Testing Non-uniform k-wise Independent Distributions over Product Spaces

Ronitt Rubinfeld*
MIT and Tel Aviv University
ronitt@csail.mit.edu

Ning Xie†
MIT
ningxie@csail.mit.edu

Abstract

A discrete distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$ is called *(non-uniform) k-wise independent* if for any set of k indexes $\{i_1, \dots, i_k\}$ and for any $z_1 \in \Sigma_{i_1}, \dots, z_k \in \Sigma_{i_k}$, $\Pr_{\boldsymbol{X} \sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] = \Pr_{\boldsymbol{X} \sim D}[X_{i_1} = z_1] \cdots \Pr_{\boldsymbol{X} \sim D}[X_{i_k} = z_k]$. We study the problem of testing (non-uniform) k-wise independent distributions over product spaces. For the uniform case we show an upper bound on the distance between a distribution D from k-wise independent distributions in terms of the sum of Fourier coefficients of D at vectors of weight at most k. Such a bound was previously known only when the underlying domain is $\{0,1\}^n$. For the non-uniform case, we give a new characterization of distributions being k-wise independent and further show that such a characterization is robust based on our results for the uniform case. These greatly generalize the results of Alon et al. [1] on uniform k-wise independence over the Boolean cubes to non-uniform k-wise independence over product spaces. Our results yield natural testing algorithms for k-wise independence with time and sample complexity sublinear in terms of the support size of the distribution when k is a constant. The main technical tools employed include discrete Fourier transform and the theory of linear systems of congruences.

*Research supported by NSF grants 0514771, 0728645,0732334 and Marie-Curie International Reintegration Grant PIRG03-GA-2008-231077.

†Research supported in part by an Akamai Presidential Fellowship and NSF grants 0514771, 0728645 and 0732334.

1 Introduction

Nowadays we are both blessed and cursed by the colossal amount of data available for processing. In many situations, simply scanning the whole data set once can be a daunting task. It is then natural to ask what we can do in *sublinear time*. For many computational questions, if instead of asking the decision version of the problems, one can relax the questions and consider the analogous property testing problems, then sublinear algorithms are often possible. See survey articles [18, 35, 27, 14].

Property testing algorithms [36, 19] are usually based on *robust characterizations* of the objects being tested. For instance, the linearity test introduced in [12] is based on the characterization that a function is linear if and only if the linearity test (for all x and y, it holds that $f(x) + f(y) = f(x + y)$) has acceptance probability 1. Moreover, the characterization is *robust* in the sense that if the linearity test accepts a function with probability close to 1, then the function must be also close to some linear function. Property testing often leads to a new understanding of well-studied problems and sheds insight on related problems.

In this work, we show robust characterizations of k-wise independent distributions over discrete product spaces and give sublinear-time testing algorithms based on these robust characterizations. Note that distributions over product spaces are in general not *product distributions*, which by definition are n-wise independent distributions (see below for definition).

The k-wise Independent Distributions. For a finite set Σ, a discrete probability distribution D over Σ^n is (non-uniform) *k-wise independent* if for any set of k indexes $\{i_1, \ldots, i_k\}$ and for all $z_1, \ldots, z_k \in \Sigma$, $\Pr_{\boldsymbol{X}\sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] = \Pr_{\boldsymbol{X}\sim D}[X_{i_1} = z_1] \cdots \Pr_{\boldsymbol{X}\sim D}[X_{i_k} = z_k]$. That is, restricting D to any k coordinates gives rise to a fully independent distribution. For the special case of $\Pr_{\boldsymbol{X}\sim D}[X_i = z] = \frac{1}{|\Sigma|}$ for every index i and every letter z in the alphabet, we refer to the distribution as *uniform* k-wise independent[1]. A distribution is *almost k-wise independent* if its restriction to any k coordinates is very close to some independent distribution. k-wise independent distributions look independent "locally" to any observer of only k coordinates, even though they may be far from the fully independent distributions "globally". Furthermore, k-wise independent distributions can be constructed with exponentially smaller support sizes than fully independent distributions. Because of these useful properties, k-wise independent distributions have many applications in both probability theory and computational complexity theory [23, 25, 28, 31].

Given samples drawn from a distribution, it is natural to ask, how many samples are required to tell whether the distribution is k-wise independent or far from k-wise independent, where by "far from k-wise independent" we mean that the distribution has a large statistical distance from *any* k-wise independent distribution. Usually the time and query complexity of distribution testing algorithms are measured against the support size of the distributions; For example, algorithms that test distributions over $\{0, 1\}^n$ with time complexity $o(2^n)$ are said to be sublinear-time testing algorithms.

Alon, Goldreich and Mansour [4] implicitly gave the first robust characterization of k-wise independence. Alon et al. [1] improved the bounds in [4] and also gave efficient testing algorithms. All of these results considered only uniform distributions over $\mathrm{GF}(2)$. Our work generalizes previous results in two ways: to distributions over arbitrary finite product spaces and to non-uniform k-wise independent distributions.

[1] In literature the term "k-wise independence" usually refers to uniform k-wise independence.

***Generic Algorithm* for Testing Uniform k-wise Independence**

1. Sample D uniformly and independently M times
2. Use these samples to estimate all the low-weight Fourier coefficients
3. **Accept** if the magnitudes of *all* the estimated Fourier coefficients are at most δ

Figure 1: A *Generic Algorithm* for testing uniform k-wise independence.

1.1 Our Results

Let $\Sigma = \{0, 1, \ldots, q-1\}$ be the alphabet[2] and let $D : \Sigma^n \to [0, 1]$ be the distribution to be tested. For any vector $\boldsymbol{a} \in \Sigma^n$, the Fourier coefficient of distribution D at $\boldsymbol{a}$ is $\hat{D}(\boldsymbol{a}) = \sum_{\boldsymbol{x} \in \Sigma^n} D(\boldsymbol{x}) e^{\frac{2\pi i}{q} \sum_{j=1}^{n} a_j x_j} = \mathbf{E}_{\boldsymbol{X} \sim D}\left[e^{\frac{2\pi i}{q} \sum_{j=1}^{n} a_j X_j} \right]$. The *weight* of $\boldsymbol{a}$ is the number of non-zero entries in $\boldsymbol{a}$. It is a folklore fact that a distribution D is uniform k-wise independent if and only if for all non-zero vectors $\boldsymbol{a}$ of weight at most k, $\hat{D}(\boldsymbol{a}) = 0$. A natural test for k-wise independence is thus the *Generic Algorithm* described in Fig. 1.

However, in order to prove that the *Generic Algorithm* works, one needs to show that the simple characterization of k-wise independence is *robust*. Here, *robustness* means that, for any distribution D, if all its Fourier coefficients at vectors of weight at most k are at most δ (in magnitude), then D is $\epsilon(\delta)$-close to some uniform k-wise independent distribution, where the closeness parameter ϵ depends, among other things, on the error parameter δ. Furthermore, the query and time complexity of the *Generic Algorithm* will depend on the underlying distance upper bound to k-wise independence. One of our main results is the following robust characterization of uniform k-wise independence. Let $\Delta(D, \mathcal{D}_{\text{kwi}})$ denote the distance between D and the set of k-wise independent distributions over $\{0, 1, \ldots, q-1\}^n$, then

$$\Delta(D, \mathcal{D}_{\text{kwi}}) \leq \sum_{0 < \text{wt}(\boldsymbol{a}) \leq k} \left| \hat{D}(\boldsymbol{a}) \right|.$$

Consequently, the sample complexity of our testing algorithm is $\tilde{O}\left(\frac{n^{2k}(q-1)^{2k}q^2}{\epsilon^2}\right)$ and the time complexity is $\tilde{O}\left(\frac{n^{3k}(q-1)^{3k}q^2}{\epsilon^2}\right)$, which are both sublinear when $k = O(1)$ and $q \leq \text{poly}(n)$. We further generalize these results to non-uniform k-wise independent distributions over product space, i.e., distributions over $\Sigma_1 \times \cdots \times \Sigma_n$, where $\Sigma_1, \ldots, \Sigma_n$ are (different) finite sets.

We remark that another related problem, namely testing *almost k-wise independence* over product spaces (see Appendix C for relevant definitions), admits a straightforward generalization of the testing algorithm given in [1], which was shown there to work only for the (uniform) binary case. We include the results in Appendix C for completeness.

Our results add a new understanding of the structures underlying (non-uniform) k-wise independent distributions and it is hoped that one may find other applications of these robust characterizations.

As is often the case, commutative rings demonstrate different algebraic structures from those of prime fields. For example, the recent improved construction [16] of 3-query locally decodable codes of Yekhanin [42] relies crucially on a set system construction [21] which holds only modulo composite numbers. Generalizing results in the binary field (or prime fields) to commutative rings often poses new technical challenges

[2] This is without loss of generality, since we are not assuming any field or ring structure of the underlying alphabet of the distribution. All the properties about distributions considered in this paper are invariant under permutations of the symbols in the alphabet.

and requires additional new ideas. We hope our results may find future applications in generalizing other results working in the Boolean domains to general domains.

1.2 Techniques

Previous Techniques. Given a distribution D over the binary field which is not k-wise independent, a k-wise independent distribution was constructed in [4] by mixing[3] D with a series of carefully chosen distributions in order to zero-out all the Fourier coefficients over subsets of size at most k. The total weight of the distributions used for mixing is an upper bound on the distance of D from k-wise independence. For a given subset S, the added distribution U_S is picked such that, on one hand it corrects the Fourier coefficient over S; on the other hand, U_S's Fourier coefficientover *any* other subset is zero. Using the orthogonality property of Hadamard matrices, they chose U_S to be the uniform distribution over all strings whose parity over S is 1 (or -1, depending on the sign of the distribution's bias over S). Although one can generalize it to work for prime fields, this construction breaks down when the alphabet size is a composite number.

For binary field a better bound is obtained in [1]. This is achieved by first working in the Fourier domain to remove all the first k-level Fourier coefficients of the input distribution. Such an operation ends up with a so-called "pseudo-distribution", since at some points the resulting function may assume negative values. Then a series of carefully chosen k-wise independent distributions are added to the pseudo-distribution to fix the negative points. This approach does not admit a direct generalization to the non-Boolean cases because, for larger domains, the pseudo-distributions are in general complex-valued. Nevertheless[4], one may use generalized Fourier expansion of real-valued functions to overcome this difficulty. We present this approach in Appendix A. However, the bound obtained from this approach is weaker than our main results for the uniform case which we discuss shortly. Moreover, the proof is "non-constructive" in the sense that we are not aware of what distributions should we mix with the input distribution to make it a k-wise independent one. This drawback seems make it hard to generalize the approach to handle the non-uniform case. In contrast, our results on non-uniform k-wise independence relies crucially on the fact that the correction procedure for the uniform case is explicit and all the distributions used for mixing are of some special structure (that is, they are uniform on all but at most k coordinates of the domain).

