon' < 1$. If $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \le \epsilon' - \epsilon$ then $D \in \mathcal{D}_{(\epsilon',k)}$.*

Proof sketch. It is easy to verify that $D \in \mathcal{D}_{(\epsilon',k)}$ by considering $D' \in \mathcal{D}_{(\epsilon,k)}$ such that $\Delta(D, D') \le \delta = \epsilon' - \epsilon$. □

We solve $\textsc{Test}(\epsilon, k)$-VS-$(\epsilon + \delta, k)$-INDEPENDENCE on D by a simple invocation to $\textsc{Test}(\epsilon, k)$-INDEPENDENCE on D.

This ends the proof of the lemma. □

4.2 Sample Complexity Bounds

In this section, we prove upper and lower bounds on the sample complexity of $\textsc{Test}(\epsilon, k)$-INDEPENDENCE.

4.2.1 Sample complexity upper bound: proof of Theorem 4.1

We give an algorithm for $\textsc{Test}(\epsilon, k)$-VS-$(\epsilon', k)$-INDEPENDENCE, and use 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 4.9 (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., $\tilde{D}(x) = U_Q(x) = \frac{|\{i \in [|Q|] \mid q_i = x\}|}{|Q|}$, 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.

The correctness of the algorithm follows via the Chernoff bound. □

4.2.2 Sample complexity lower bound: proof of Theorem 4.2

In this section we study the lower bound on sample complexity for the problem $\textsc{Test}(\epsilon, k)$-INDEPENDENCE to within distance δ.

We first study the minimum support size of a uniform distribution D which is (ϵ, k)-wise independent, lower-bounding it by $\Omega\left(\frac{k}{(2^k\epsilon)^2 \log \frac{1}{2^k\epsilon}} \log n\right)$. We apply this bound to distributions that are $(\epsilon + \delta, k)$-wise independent and deduce, via Lemma 4.6, the minimum support size for distributions δ-close to (ϵ, k)-wise independence. The rest of the proof for the query complexity lower bound follows closely the one for testing k-wise independence, appearing in Theorem 3.9.

To prove our bound on the minimum support, we use the following theorem that appears in [1].

Theorem 4.10 ([1])**.** *Let B be an n by n real matrix with $b_{i,i} = 1$ for all i and $|b_{i,j}| \le \epsilon$ for all $i \ne j$. If the rank of B is d, and $\frac{1}{\sqrt{n}} \le \epsilon \le \frac{1}{2}$, then $d > \Omega\left(\frac{1}{\epsilon^2 \log \frac{1}{\epsilon}} \log n\right)$.*

Theorem 4.11 (Minimum Support Size)**.** *Let $\frac{1}{n^{k/4}} < \epsilon < \frac{1}{2^k}$. The minimum support size of a uniform distribution D which is (ϵ, k)-wise independent is $\Omega\left(\frac{k}{2^{2k}\epsilon^2 \log \frac{1}{2^k\epsilon}} \log n\right)$.*

Proof. Consider a uniform 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 $\emptyset \ne I \subset [n]$, $|I| \le k$ is denoted as $M_{D,I}$ and it is an $s \times |I|$ matrix that contains the relevant columns of M_D.

For $\emptyset \ne I \subset [n]$, $|I| \le \frac{k}{2}$, 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, $\emptyset \ne I \subset [n]$, $|I| \le \frac{k}{2}$ 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 independence, it follows that for every $\emptyset \ne I, I' \subset [n]$, $|I|, |I'| \le \frac{k}{2}$, $I \ne I'$, we have $\left|\frac{2w(v_{M,I} \oplus v_{M,I'}) - s}{s}\right| \le 2^k\epsilon$.

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 $B_{I,I'} = \frac{2w(v_{M,I} \oplus v_{M,I'}) - s}{s}$. Note that $B = [2(C^t \cdot C + C'^t \cdot C') - sJ]/s$. Since the rows of C and the all 1 row span the rows of C' and of J, and the rank of C is clearly at most s, the rank of B is at most $s + 1$. From the definition of (ϵ, k)-wise independence we obtain that $B_{I,I'} = 1$ and $|B_{I,I'}| \le 2^k\epsilon$ for $I \ne I'$. Hence by Theorem 4.10 we obtain $Rank(B) > \Omega\left(\frac{k}{2^{2k}\epsilon^2 \log \frac{1}{2^k\epsilon}} \log n\right)$.

However, as mentioned above, $s + 1 \ge Rank(B)$. Hence we obtain the claimed lower bound on s. □

Corollary 4.12. *Let $\frac{1}{n^{k/4}} < \epsilon < \frac{1}{2}$, $0 < \delta < \frac{1}{2^k} - \epsilon$. The minimum support of a uniform distribution D for which $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \le \delta$ is $\Omega\left(\frac{k \log n}{2^{2k}(\epsilon+\delta)^2 \log \frac{1}{2^k(\epsilon+\delta)}}\right)$.*

Proof sketch. The proof follows by applying Proposition 4.8, and the theorem above. □

The above corollary implies Theorem 4.2. We defer the proof of the theorem to the full version of the paper.

4.3 Time Complexity Bounds

In this section, we discuss the plausibility of the hidden clique conjecture (Conjecture 4.3), and present the proof of the time complexity lower bound, Theorem 4.4, based on this conjecture.

4.3.1 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 $(1-o(1))\log_2 n$ 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].

4.3.2 Time complexity lower bound: proof of Theorem 4.4

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 4.13 (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 adversarially. 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 $1-1/n^3$. (Probability is over both the choice of G and the random coins of the algorithm.)*

We prove the following lemma in Section 4.3.3.

Lemma 4.14. *For $t(n) > \Omega(\log n)$, if* HC-FIND$[t(n)]$ *holds, then* HC-DECIDE$[t(n)/3]$ *also holds.*

Given this lemma, it is now 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 4.15. *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 Theorem 4.4. ◻

4.3.3 Hardness of hidden clique: finding vs deciding

Proof of Lemma 4.14. 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.

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

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

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.

Figure 2: Algorithm for finding a hidden clique using a distinguisher $M(G_v)$ that decides whether G_v is from $\mathcal{G}_{n_v,1/2}$ or from $\mathcal{G}_{n_v,1/2,t/3}$.

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 $\Pr[n_v \le 0.4n] \le e^{-\Omega(n)}$, and $\Pr[t' \le t(n_v)/3] \le \Pr[t' \le t(n)/3] \le e^{-\Omega(t(n))}$. 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 $\Pr[n_v \le n/3] \le e^{-\Omega(n)}$. 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 $1-1/n$. $\square$

Acknowledgments. We would like to thank Piotr Indyk and Kenneth Regan for fruitful discussions and careful reading of an early draft of this paper.

5. 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.