Uniform Distributions. Our results on uniform k-wise independent distributions extend the framework in [4]. As noted before, the key property used to mend a distribution into k-wise independent is the *orthogonality* relation between any pair of vectors. We first observe that all prime fields also enjoy this nice feature after some slight modifications. More specifically, for any two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{Z}_p^n$ that are *linearly independent*, the set of strings with $\sum_{i=1}^n a_i x_i \equiv j \pmod{\mathrm{p}}$ are *uniformly* distributed over $S_{\boldsymbol{b},\ell} \stackrel{\text{def}}{=} \{\boldsymbol{x} : \sum_{i=1}^n b_i x_i \equiv \ell \pmod{\mathrm{p}}\}$ for every $0 \leq \ell \leq p-1$. We will call this the *strong orthogonality* between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The case when $q = |\Sigma|$ is not a prime is less straightforward. The main difficulty is that the strong orthogonality between pairs of vectors no longer holds, even when they are linearly independent. Suppose we wish to use some distribution $U_{\boldsymbol{a}}$ to correct the bias over $\boldsymbol{a}$. A simple but important observation is that we only need that $U_{\boldsymbol{a}}$'s Fourier coefficient at $\boldsymbol{b}$ to be zero, which is a much weaker requirement than the property of being strongly orthogonal between $\boldsymbol{a}$ and $\boldsymbol{b}$. Using a classical result in linear systems of congruences due to Smith [39], we are able to show that, when $\boldsymbol{a}$ satisfies $\gcd(a_1, \ldots, a_n) = 1$ and $\boldsymbol{b}$ is not a multiple of $\boldsymbol{a}$, the set of strings with $\sum_{i=1}^n a_i x_i \equiv j \pmod{\mathrm{q}}$ are *uniformly* distributed over $S_{\boldsymbol{b},\ell}$ for ℓ's that lie in a *subgroup* of $\mathbb{Z}_q$ (compared with uniform distribution over the whole group $\mathbb{Z}_p$ for

[3] Here "mixing" means replacing the distribution D with a convex combination of D and some other distribution.

[4] We thank an anonymous referee for pointing this out.

the prime field case). We refer to this as *weak orthogonality* between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. To zero-out the Fourier coefficients at $\boldsymbol{a}$, we instead bundle the Fourier coefficient at $\boldsymbol{a}$ with the Fourier coefficients at $\ell\boldsymbol{a}$ for every $\ell = 2, \dots, q-1$, and think of them as the Fourier coefficients of some function over the one-dimensional space $\mathbb{Z}_q$. This allows us to upper bound the total weight required to simultaneously correct *all* the Fourier coefficients at $\boldsymbol{a}$ and its multiples using only $U_{\boldsymbol{a}}$. We also generalize the result to product spaces with different alphabets at different coordinates $\Omega = \Sigma_1 \times \cdots \times \Sigma_n$.

Non-uniform Distributions. One possible way of extending the upper bounds for the uniform case to the non-uniform case would be to map non-uniform probabilities to uniform probabilities over a larger domain. For example, consider when $q = 2$ a distribution D with $\Pr_D[x_i = 0] = 0.501$ and $\Pr_D[x_i = 1] = 0.499$. We could map $x_i = 0$ and $x_i = 1$ uniformly to $\{1, \dots, 501\}$ and $\{502, \dots, 1000\}$, respectively and test if the transformed distribution D' over $\{1, \dots, 1000\}$ is k-wise independent. Unfortunately, this approach produces a huge overhead on the distance upper bound, due to the fact that the alphabet size increases depends on the closeness of marginal probabilities over different letters in the alphabet. However, in the previous example we would expect D behaves very much like the uniform case rather than with an additional factor of 1000 blowup in the alphabet size. Instead we take the following approach. Consider a compressing/stretching factor for each marginal probability $\Pr_D[x_i = z]$, where $z \in \Sigma$ and $1 \le i \le n$. Specifically, let $\theta_i(z) \overset{\text{def}}{=} \frac{1}{q \Pr_D[x_i = z]}$ so that $\theta_i(z) \Pr_D[x_i = z] = \frac{1}{q}$, the probability that $x_i = z$ in the uniform distribution. If we multiply $D(\boldsymbol{x})$ for each $\boldsymbol{x}$ in the domain by a product of all such factors at each coordinate, the non-uniform k-wise independent distribution will be transformed into a uniform one. The hope is that distributions *close to* non-uniform k-wise independent will also be transformed into distributions that are *close to* uniform k-wise independent. However, this could give rise to exponentially large distribution weight at some points in the domain, making the task of estimating the relevant Fourier coefficients intractable. Intuitively, for testing k-wise independence purposes, all we need to know are the "local" weight distributions. To be more specific, for a vector $\boldsymbol{a} \in \Sigma^n$, the *support set* or simply *support* of $\boldsymbol{a}$ is $\mathrm{supp}(\boldsymbol{a}) \overset{\text{def}}{=} \{i \in [n] : a_i \neq 0\}$. For every non-zero vector $\boldsymbol{a}$ of weight at most k, we define a new *non-uniform Fourier coefficient* at $\boldsymbol{a}$ as follows. First we project D to $\mathrm{supp}(\boldsymbol{a})$ and then, for every point in the support of the projected distribution, multiply the mass of probability by a product of the compressing/stretching factors at each coordinates in $\mathrm{supp}(\boldsymbol{a})$. We denote this transformed distribution by $D'_{\mathrm{supp}(\boldsymbol{a})}$. Finally the non-uniform Fourier coefficient of D at $\boldsymbol{a}$ is defined to be the (uniform) Fourier coefficient of $D'_{\mathrm{supp}(\boldsymbol{a})}$ at $\boldsymbol{a}$. We then show a new characterization that D is non-uniform k-wise independent *if and only if* all the low-degree non-uniform Fourier coefficients of D are zero. This enables us to apply the Fourier coefficient correcting approach developed for the uniform case to the non-uniform case. Loosely speaking, for any vector $\boldsymbol{a}$, we can find a (small-weight) distribution $\mathscr{W}_{\boldsymbol{a}}$ such that mixing $D'_{\mathrm{supp}(\boldsymbol{a})}$ with $\mathscr{W}_{\boldsymbol{a}}$ zeroes-out the (uniform) Fourier coefficient at $\boldsymbol{a}$, which is, by definition, the non-uniform Fourier coefficient of D at $\boldsymbol{a}$. But this $\mathscr{W}_{\boldsymbol{a}}$ is the distribution to mix with the "transformed" distribution, i.e., $D'_{\mathrm{supp}(\boldsymbol{a})}$. To find out the distribution works for D, we apply an *inverse* compressing/stretching transformation to $\mathscr{W}_{\boldsymbol{a}}$ to get $\tilde{\mathscr{W}}_{\boldsymbol{a}}$. It turns out that mixing $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ with the original distribution D not only corrects D's non-uniform Fourier coefficient at $\boldsymbol{a}$ but also dose not increase D's non-uniform Fourier coefficients at any other vectors except those vectors whose supports are strictly contained in $\mathrm{supp}(\boldsymbol{a})$. Moreover, the transformation from $\mathscr{W}_{\boldsymbol{a}}$ to $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ may incur at most a constant (independent of n) blowup in total weight. Therefore we can start from vectors of weight k and correct the non-uniform Fourier coefficients level by level until we finish correcting vectors of weight 1 and finally obtain a k-wise independent distribution. Bounding the total weight added during this process gives an upper bound on the distance between D and non-uniform k-wise independence. We hope that the notion of non-uniform Fourier coefficients may find other applications when non-uniform independence is involved.

1.3 Other Related Research

There are many works on k-wise independence, most focus on various *constructions* of k-wise independence or distributions that approximate k-wise independence. k-wise independent random variables were first studied in probability theory [23] and then in complexity theory [13, 2, 28, 29] mainly for derandomization purposes. Constructions of almost k-wise independent distributions were studied in [31, 3, 6, 17, 10]. Construction results of non-uniform k-wise independent distributions were given in [24, 26].

There has been much activity on property testing of distributions. Some examples include testing uniformity [20, 8], independence [7], monotonicity and being unimodal [9], estimating the support sizes [34] and testing a weaker notion than k-wise independence, namely, "almost k-wise independence" [1].

Many other techniques have also been developed to generalize results from Boolean domains to arbitrary domains [15, 30, 11].

1.4 Organization

We first give some necessary definitions and preliminary facts in Section 2. Then we study testing uniform k-wise independent distributions in Section 3. The case of non-uniform k-wise independence is treated in Section 4. Some proofs as well as the discussion of testing almost k-wise independence are deferred to the Appendices.

2 Preliminaries

Let n and m be two natural numbers with $m > n$. We write $[n]$ for the set $\{1, \dots, n\}$ and $[n, m]$ for the set $\{n, n+1, \dots, m\}$. Throughout this paper, Σ always stands for a finite set. Without loss of generality, we assume that $\Sigma = \{0, 1, \dots, q-1\}$, where $q = |\Sigma|$.

We use bold letters to denote vectors in Σ^n, for example, $\boldsymbol{a}$ stands for the vector $(a_1, \dots, a_n)$ with $a_i \in \Sigma$ being the i^{th} component of $\boldsymbol{a}$. For two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in Σ^n, their inner product is $\boldsymbol{a} \cdot \boldsymbol{b} \overset{\text{def}}{=} \sum_{i=1}^{n} a_i b_i \pmod{\text{q}}$. The support of $\boldsymbol{a}$ is the set of indexes at which $\boldsymbol{a}$ is non-zero. That is, $\text{supp}(\boldsymbol{a}) = \{i \in [n] : a_i \neq 0\}$. The weight of a vector $\boldsymbol{a}$ is the cardinality of its support. Let $1 \leq k \leq n$ be an integer. We use $M(n, k, q) \overset{\text{def}}{=} \binom{n}{1}(q-1) + \cdots + \binom{n}{k}(q-1)^k$ to denote the total number of non-zero vectors in Σ^n of weight at most k. Note that $M(n, k, q) = \Theta(n^k (q-1)^k)$ for $k = O(1)$.

We assume that there is an underlying probability distribution D from which we can receive independent, identically distributed (i.i.d) samples. The domain Ω of every distribution we consider in this paper will always be finite and in general is of the form $\Omega = \Sigma_1 \times \cdots \times \Sigma_n$, where $\Sigma_1, \dots, \Sigma_n$ are finite sets. A point $\boldsymbol{x}$ in Ω is said to be *in the support* of a distribution D if $D(\boldsymbol{x}) > 0$.

Let D_1 and D_2 be two distributions over the same domain Ω. The statistical distance between D_1 and D_2 is $\Delta(D_1, D_2) = \frac{1}{2} \sum_{x \in \Omega} |D_1(x) - D_2(x)|$. One can check that statistical distance is a metric and in particular satisfies the triangle inequality. We use statistical distance as the main metric to measure closeness between distributions in this paper. For any $0 \leq \epsilon \leq 1$, one may define a new distribution to be the convex combination of D_1 and D_2: $D' = \frac{1}{1+\epsilon} D_1 + \frac{\epsilon}{1+\epsilon} D_2$. It then follows that $\Delta(D', D_1) \leq \frac{\epsilon}{1+\epsilon} \leq \epsilon$. Sometimes we abuse notation and call the non-negative function ϵD_1 a *weighted* distribution (in particular, a *small-weight distribution* if ϵ is small).

Let $S = \{i_1, \dots, i_k\} \subseteq [n]$ be an index set. Let $\boldsymbol{x}$ be an n-dimensional vector. We write $\boldsymbol{x}_S$ to denote the

k-dimensional vector obtained from projecting $\boldsymbol{x}$ to the indexes in S. Similarly, the *projection distribution* of a discrete distribution D over Σ^n with respect to S, denoted by D_S, is the distribution obtained by restricting to the coordinates in S. Namely, $D_S : \Sigma^k \to [0,1]$ is a distribution such that $D_S(z_{i_1} \cdots z_{i_k}) = \sum_{\boldsymbol{x}_S=(z_{i_1},\ldots,z_{i_k})} D(\boldsymbol{x})$. For brevity, we sometimes write $D_S(z_j : j \in S)$ for $D_S(z_{i_1} \cdots z_{i_k})$.

The k-wise Independent Distributions. Let $D : \Sigma_1 \times \cdots \times \Sigma_n \to [0,1]$ be a distribution. We say D is the *uniform* distribution if for every $\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n$, $\Pr_{\boldsymbol{X}\sim D}[\boldsymbol{X} = \boldsymbol{x}] = \frac{1}{q_1 \cdots q_n}$, where $q_i = |\Sigma_i|$. D is *k-wise independent* if for any set of k indexes $\{i_1, \ldots, i_k\}$ and for any $z_1 \cdots z_k \in \Sigma_{i_1} \times \cdots \times \Sigma_{i_k}$, $\Pr_{\boldsymbol{X}\sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] = \Pr_{\boldsymbol{X}\sim D}[X_{i_1} = z_1] \times \cdots \times \Pr_{\boldsymbol{X}\sim D}[X_{i_k} = z_k]$. D is *uniform k-wise independent* if, on top of the previous condition, we have $\Pr_{\boldsymbol{X}\sim D}[X_i = z_j] = \frac{1}{|\Sigma_i|}$ for every $1 \le i \le n$ and every $z_j \in \Sigma_i$. Let $\mathcal{D}_{\text{kwi}}$ denote the set of all uniform k-wise independent distributions. The distance between D and $\mathcal{D}_{\text{kwi}}$, denoted by $\Delta(D, \mathcal{D}_{\text{kwi}})$, is the minimum statistical distance between D and any uniform k-wise independent distribution, i.e., $\Delta(D, \mathcal{D}_{\text{kwi}}) \overset{\text{def}}{=} \min_{D' \in \mathcal{D}_{\text{kwi}}} \Delta(D, D')$.

Discrete Fourier Transform. For background on discrete Fourier transform in computer science, the reader is referred to [40, 41]. Let $f : \Sigma_1 \times \cdots \times \Sigma_n \to \mathbb{C}$ be any function defined over the discrete product space, we define the Fourier transform of D to be, for every $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$,

$$\hat{f}(\boldsymbol{a}) = \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n} f(\boldsymbol{x}) e^{2\pi i (\frac{a_1 x_1}{q_1} + \cdots + \frac{a_n x_n}{q_n})}. \tag{1}$$

$\hat{f}(\boldsymbol{a})$ is called f's Fourier coefficient at $\boldsymbol{a}$. If the weight of $\boldsymbol{a}$ is k, we then refer to $\hat{f}(\boldsymbol{a})$ as a *degree-k* or *level-k Fourier coefficient*.

One can easily verify that the inverse Fourier transform is

$$f(\boldsymbol{x}) = \frac{1}{q_1 \cdots q_n} \sum_{\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n} \hat{f}(\boldsymbol{a}) e^{-2\pi i (\frac{a_1 x_1}{q_1} + \cdots + \frac{a_n x_n}{q_n})}. \tag{2}$$

Note that if $\Sigma_i = \Sigma$ for every $1 \le i \le n$ (which is the main focus of this paper), then $\hat{f}(\boldsymbol{a}) = \sum_{\boldsymbol{x} \in \Sigma^n} f(\boldsymbol{x}) e^{\frac{2\pi i}{q} \boldsymbol{a} \cdot \boldsymbol{x}}$ and $f(\boldsymbol{x}) = \frac{1}{|\Sigma|^n} \sum_{\boldsymbol{a} \in \Sigma^n} \hat{f}(\boldsymbol{a}) e^{-\frac{2\pi i}{q} \boldsymbol{a} \cdot \boldsymbol{x}}$.

We will use the following two simple facts about discrete Fourier transform which are straightforward to prove.

Fact 2.1. *For any integer ℓ which is not congruent to 0 modulo q, $\sum_{j=0}^{q-1} e^{\frac{2\pi i}{q} \ell j} = 0$.*

Fact 2.2. *Let d, ℓ_0 be integers such that $d|q$ and $0 \le \ell_0 \le d-1$. Then $\sum_{\ell=0}^{\frac{q}{d}-1} e^{\frac{2\pi i}{q}(\ell_0 + d\ell)} = 0$.*

Proposition 2.3. *Let D be a distribution over $\Sigma_1 \times \cdots \times \Sigma_n$. Then D is the uniform distribution if and only if for any non-zero vector $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$, $\hat{D}(\boldsymbol{a}) = 0$.*

Proof. First note that $\hat{D}(0) = \sum_{\boldsymbol{x}} D(x) = 1$. Therefore, if $\hat{D}(\boldsymbol{a}) = 0$ for all non-zero $\boldsymbol{a}$, then by the inverse Fourier transform (2),

$$D(\boldsymbol{x}) = \frac{1}{q_1 \cdots q_n} \hat{D}(\boldsymbol{0}) = \frac{1}{q_1 \cdots q_n}.$$

For the converse, let $\boldsymbol{a}$ be any non-zero vector. Without loss of generality, suppose $a_1 \neq 0$. Since $D(\boldsymbol{x}) = \frac{1}{q_1 \cdots q_n}$ for all $\boldsymbol{x}$, we have

$$\begin{aligned}\hat{D}(\boldsymbol{a}) &= \frac{1}{q_1 \cdots q_n} \sum_{\boldsymbol{x}} e^{2\pi i(\frac{a_1 x_1}{q_1} + \cdots + \frac{a_n x_n}{q_n})} \\ &= \frac{1}{q_1 \cdots q_n} \sum_{x_2,\ldots,x_n} e^{2\pi i(\frac{a_2 x_2}{q_2} + \cdots + \frac{a_n x_n}{q_n})} \sum_{x_1=0}^{q_1-1} e^{\frac{2\pi i}{q_1} a_1 x_1} \\ &= 0. \qquad \text{(by Fact 2.1)}\end{aligned}$$

□

By applying Proposition 2.3 to distributions obtained from restriction to any k indexes, we have the following characterization of k-wise independent distributions over product spaces, which is the basis of all the testing algorithms in this paper.

Corollary 2.4. *A distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$ is k-wise independent if and only if, for all non-zero vectors $\boldsymbol{a}$ of weight at most k, $\hat{D}(\boldsymbol{a}) = 0$.*

Other Definitions and Notation. We are going to use the following notation extensively in this paper.

Definition 2.5. *Let D be a distribution over Σ^n. For every $\boldsymbol{a} \in \Sigma^n$ and every $0 \leq j \leq q-1$, let $P^D_{\boldsymbol{a},j} \stackrel{\text{def}}{=} \Pr_{\boldsymbol{X}\sim D}[\boldsymbol{a} \cdot \boldsymbol{X} \equiv j \pmod{\text{q}}]$. When the distribution D is clear from context, we often omit the superscript D and simply write $P_{\boldsymbol{a},j}$.*

It is worth noticing that the Fourier transform (1) can be rewritten as

$$\hat{D}(\boldsymbol{a}) = \sum_{j=0}^{q-1} \Pr_{\boldsymbol{X}\sim D}[\boldsymbol{a} \cdot \boldsymbol{X} \equiv j \pmod{\text{q}}] e^{\frac{2\pi i}{q} j} = \sum_{j=0}^{q-1} P_{\boldsymbol{a},j} e^{\frac{2\pi i}{q} j}. \tag{3}$$

For any non-zero vector $\boldsymbol{a} \in \Sigma^n$ and any integer $0 \leq j \leq q-1$, let $S_{\boldsymbol{a},j} \stackrel{\text{def}}{=} \{\boldsymbol{x} \in \Sigma^n : \sum_{i=1}^n a_i x_i \equiv j \pmod{\text{q}}\}$. Finally we write $U_{\boldsymbol{a},j}$ for the uniform distribution over $S_{\boldsymbol{a},j}$.

3 Testing Uniform k-wise Independent Distributions

3.1 Warm-up: Distributions over $\mathbb{Z}_p^n$

We begin our study with testing k-wise independent distributions when the alphabet size is a prime. Our main result is that in this case the distance between a distribution and k-wise independence can be upper bounded by the sum of the biases (to be defined later) of the distribution, slightly generalizing an idea of Alon, Goldreich and Mansour [4] use for the binary field case.

Let D be a discrete distribution over $\mathbb{Z}_p^n$, where p is a prime number.

Definition 3.1. *Let $\boldsymbol{a} \in \mathbb{Z}_p^n$ be a non-zero vector. We say D is* unbiased *over $\boldsymbol{a}$ if $P^D_{\boldsymbol{a},\ell} = 1/p$ for every $0 \leq \ell \leq p-1$. The $\mathrm{MaxBias}(\boldsymbol{a})$ of a distribution D is defined to be $\mathrm{MaxBias}_D(\boldsymbol{a}) \stackrel{\text{def}}{=} \max_{0 \leq j < p} P^D_{\boldsymbol{a},j} - \frac{1}{p}$.*

Note that the MaxBias is non-negative for any distribution. It is well-known that, for prime number p, $\hat{D}(\boldsymbol{a})$ in (3) is zero if and only $P_{\boldsymbol{a},j} = 1/p$ for every $0 \leq j \leq p-1$. Combining this with the fact that D is unbiased over $\boldsymbol{a}$ if and only if $\mathrm{MaxBias}_D(\boldsymbol{a})$ is zero, we thus have the following simple characterization of k-wise independence in terms of MaxBias.

Proposition 3.2. *D is k-wise independent iff for all non-zero $\boldsymbol{a} \in \mathbb{Z}_p^n$ with* $\mathrm{wt}(\boldsymbol{a}) \leq k$, $\mathrm{MaxBias}_D(\boldsymbol{a}) = 0$.

We say two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are *linearly dependent* if there exits some $c \in \mathbb{Z}_p^*$ such that $\boldsymbol{b} = c\boldsymbol{a}$.

Claim 3.3. *If $\boldsymbol{a}$ and $\boldsymbol{b}$ are linearly dependent, then* $\mathrm{MaxBias}_D(\boldsymbol{a}) = \mathrm{MaxBias}_D(\boldsymbol{b})$.

Proof. Suppose $\mathrm{MaxBias}_D(\boldsymbol{a})$ is attained at j, i.e., $\mathrm{MaxBias}_D(\boldsymbol{a}) = P_{\boldsymbol{a},j} - \frac{1}{p}$. Then $\mathrm{MaxBias}_D(\boldsymbol{b}) \geq P_{\boldsymbol{b},cj (\mathrm{mod\ p})} - \frac{1}{p} = P_{\boldsymbol{a},j} - \frac{1}{p} = \mathrm{MaxBias}_D(\boldsymbol{a})$. Similarly, since c^{-1} exists, we also have $\mathrm{MaxBias}_D(\boldsymbol{a}) \geq \mathrm{MaxBias}_D(\boldsymbol{b})$; consequently, $\mathrm{MaxBias}_D(\boldsymbol{a}) = \mathrm{MaxBias}_D(\boldsymbol{b})$. □

For each $\boldsymbol{a} \in \mathbb{Z}_p^n$ there are another $p-2$ vectors (namely, by taking $c = 2, \ldots, p-1$) that are linearly dependent with $\boldsymbol{a}$.

Lemma 3.4. *Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{Z}_p^n$ be two non-zero, linearly independent vectors, then for any* $0 \leq r_a, r_b \leq p-1$,

$$\Pr_{\boldsymbol{x} \in \mathbb{Z}_p^n}\left[\sum_{i=1}^{n} a_i x_i \equiv r_a \ (\mathrm{mod\ p}) \ \wedge \ \sum_{i=1}^{n} b_i x_i \equiv r_b \ (\mathrm{mod\ p})\right] = \frac{1}{p^2}$$

Proof. This follows from the well-known fact that the number of solutions to a system of 2 linearly independent linear equations over $\mathbb{Z}_p$ in n variables is p^{n-2}, independent of the vectors of free coefficients. □

Definition 3.5 (Strong Orthogonality)**.** *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two non-zero vectors in $\mathbb{Z}_p^n$. We say $\boldsymbol{a}$ is* strongly orthogonal *to $\boldsymbol{b}$ if $U_{\boldsymbol{a},j}$ is* unbiased *over $\boldsymbol{b}$ for every $0 \leq j \leq p-1$. That is,* $\Pr_{\boldsymbol{X} \sim U_{\boldsymbol{a},j}}[\boldsymbol{b} \cdot \boldsymbol{X} \equiv \ell \ (\mathrm{mod\ p})] = 1/p$, *for all* $0 \leq j, \ell \leq p-1$.

Corollary 3.6. *For any non-zero vector $\boldsymbol{a}$ and any non-zero vector $\boldsymbol{b}$ which is linearly independent of $\boldsymbol{a}$, $\boldsymbol{a}$ is strongly orthogonal to $\boldsymbol{b}$.*

Proof. Clearly we have $|S_{\boldsymbol{a},j}| = p^{n-1}$ for all non-zero $\boldsymbol{a}$ and all j. Then by Lemma 3.4, the p^{n-1} points in $S_{\boldsymbol{a},j}$ are evenly distributed over each of the p sets $S_{\boldsymbol{b},\ell}$, $0 \leq \ell \leq p-1$. □

Now we are ready to prove the following main result of this section.

Theorem 3.7. *Let D be a distribution over $\mathbb{Z}_p^n$. Then* $\Delta(D, \mathcal{D}_{kwi}) \leq \frac{p}{p-1} \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} \mathrm{MaxBias}_D(\boldsymbol{a})$.

Note that this generalizes the result of [4] for GF(2) to GF(p) for any prime p. When $p = 2$, we recover the same (implicit) bound there (our MaxBias is exactly half of their "Bias").

We first give a brief overview of the proof. We are going to prove the Theorem by constructing a k-wise independent distribution that is close to D. Generalizing the approach in [4], we start from D, step by step, zeroing-out $\mathrm{MaxBias}_D(\boldsymbol{a})$ for all vectors $\boldsymbol{a}$ of weight at most k. By Proposition 3.2, the resulting distribution will be a k-wise independent distribution. At each step, we pick any $\boldsymbol{a}$ with $\mathrm{MaxBias}_D(\boldsymbol{a}) > 0$, and the way we zero-out $\mathrm{MaxBias}_D(\boldsymbol{a})$ is to apply a convex combinations between the old distribution and some carefully chosen distribution to get a new distribution. By the strong orthogonality between linearly independent vectors (c.f. Corollary 3.6), if, for every $0 \leq j \leq q-1$, we mix the uniform distribution over all strings in $S_{\boldsymbol{a},j}$ with some appropriate weight (this weight can be zero) with D, we not only zero-out the MaxBias at $\boldsymbol{a}$ but also guarantee that, for any $\boldsymbol{b}$ that is linearly independent from $\boldsymbol{a}$, $\mathrm{MaxBias}_D(\boldsymbol{b})$ is not increased (therefore the MaxBias of zeroed-out vectors will remain zero throughout the correcting steps). This enables us to repeat the zeroing-out process for all other vectors of weight at most k and finally obtain a k-wise independent distribution.

Proof of Theorem 3.7. First we partition all the non-zero vectors of weight at most k into families of linearly dependent vectors, say $F_1, F_2, \ldots$, etc. Pick any vector $\boldsymbol{a}$ from F_1. If $\mathrm{MaxBias}_D(\boldsymbol{a}) = 0$, we move on to the next family of vectors. Now suppose $\mathrm{MaxBias}_D(\boldsymbol{a}) > 0$, and without loss of generality, assume that $P_{\boldsymbol{a},0} \leq P_{\boldsymbol{a},1} \leq \cdots \leq P_{\boldsymbol{a},p-1}$. Let $\epsilon_j = P_{\boldsymbol{a},j} - \frac{1}{p}$. Since $\sum_{j=0}^{p-1} P_{\boldsymbol{a},j} = 1$, we have $\epsilon_0 + \cdots + \epsilon_{p-1} = 0$. Also note that $\mathrm{MaxBias}_D(\boldsymbol{a}) = \epsilon_{p-1}$.

Now we define a new distribution D' as

$$D' = \frac{1}{1+\epsilon} D + \frac{\epsilon_{p-1} - \epsilon_0}{1+\epsilon} U_{\boldsymbol{a},0} + \cdots + \frac{\epsilon_{p-1} - \epsilon_{p-2}}{1+\epsilon} U_{\boldsymbol{a},p-2},$$

where $\epsilon = (\epsilon_{p-1} - \epsilon_0) + \cdots + (\epsilon_{p-1} - \epsilon_{p-2})$. Now by triangle inequality,

$$\begin{aligned} \Delta(D, D') \leq \epsilon &= (\epsilon_{p-1} - \epsilon_0) + \cdots + (\epsilon_{p-1} - \epsilon_{p-2}) \\ &= p\epsilon_{p-1} = p\mathrm{MaxBias}_D(\boldsymbol{a}). \end{aligned}$$

Moreover, due to Corollary 3.6, since $U_{\boldsymbol{a},j}$ is unbiased over $\boldsymbol{b}$ for every $0 \leq j < p$, we have for any vector $\boldsymbol{b}$ that is not in the same family with $\boldsymbol{a}$ (i.e., in $F_2, \ldots$, etc.),

$$\mathrm{MaxBias}_{D'}(\boldsymbol{b}) = \frac{1}{1+\epsilon} \mathrm{MaxBias}_D(\boldsymbol{b}) \leq \mathrm{MaxBias}_D(\boldsymbol{b}).$$

In particular, if $\mathrm{MaxBias}_D(\boldsymbol{b})$ is zero, then after zeroing-out the bias at $\boldsymbol{a}$, $\mathrm{MaxBias}_{D'}(\boldsymbol{b})$ remains zero.

Note that once we zero-out the MaxBias over $\boldsymbol{a}$, then by Claim 3.3, the biases over all other $p-2$ vectors in F_1 vanish as well (that is, we only need to perform one zeroing-out for the $p-1$ vectors in the same family). Repeating this process for all other families of vectors, we reach a distribution D_f that is unbiased over all vectors of weight at most k. By Proposition 3.2 D_f is k-wise independent, and the distance between D_f and D is at most as claimed in the theorem. □

3.2 Distributions over $\mathbb{Z}_q^n$

We now address the main problem of this Section, that is, testing k-wise independent distributions over domains of the form $\mathbb{Z}_q^n$ with q composite, where q is the size of the alphabet. Recall that a distribution D over $\mathbb{Z}_q^n$ is k-wise independent if and only if for all non-zero vectors $\boldsymbol{a}$ of weight at most k, $\hat{D}(\boldsymbol{a}) = 0$. Our main result in this Section is to show that, like in the prime field case, for every non-zero vector $\boldsymbol{a}$ of weight at most k, the following holds: There exists a (small-weight) distribution such that mixing it with D zeroes-out the Fourier coefficient at $\boldsymbol{a}$ and does not increase the Fourier coefficient at any other vector.

Unless stated otherwise, all arithmetic operations in this section are performed modulo q; For instance, we use $\boldsymbol{a} = \boldsymbol{b}$ to mean that $a_i \equiv b_i \pmod{\mathrm{q}}$ for each $1 \leq i \leq n$.

Definition 3.8 (Prime Vector)**.** *Let $\boldsymbol{a} = (a_1, \ldots, a_n)$ be a non-zero vector in $\mathbb{Z}_q^n$. $\boldsymbol{a}$ is called a* prime vector *if* $\gcd(a_1, \ldots, a_n) = 1$. *If $\boldsymbol{a}$ is a prime vector, then we refer to the set of vectors $\{2\boldsymbol{a}, \ldots, (q-1)\boldsymbol{a}\}$ (note that all these vectors are distinct) as the* siblings *of $\boldsymbol{a}$. A prime vector and its sibling vectors are collectively referred to as a* family of vectors.

Note that families of vectors do *not* form a partition of the set of all the vectors. For example when $n = 2$ and $q = 6$, vector $(4, 0)$ is a sibling of both $(1, 0)$ and $(2, 3)$, but the latter two vectors are not siblings of each other. Furthermore, there can be more than one prime vectors in a family of vectors, e.g., for $q = 6$ again, $(2, 3)$ and $(4, 3)$ are siblings while they are both prime vectors.

Recall that we use $S_{\boldsymbol{a},j}$ to denote the set $\{\boldsymbol{x} \in \mathbb{Z}_q^n : \sum_{i=1}^n a_i x_i \equiv j \pmod{\mathrm{q}}\}$.

Proposition 3.9. *If $\boldsymbol{a}$ is a prime vector, then $|S_{\boldsymbol{a},j}| = q^{n-1}$ for any $0 \leq j \leq q-1$.*

Proof. Since $\gcd(a_1, \ldots, a_n) = 1$, there exist integers $z_1, \ldots, z_n$ such that $a_1z_1 + \cdots + a_nz_n = 1$. Note that for any $\boldsymbol{z} \in \mathbb{Z}_q^n$ the map $h_{\boldsymbol{z}}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{z}$ is injective. Now if $\boldsymbol{x} \in S_{\boldsymbol{a},0}$, then $h_{\boldsymbol{z}}(\boldsymbol{x}) = (x_1+z_1, \ldots, x_n+z_n) \in S_{\boldsymbol{a},1}$. Therefore $|S_{\boldsymbol{a},0}| \leq |S_{\boldsymbol{a},1}|$. Similarly we have $|S_{\boldsymbol{a},1}| \leq |S_{\boldsymbol{a},2}| \leq \cdots \leq |S_{\boldsymbol{a},q-1}| \leq |S_{\boldsymbol{a},0}|$. Since the sets $S_{\boldsymbol{a},0}, \ldots, S_{\boldsymbol{a},q-1}$ form a partition of $\mathbb{Z}_q^n$, it follows that $|S_{\boldsymbol{a},0}| = |S_{\boldsymbol{a},1}| = \cdots = |S_{\boldsymbol{a},q-1}| = q^{n-1}$. □

3.2.1 Linear Systems of Congruences

A *linear system of congrences* is a set of linear modular arithmetic equations in some variables. We will be particularly interested in the case that all modular arithmetic equations are modulo q. If the number of variables is k, then a solution to the system of congruences is a vector in $\mathbb{Z}_q^k$. Two solutions $\boldsymbol{x}, \boldsymbol{x}'$ in $\mathbb{Z}_q^k$ are *congruent* to each other if $\boldsymbol{x} = \boldsymbol{x}'$ and *incongruent* otherwise.

We record some useful results on linear systems of congruences in this section. For more on this, the interested reader is referred to [22] and [39]. These results will be used in the next section to show some important orthogonality properties of vectors in $\mathbb{Z}_q^n$. In this section, all matrices are integer-valued. Let M be a $k \times n$ matrix with $k \leq n$. The *greatest divisor* of M is the greatest common divisor (gcd) of the determinants of all the $k \times k$ sub-matrices of M. M is a *prime matrix* if the greatest divisor of M is 1.

Lemma 3.10 ([39]). *Let M be a $(k+1) \times n$ matrix. If the sub-matrix consisting of the first k rows of M is a prime matrix and M has greatest divisor d, then there exist integers $u_1, \ldots, u_k$ such that*

$$u_1M_{1,j} + u_2M_{2,j} + \ldots + u_kM_{k,j} \equiv M_{k+1,j} \pmod{\mathrm{d}},$$

for every $1 \leq j \leq n$.

Consider the following system of linear congruent equations:

$$\begin{cases} M_{1,1}x_1 + M_{1,2}x_2 + \cdots + M_{1,n}x_n \equiv M_{1,n+1} \pmod{\mathrm{q}} \\ \quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ M_{k,1}x_1 + M_{k,2}x_2 + \cdots + M_{k,n}x_n \equiv M_{k,n+1} \pmod{\mathrm{q}}. \end{cases} \tag{4}$$

Let M denote the $k \times n$ matrix consisting of the coefficients of the linear system of equations and let $\tilde{M}$ denote the corresponding augmented matrix of M, that is, the $k \times (n+1)$ matrix with one extra column consisting of the free coefficients.

Definition 3.11. *Let M be the coefficient matrix of* (4) *and $\tilde{M}$ be the augmented matrix. Suppose $k < n$ so that system (4) is a defective system of equations. Define $Y_k, Y_{k-1}, \ldots, Y_1$, respectively, to be the greatest common divisors of the determinants of all the $k \times k$, $(k-1) \times (k-1), \ldots, 1 \times 1$, respectively, sub-matrices of M. Analogously define $Z_k, Z_{k-1}, \ldots, Z_1$ for the augmented matrix $\tilde{M}$. Also we set $Y_0 = 1$ and $Z_0 = 1$. Finally let $s = \prod_{j=1}^{k} \gcd(q, \frac{Y_j}{Y_{j-1}})$ and $t = \prod_{j=1}^{k} \gcd(q, \frac{Z_j}{Z_{j-1}})$.*

The following Theorem of Smith gives the necessary and sufficient conditions for a system of congruent equations to have solutions.

Theorem 3.12 ([39]). *If $k < n$, then the (defective) linear system of congruences (4) has solutions if and only if $s = t$. Moreover, if this condition is met, the number of incongruent solutions is sq^{n-k}.*

3.2.2 Weak Orthogonality between Families of Vectors

To generalize the proof idea of the $\mathrm{GF}(2)$ case (and also the prime field case, cf. Section 3.1) to commutative rings $\mathbb{Z}_q$ for arbitrary q, it seems crucial to relax the requirement that linearly independent vectors are strongly orthogonal. Rather, we introduce the notion of weak orthogonality between a pair of vectors.

Definition 3.13 (Weakly Orthogonality)**.** *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors in $\mathbb{Z}_q^n$. We say $\boldsymbol{a}$ is* weakly orthogonal *to $\boldsymbol{b}$ if for all $0 \le j \le q-1$, $\hat{U}_{\boldsymbol{a},j}(\boldsymbol{b}) = 0$.*

We remark that strong orthogonality implies weak orthogonality while the converse is not necessarily true. In particular, strong orthogonality does not hold in general for linearly independent vectors in $\mathbb{Z}_q^n$. However, for our purpose of constructing k-wise independent distributions, weak orthogonality between pairs of vectors suffices.

The following Lemma is the basis of our upper bound on the distance from a distribution to k-wise independence. This Lemma enables us to construct a small-weight distribution using an appropriate convex combination of $\{U_{\boldsymbol{a},j}\}_{j=0}^{q-1}$, which on the one hand zeroes-out all the Fourier coefficients at $\boldsymbol{a}$ and its sibling vectors, on the other hand has zero Fourier coefficient at all other vectors. The proof of the Lemma relies crucially on the results in Section 3.2.1 about linear system of congruences.

Lemma 3.14 (Main)**.** *Let $\boldsymbol{a}$ be a non-zero prime vector and $\boldsymbol{b}$ any non-zero vector that is not a sibling of $\boldsymbol{a}$. Then $\boldsymbol{a}$ is weakly orthogonal to $\boldsymbol{b}$.*

Proof. Consider the following system of linear congruences:

$$\begin{cases} a_1x_1 + a_2x_2 + \cdots + a_nx_n \equiv a_0 \pmod{\mathrm{q}} \\ b_1x_1 + b_2x_2 + \cdots + b_nx_n \equiv b_0 \pmod{\mathrm{q}}. \end{cases} \tag{5}$$

Following our previous notation, let $M = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix}$ and $\tilde{M} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n & a_0 \\ b_1 & b_2 & \cdots & b_n & b_0 \end{bmatrix}$. Since $\boldsymbol{a}$ is a prime vector, $Y_1 = Z_1 = 1$. We next show that Y_2 can not be a multiple of q.

Claim 3.15. *Let $M = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix}$. The determinants of all 2×2 sub-matrices of M are congruent to 0 modulo q if and only if $\boldsymbol{a}$ and $\boldsymbol{b}$ are sibling vectors.*

Proof. If $\boldsymbol{a}$ and $\boldsymbol{b}$ are sibling vectors, then it is clear that the determinants of all the sub-matrices are congruent to 0 modulo q. For the *only if* direction, we may assume that $\boldsymbol{a}$ is a prime vector, since otherwise we can divide the first row of M by the common divisor. all we need to prove is that $\boldsymbol{b} = c\boldsymbol{a}$ for some integer c. First suppose that the determinants of all 2×2 sub-matrices of M are 0. Then it follows that $\frac{b_1}{a_1} = \cdots = \frac{b_n}{a_n} = c$. If c is not an integer, then $c = \frac{u}{v}$, where u, v are integers and $\gcd(u, v) = 1$. But this implies $v|a_i$ for every $1 \le i \le n$, contradicting our assumption that $\boldsymbol{a}$ is a prime vector. Now if not all of the determinants are 0, it must be the case that the greatest divisor of the determinants of all 2×2 sub-matrices, say d', is a multiple of q. By Lemma 3.10, there is an integer c such that $ca_i \equiv b_i \pmod{\mathrm{d'}}$ for every $1 \le i \le n$. Consequently, $b_i \equiv ca_i \pmod{\mathrm{q}}$ for every i and hence $\boldsymbol{b}$ is a sibling of $\boldsymbol{a}$. □

Let $d = \gcd(q, Y_2)$. Clearly $1 \le d \le q-1$ and, according to Claim 3.15, $d|q$. Applying Theorem 3.12 with $k = 2$ to (5), the two linear congruences are solvable if and only if $d = \gcd(q, Y_2) = \gcd(q, Z_2)$. If this is the case, the total number of incongruent solutions is dq^{n-2}. Furthermore, if we let h denote the greatest common divisor of the determinants of all 2×2 sub-matrices of $\tilde{M}$, then $d|h$. By Lemma 3.10, there is an integer u such that $b_0 \equiv ua_0 \pmod{\text{h}}$. It follows that $d|(b_0 - ua_0)$. Let us consider a fixed a_0 and write $\ell_0 = ua_0 \pmod{d}$. Since $\boldsymbol{a}$ is a prime vector, by Proposition 3.9, there are in total q^{n-1} solutions to (5). But for any specific b_0 that has solutions to (5), there must be dq^{n-2} solutions to (5) and in addition $d|q$. Since there are exactly q/d b_0's in $\{0, \ldots, q-1\}$, we conclude that (5) has solutions for b_0 if and only if $b_0 = \ell_0 + d\ell$, where ℓ_0 is some constant and $\ell = 0, \ldots, \frac{q}{d} - 1$. Finally we have

$$\begin{aligned}\hat{U}_{\boldsymbol{a},j}(\boldsymbol{b}) &= \sum_{\boldsymbol{x}\in\mathbb{Z}_q^n} U_{\boldsymbol{a},j}(\boldsymbol{x}) e^{\frac{2\pi i}{q}\boldsymbol{b}\cdot\boldsymbol{x}} = \frac{1}{q^{n-1}} \sum_{\boldsymbol{a}\cdot\boldsymbol{x}\equiv j \pmod{\text{q}}} e^{\frac{2\pi i}{q}\boldsymbol{b}\cdot\boldsymbol{x}} \\ &= \frac{d}{q} \sum_{b_0 : b_0 = \ell_0 + d\ell} e^{\frac{2\pi i}{q} b_0} = 0. \qquad \text{(by Fact 2.2)}\end{aligned}$$

This finishes the proof of Lemma 3.14. □

3.2.3 Correcting the Fourier Coefficients of Sibling Vectors

Now we show how to zero-out a distribution's Fourier coefficient at every vector in a family. Let D be a distribution over $\mathbb{Z}_q^n$. By (3), for every $1 \le \ell \le q-1$, the Fourier coefficient of a vector $\ell\boldsymbol{a}$ can be rewritten as $\hat{D}(\ell\boldsymbol{a}) = \sum_{j=0}^{q-1} P_{\boldsymbol{a},j} e^{\frac{2\pi i}{q}\ell j}$. Recall that $\mathrm{MaxBias}(\boldsymbol{a}) = \max_{0\le j\le q-1} P_{\boldsymbol{a},j} - \frac{1}{q}$.

Claim 3.16. *We have that* $\mathrm{MaxBias}(\boldsymbol{a}) \le \frac{1}{q}\sum_{\ell=1}^{q-1} \left|\hat{D}(\ell\boldsymbol{a})\right|$.

Proof. Since $\hat{D}(\ell\boldsymbol{a}) = \sum_{j=0}^{q-1} P_{\boldsymbol{a},j} e^{\frac{2\pi i}{q}\ell j}$, by the inverse Fourier transform (2), for every $0 \le j \le q-1$,

$$P_{\boldsymbol{a},j} = \frac{1}{q}\sum_{\ell=0}^{q-1} \hat{D}(\ell\boldsymbol{a}) e^{-\frac{2\pi i}{q}\ell j}.$$

Since $\hat{D}(0) = 1$, we have for every $0 \le j \le q-1$,

$$\begin{aligned}\left|P_{\boldsymbol{a},j} - \frac{1}{q}\right| &= \frac{1}{q}\left|\sum_{\ell=1}^{q-1} \hat{D}(\ell\boldsymbol{a}) e^{-\frac{2\pi i}{q}\ell j}\right| \\ &\le \frac{1}{q}\sum_{\ell=1}^{q-1}\left|\hat{D}(\ell\boldsymbol{a}) e^{-\frac{2\pi i}{q}\ell j}\right| \le \frac{1}{q}\sum_{\ell=1}^{q-1}\left|\hat{D}(\ell\boldsymbol{a})\right|.\end{aligned}$$ □

Now we are ready to prove the main Theorem of this section.

Theorem 3.17. *Let D be a distribution over $\mathbb{Z}_q^n$, then* [5]

$$\Delta(D, \mathcal{D}_{\text{kwi}}) \le \sum_{0<\text{wt}(\boldsymbol{a})\le k} \left|\hat{D}(\boldsymbol{a})\right|. \qquad (6)$$

In particular, $\Delta(D, \mathcal{D}_{\text{kwi}}) \le M(n,k,q) \max_{0<\text{wt}(\boldsymbol{a})\le k} \left|\hat{D}(\boldsymbol{a})\right|$.

[5] It is easy to verify that the same bound holds for prime field case if we transform the bound in MaxBias there into a bound in terms of Fourier coefficients. Conversely we can equivalently write the bound of the distance from k-wise independence in terms of MaxBias at *prime vectors*. However, we believe that stating the bound in terms of Fourier coefficients is more natural and generalizes more easily.

Proof. Let $\boldsymbol{a}$ be a prime vector and $\hat{D}(\boldsymbol{a}), \hat{D}(\boldsymbol{2a}), \ldots, \hat{D}((q-1)\boldsymbol{a})$ be the Fourier coefficients of $\boldsymbol{a}$ and all the siblings of $\boldsymbol{a}$. Now construct a new distribution D' over $\mathbb{Z}_q^n$ as

$$D' = \frac{1}{1+\epsilon} D + \frac{1}{1+\epsilon} \sum_{j=0}^{q-1} v(j) U_{\boldsymbol{a},j},$$

where $\epsilon = \sum_{j=0}^{q-1} v(j)$ and $\{v(j)\}_{j=0}^{q-1}$ are a set of non-negative real numbers will be chosen later. It is easy to check that D' is indeed a distribution. Moreover, by Lemma 3.14 and linearity of Fourier transform, for every $\boldsymbol{b}$ that is not a sibling of $\boldsymbol{a}$,

$$\left|\hat{D}'(\boldsymbol{b})\right| = \frac{1}{1+\epsilon}\left|\hat{D}(\boldsymbol{b})\right| \leq \left|\hat{D}(\boldsymbol{b})\right|.$$

Without loss of generality, assume that $P_{\boldsymbol{a},0} \leq \cdots \leq P_{\boldsymbol{a},q-1}$. That is, $\mathrm{MaxBias}(\boldsymbol{a}) = P_{\boldsymbol{a},q-1} - \frac{1}{q}$. If we choose $v(j) = P_{\boldsymbol{a},q-1} - P_{\boldsymbol{a},j}$, then clearly $v(j)$ is non-negative for every $0 \leq j \leq q-1$. Furthermore, by our construction $P_{\boldsymbol{a},j}^{D'} = \frac{1}{q}$ for every j. Therefore by Fact 2.1, $\hat{D}'(\ell\boldsymbol{a}) = 0$ for every $1 \leq \ell \leq q-1$. Since $\sum_{j=0}^{q-1} P_{\boldsymbol{a},j} = 1$, it follows that $\sum_{j=0}^{q-1} v(j) = q\mathrm{MaxBias}(\boldsymbol{a})$. By Claim 3.16,

$$\Delta(D, D') \leq \epsilon = \sum_{j=0}^{q-1} v(j) \leq \sum_{\ell=1}^{q-1} \left|\hat{D}(\ell\boldsymbol{a})\right|.$$

Finally note that although some vectors are siblings of more than one prime vector, after its Fourier coefficient is zeroed-out, due to the fact that the distance bound in (6) is the sum of the magnitudes of Fourier coefficients at every vector in the family, the zeroed-out vector will no longer contribute to the weight added to the distribution at any later stages. This completes the proof of the Theorem. □

3.2.4 Testing Algorithm and its Analysis

In this section we prove the following theorem by presenting and analyzing a test algorithm for k-wise independence over $\mathbb{Z}_q^n$.

Theorem 3.18. *There is an algorithm that tests k-wise independence over $\{0, \ldots, q-1\}^n$ with query complexity $\tilde{O}(\frac{n^{2k}(q-1)^{2k}q^2}{\epsilon^2})$ and time complexity $\tilde{O}(\frac{n^{3k}(q-1)^{3k}q^2}{\epsilon^2})$.*

Our test algorithm is by plugging the upper bound on distance to k-wise independence in Theorem 3.17 to the *Generic Algorithm* shown in Fig. 1. As is illustrated in Figure 2, given a distribution D over $\{0, \ldots, q-1\}^n$, the algorithm tests if D is k-wise independent or ϵ-far from k-wise independent distributions. We remark that an improved upper bound on the distance from k-wise independence would immediately imply improved query and time bounds of the testing algorithm.

Now the sample and time complexity upper bounds in Theorem 3.18 are straightforward to check. The correctness of the testing algorithm is proved in the following Proposition.

Proposition 3.19. *Let D be a distribution over $\{0, \ldots, q-1\}^n$. If $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \leq \frac{\epsilon}{3qM(n,k,q)}$, then with probability at least $2/3$, **Test-Uniform-KWI**(D, k, q, ϵ) outputs **"Yes"**; if $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) > \epsilon$, then with probability at least $2/3$, **Test-Uniform-KWI**(D, k, q, ϵ) outputs **"No"**.*

We will use the following Chernoff bound in our analysis of the algorithm.

```
Test-Uniform-KWI($D, k, q, \epsilon$)
    1. Sample $D$ uniformly and independently $O\left(\frac{q^2 M(n,k,q)^2}{\epsilon^2}\log\left(qM(n,k,q)\right)\right)$ times to obtain a set $Q$
    2. Use $Q$ to estimate, for each non-zero vector $\boldsymbol{a}$ of weight at most $k$, $\hat{D}(\boldsymbol{a})$
    3. If $\max_{\boldsymbol{a}}\left|\hat{D}(\boldsymbol{a})\right| \le \frac{2\epsilon}{3M(n,k,q)}$, return "Yes"; else return "No".
```

Figure 2: Algorithm for testing if a distribution D over Σ^n is uniform k-wise independent.

Theorem 3.20 (Chernoff Bound)**.** *Let* $X_1, \dots, X_m$ *be i.i.d.* 0-1 *random variables with* $\mathbf{E}[X_i] = \mu$. *Let* $\bar{\mu} = \frac{1}{m}\sum_{i=1}^{m} X_i$. *Then for all* γ, $0 < \gamma < 1$, *we have* $\Pr[|\bar{\mu} - \mu| \ge \gamma\mu] \le 2 \cdot e^{-\frac{\gamma^2 \mu m}{3}}$.

Proof of Proposition 3.19. First note that we only need to estimate $\{P_{\boldsymbol{a},j}\}_j$ for all prime vectors, since the Fourier coefficient at any other vector is determined by the probabilities $\{P_{\boldsymbol{a},j}\}_j$ for some prime vector $\boldsymbol{a}$.

Define a 0-1 indicator variable $I_{\boldsymbol{a},j}(\boldsymbol{x})$ which is 1 if $\boldsymbol{a} \cdot \boldsymbol{x} \equiv j \pmod{\mathrm{q}}$ and 0 otherwise. Clearly $\bar{I}_{\boldsymbol{a},j} \overset{\text{def}}{=} \mathbf{E}[I_{\boldsymbol{a},j}] = P_{\boldsymbol{a},j}$. Let $m = |Q| = O\left(\frac{q^2 M(n,k,q)^2}{\epsilon^2}\log(qM(n,k,q))\right)$ and $\bar{P}_{\boldsymbol{a},j} = \frac{1}{m}\sum_{x \in Q} I_{\boldsymbol{a},j}(\boldsymbol{x})$; that is, $\bar{P}_{\boldsymbol{a},j}$ is the empirical estimate of $P_{\boldsymbol{a},j}$. Since $P_{\boldsymbol{a},j} \le 1$, by Chernoff bound, $\Pr[|\bar{P}_{\boldsymbol{a},j} - P_{\boldsymbol{a},j}| > \frac{\epsilon}{3qM(n,k,q)}] < \frac{1}{qM(n,k,q)}$. By union bound, with probability at least $2/3$, for every prime vector $\boldsymbol{a}$ of weight at most k and every $0 \le j < q$, $|\bar{P}_{\boldsymbol{a},j} - P_{\boldsymbol{a},j}| \le \frac{\epsilon}{3qM(n,k,q)}$.

The following fact provides an upper bound on the error in estimating the Fourier coefficient at vector $\boldsymbol{a}$ in terms of the errors from estimating $P_{\boldsymbol{a},j}$.

Fact 3.21. *Let* $f, g : \{0, \dots, q-1\} \to \mathbb{R}$ *with* $|f(j) - g(j)| \le \epsilon$ *for every* $0 \le j \le q-1$. *Then* $\left|\hat{f}(\ell) - \hat{g}(\ell)\right| \le q\epsilon$ *for all* $0 \le \ell \le q-1$.

Proof. Let $h = f - g$, then $|h(j)| \le \epsilon$ for every j. Therefore, $|\hat{f}(\ell) - \hat{g}(\ell)| = |\hat{h}(\ell)| = |\sum_{j=0}^{q-1} h(j) e^{\frac{2\pi i}{q}\ell j}| \le \sum_{j=0}^{q-1}|h(j)| \le \sum_{j=0}^{q-1}\epsilon = q\epsilon$. □

Let $\bar{\hat{D}}(\boldsymbol{a})$ be the estimated Fourier coefficient computed from $\bar{P}_{\boldsymbol{a},j}$. Fact 3.21 and (3) then imply that, with probability at least $2/3$, $\left|\bar{\hat{D}}(\boldsymbol{a}) - \hat{D}(\boldsymbol{a})\right| \le \frac{\epsilon}{3M(n,k,q)}$ for every $\boldsymbol{a}$.

Now if $\Delta(D, \mathcal{D}_{\text{kwi}}) \le \frac{\epsilon}{3qM(n,k,q)}$, then clearly $\left|P_{\boldsymbol{a},j} - \frac{1}{q}\right| \le \frac{\epsilon}{3qM(n,k,q)}$ for every prime vector $\boldsymbol{a}$ and $0 \le j \le q-1$, since $P_{\boldsymbol{a},j} = 1/q$ holds for every $\boldsymbol{a}$ and j for any k-wise independent distribution and no (randomized) algorithm can increase the statistical difference between two distributions [37]. By Fact 3.21, we have $\max_{\boldsymbol{a}}\left|\hat{D}(\boldsymbol{a})\right| \le \frac{\epsilon}{3M(n,k,q)}$. Taking the error from estimation into account, $\left|\bar{\hat{D}}(\boldsymbol{a})\right| \le \frac{2\epsilon}{3M(n,k,q)}$ holds with probability at least $2/3$. Therefore with probability at least $2/3$, the algorithm returns **"Yes"**.

If $\Delta(D, \mathcal{D}_{\text{kwi}}) \ge \epsilon$, then by Theorem 3.17, $\max_{\boldsymbol{a}}\left|\hat{D}(\boldsymbol{a})\right| \ge \frac{\epsilon}{M(n,k,q)}$. Again with probability at least $2/3$, $\max_{\boldsymbol{a}}\left|\bar{\hat{D}}(\boldsymbol{a})\right| \ge \frac{2\epsilon}{3M(n,k,q)}$ and the algorithm returns **"No"**. □

3.3 Distributions over Product Spaces

Now we generalize the underlying domains from $\mathbb{Z}_q^n$ to product spaces. Let $\Sigma_1, \ldots, \Sigma_n$ be n finite sets. Without loss of generality, let $\Sigma_i = \{0, 1, \ldots, q_i - 1\}$. In this section, we consider distributions over product space $\Omega = \Sigma_1 \times \cdots \times \Sigma_n$. For a set of integers $\{q_1, \ldots, q_n\}$, denote their *least common multiple* (lcm) by $\mathrm{lcm}(q_1, \ldots, q_n)$. Let $Q \overset{\mathrm{def}}{=} \mathrm{lcm}(q_1, \ldots, q_n)$ and in addition, for every $1 \le i \le n$, set $M_i = \frac{Q}{q_i}$. Then we can rewrite the Fourier coefficient of a distribution D over Ω at a vector $\boldsymbol{a} \in \Omega$ in (1) as

$$\begin{aligned}\hat{D}(\boldsymbol{a}) &= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n} D(\boldsymbol{x}) e^{\frac{2\pi i}{Q}(M_1 a_1 x_1 + \cdots + M_n a_n x_n)} \\ &= \sum_{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_n} D(\boldsymbol{x}) e^{\frac{2\pi i}{Q}(a'_1 x_1 + \cdots + a'_n x_n)},\end{aligned}$$

where $a'_i \equiv M_i a_i \pmod{\mathrm{Q}}$ for every $1 \le i \le n$. This suggests that we may view D as a distribution over Σ^n with *effective alphabet size* $|\Sigma| = Q = \mathrm{lcm}(q_1, \ldots, q_n)$ and consider the following maps from vectors in $\Sigma_1 \times \cdots \times \Sigma_n$ to vectors in $\mathbb{Z}_Q^n$:

$$\mathcal{H} : (a_1, \ldots, a_n) \mapsto (M_1 a_1 \pmod{\mathrm{Q}}, \ldots, M_n a_n \pmod{\mathrm{Q}}). \tag{7}$$

Then we only need to consider the Fourier coefficients at vectors $\boldsymbol{a}' \overset{\mathrm{def}}{=} \mathcal{H}(\boldsymbol{a}) = (a'_1, \ldots, a'_n) \in \mathbb{Z}_Q^n$ (that is, vectors in $\mathbb{Z}_Q^n$ whose i^{th} component is a multiple of M_i for every i). Note that in general $M = \mathrm{lcm}(q_1, \ldots, q_n)$ could be an exponentially large number and is therefore not easy to handle in practice[6]. However, this difficulty can be overcome by observing the following simple fact. Since we are only concerned with vectors of weight at most k, we may take different effective alphabet sizes for different index subsets of size k. For example, suppose a k-subset is $S = \{i_1, \ldots, i_k\}$, then the effective alphabet size of this index set is $|\Sigma_S| = \mathrm{lcm}(q_{i_1}, \ldots, q_{i_k})$, which is at most a polynomial in n if we assume k is a constant and each q_i is polynomially bounded.

Our main result for distributions over product spaces is the following theorem.

Theorem 3.22. *Let D be a distribution over $\Sigma_1 \times \cdots \times \Sigma_n$. Then $\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \le \sum_{0 < \mathrm{wt}(\boldsymbol{a}) \le k} \left| \hat{D}(\boldsymbol{a}) \right|$.*

We now sketch the proof of Theorem 3.22.

A vector $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$ is a *prime vector* if $\gcd(a_1, \ldots, a_n) = 1$. For any integer $\ell > 0$, let $\ell\boldsymbol{a} \overset{\mathrm{def}}{=} (\ell a_1 \pmod{q_1}, \ldots, \ell a_n \pmod{q_n})$ be a multiple of $\boldsymbol{a}$. Let $\boldsymbol{a}$ be a prime vector. Then vectors in the set $\{2\boldsymbol{a}, \ldots, (Q-1)\boldsymbol{a}\}$ are called the *sibling vectors* of $\boldsymbol{a}$. Note that these $Q - 1$ vectors may not be all distinct.

The main difficulty of applying our result for distributions over $\mathbb{Z}_q^n$ to distributions over product space is that the mapping in (7) is not surjective. In particular, after the mapping some families of vectors may have no prime vector in it. To handle this problem, we slightly generalize the result of weakly orthogonality in Lemma 3.14 to non-prime vectors. Specifically, we say a non-zero vector $\boldsymbol{a}$ (not necessarily prime) is *weakly orthogonal* to vector $\boldsymbol{b}$ if $\hat{U}_{\boldsymbol{a},\ell}(\boldsymbol{b}) = 0$ for all ℓ such that $S_{\boldsymbol{a},\ell}$ is non-empty.

Lemma 3.23. *Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors in $\mathbb{Z}_q^n$. If $\boldsymbol{b}$ is not a sibling of $\boldsymbol{a}$, then vector $\boldsymbol{a}$ is weakly orthogonal to $\boldsymbol{b}$.*

[6] Recall that the testing algorithm requires estimating all the low-degree Fourier coefficients, where each Fourier coefficient is an exponential sum with M as the denominator.

Proof. Clearly we only need to prove the case when $\boldsymbol{a}$ is not a prime vector. Let $\tilde{\boldsymbol{a}}$ be any prime vector that is a sibling of $\boldsymbol{a}$ and suppose $\boldsymbol{a} = d\tilde{\boldsymbol{a}}$. Now $S_{\boldsymbol{a},\ell}$ is non-empty only if $\ell \equiv \ell' d \pmod{q}$ for some integer ℓ'. Note that $S_{\boldsymbol{a},\ell' d} = \cup_{j:jd\equiv \ell' d \pmod{q}} S_{\tilde{\boldsymbol{a}},j}$. Since the sets $\{S_{\tilde{\boldsymbol{a}},j}\}_{j=0}^{q-1}$ are pair-wise disjoint, it follows that $U_{\boldsymbol{a},\ell' d} = \frac{1}{\gcd(d,q)} \sum_{j:jd\equiv \ell' d \pmod{q}} U_{\tilde{\boldsymbol{a}},j}$, where $\gcd(d,q)$ is the number of incongruent j's satisfying $jd \equiv \ell' d \pmod{q}$. Now by Lemma 3.14, if $\boldsymbol{b}$ is not a sibling of $\tilde{\boldsymbol{a}}$, then $\hat{U}_{\tilde{\boldsymbol{a}},j}(\boldsymbol{b}) = 0$ for every j. It follows that $\hat{U}_{\boldsymbol{a},\ell d}(\boldsymbol{b}) = 0$. □

Note that for any integer $\ell > 0$ and every $1 \leq i \leq n$, $\ell a_i \equiv b_i \pmod{q_i}$ if and only if $\ell a_i m_i \equiv b_i m_i \pmod{Q}$, it follows that the map $\mathcal{H}$ preserves the sibling relationship between vectors. Now Lemma 3.23 implies that if we map the vectors in $\Sigma_1 \times \cdots \times \Sigma_n$ to vectors in $\mathbb{Z}_Q^n$ as in (7), then we can perform the same zeroing-out process as before: For each family of vectors, zero-out all the Fourier coefficients at the vectors in this family without increasing the magnitudes of the Fourier coefficients everywhere else. This will ends up with a k-wise independent distribution over the product space $\Sigma_1 \times \cdots \times \Sigma_n$.

Next we bound the total weight required to zero-out a family of vectors. Let S be any k-subset of $[n]$. Without loss of generality, we may take $S = [k]$. Let $q_S = \operatorname{lcm}(q_1, \ldots, q_k)$ and let $m_i = \frac{q_S}{q_i}$ for each $1 \leq i \leq k$. Let $\boldsymbol{a} \in \Sigma_1 \times \cdots \times \Sigma_n$ be a prime vector with support contained in $[k]$. Then

$$\begin{aligned}
\hat{D}(\boldsymbol{a}) &= \sum_{\boldsymbol{x}\in\Sigma_1\times\cdots\times\Sigma_k} D_S(\boldsymbol{x}) e^{2\pi i(\frac{a_1 x_1}{q_1}+\cdots+\frac{a_k x_k}{q_k})} \\
&= \sum_{\boldsymbol{x}\in\Sigma_1\times\cdots\times\Sigma_k} D_S(\boldsymbol{x}) e^{\frac{2\pi i}{q_S}(m_1 a_1 x_1+\cdots+m_k a_k x_k)} \\
&= \sum_{\boldsymbol{x}\in\Sigma_1\times\cdots\times\Sigma_k} D_S(\boldsymbol{x}) e^{\frac{2\pi i}{q_S}(a'_1 x_1+\cdots+a'_k x_k)},
\end{aligned}$$

where, as before, we define $\boldsymbol{a}' = (a'_1, \ldots, a'_k)$ with $a'_i = m_i a_i \pmod{q_S}$ for $1 \leq i \leq k$.

Let $d = \gcd(m_1 a_1 \pmod{q_S}, \ldots, m_k a_k \pmod{q_S}) = \gcd(a'_1, \ldots, a'_k)$ and set $S_{\boldsymbol{a}',j} = \{\boldsymbol{x} \in \Sigma_1 \times \cdots \times \Sigma_k : a'_1 x_1 + \cdots + a'_k x_k \equiv j \pmod{q_S}\}$. Clearly $S_{\boldsymbol{a}',j}$ is non-empty only if $d|j$.

Claim 3.24. *Let $\boldsymbol{a}$ be a vector in $\Sigma_1 \times \cdots \times \Sigma_k$ with $d = \gcd(a'_1, \ldots, a'_k)$. Then $|S_{\boldsymbol{a}',\ell d}| = \frac{d q_1 \cdots q_k}{q_S}$ for every $0 \leq \ell \leq \frac{q_S}{d} - 1$.*

Proof. Since $d = \gcd(a'_1, \ldots, a'_k)$, if we let $b_i = \frac{a'_i}{d}$ for each $1 \leq i \leq k$, then $\gcd(b_1, \ldots, b_k) = 1$. Now applying the same argument as in the proof of Proposition 3.9 gives the desired result. □

Now for every $1 \leq \ell \leq \frac{q_S}{d} - 1$ and put $q^* \stackrel{\text{def}}{=} \frac{q_S}{d}$, we have

$$\begin{aligned}
\hat{D}(\ell\boldsymbol{a}) &= \sum_{\boldsymbol{x}\in\Sigma_1\times\cdots\times\Sigma_k} D_S(\boldsymbol{x}) e^{2\pi i(\frac{\ell a_1 x_1}{q_1}+\cdots+\frac{\ell a_k x_k}{q_k})} \\
&= \sum_{\boldsymbol{x}\in\Sigma_1\times\cdots\times\Sigma_k} D_S(\boldsymbol{x}) e^{\frac{2\pi i}{q_S}\ell \boldsymbol{a}'\cdot\boldsymbol{x}} = \sum_{j=0}^{\frac{q_S}{d}-1} \Pr_{\boldsymbol{X}\sim D}[\boldsymbol{a}'\cdot\boldsymbol{X} \equiv jd \pmod{q_S}] e^{\frac{2\pi i}{q_S}\ell j d} \\
&= \sum_{j=0}^{\frac{q_S}{d}-1} w(j) e^{\frac{2\pi i}{q_S}\ell j d} = \sum_{j=0}^{q^*-1} w(j) e^{\frac{2\pi i}{q^*}\ell j},
\end{aligned}$$

where $w(j) \overset{\text{def}}{=} P_{\boldsymbol{a}',jd}$. That is, each of the Fourier coefficients $\hat{D}(\boldsymbol{a}), \hat{D}(2\boldsymbol{a}), \dots, \hat{D}((q^* - 1)\boldsymbol{a})$ can be written as a one-dimensional Fourier transform of a function (namely, $w(j)$) over $\mathbb{Z}_{q^*}$. Then following the same proofs as those in Sec. 3.2.3, we have that the total weight to zero-out the Fourier coefficients at $\boldsymbol{a}$ and its siblings is at most $\sum_{\ell=1}^{\frac{q_S}{d}-1} \left|\hat{D}(\ell\boldsymbol{a})\right|$. This in turn gives the upper bound stated in Theorem 3.22 on the distance between D and k-wise independence over product spaces.

4 Testing Non-uniform k-wise Independent Distributions

In this section we focus on non-uniform k-wise independent distributions. For ease of exposition, we only prove our results for the case when the underlying domain is Σ^n with $q = |\Sigma|$. Our approach here generalizes easily to distributions over product spaces.

Recall that a distribution $D : \Sigma^n \to [0, 1]$ is k-wise independent if for any index subset $S \subset [n]$ of size k, $S = \{i_1, \dots, i_k\}$, and for any $z_1 \cdots z_k \in \Sigma^k$, $D_S(z_1 \cdots z_k) = \Pr_D[X_{i_1} = z_1] \cdots \Pr_D[X_{i_k} = z_k]$. Our strategy of showing an upper bound on the distance between D and non-uniform k-wise independence is to reduce the non-uniform problem to the uniform case and then apply Theorem 3.17.

4.1 Non-uniform Fourier Coefficients

In the following we define a set of multipliers which are used to transform non-uniform k-wise independent distributions into uniform ones. Let $p_i(z) \overset{\text{def}}{=} \Pr_D[X_i = z]$. We assume that $0 < p_i(z) < 1$ for every $i \in [n]$ and every $z \in \Sigma$ (this is without loss of generality since if some $p_i(z)$'s are zero, then it reduces to the case of distributions over product spaces). Let $\theta_i(z) \overset{\text{def}}{=} \frac{1}{qp_i(z)}$. Intuitively, one may think $\theta_i(z)$'s as a set of compressing/stretching factors which transform a non-uniform k-wise distribution into a uniform one. For convenience of notation, if $S = \{i_1, \dots, i_\ell\}$ and $\boldsymbol{z} = z_{i_1} \cdots z_{i_\ell}$, we write $\theta_S(\boldsymbol{z})$ for the product $\theta_{i_1}(z_{i_1}) \cdots \theta_{i_\ell}(z_{i_\ell})$.

Definition 4.1 (Non-uniform Fourier Coefficients)**.** *Let D be a distribution over Σ^n. Let $\boldsymbol{a}$ be a non-zero vector in Σ^n with* $\operatorname{supp}(\boldsymbol{a})$ *being its support set and $D_{\operatorname{supp}(\boldsymbol{a})}$ be the projection distribution of D with respect to* $\operatorname{supp}(\boldsymbol{a})$. *Set $D'_{\operatorname{supp}(\boldsymbol{a})}(\boldsymbol{z}) = \theta_{\operatorname{supp}(\boldsymbol{a})}(\boldsymbol{z}) D_{\operatorname{supp}(\boldsymbol{a})}(\boldsymbol{z})$, which is the transformed distribution*[7] *of the projection distribution $D_{\operatorname{supp}(\boldsymbol{a})}$. Then, $\hat{D}^{\text{non}}(\boldsymbol{a})$, the* non-uniform Fourier coefficient *of D at $\boldsymbol{a}$ is defined as*

$$\hat{D}^{\text{non}}(\boldsymbol{a}) \overset{\text{def}}{=} \hat{D'}_{\operatorname{supp}(\boldsymbol{a})}(\boldsymbol{a}) = \sum_{\boldsymbol{z} \in \Sigma^{\operatorname{supp}(\boldsymbol{a})}} D'_{\operatorname{supp}(\boldsymbol{a})}(\boldsymbol{z}) e^{\frac{2\pi i}{q} \boldsymbol{a} \cdot \boldsymbol{z}}. \tag{8}$$

Remarks 4.2. *Note that we will always refer to $\hat{D}^{\text{non}}$ collectively as a set of (complex) numbers that will be used to indicate the distance between distribution D and non-uniform k-wise independence. Strictly speaking, $\hat{D}^{\text{non}}$ are not Fourier coefficients since in general there is no function whose (low degree) Fourier coefficients are exactly $\hat{D}^{\text{non}}$.*

To summarize, let us define a function

$$\mathcal{F} : (\mathbb{R}^{\geq 0})^{\Sigma^n} \times \Sigma^k \to (\mathbb{R}^{\geq 0})^{\Sigma^k}$$

[7] Note that in general $D'_{\operatorname{supp}(\boldsymbol{a})}$ is not a distribution, since although it is non-negative everywhere but $\sum_{\boldsymbol{x}} D'_{\operatorname{supp}(\boldsymbol{a})}(\boldsymbol{x}) = 1$ may not hold.

which maps a distribution D over Σ^n and a vector $\boldsymbol{a} \in \Sigma^n$ of weight k to a non-negative function over $\Sigma^{|\mathrm{supp}(\boldsymbol{a})|}$. That is, for every $\boldsymbol{z} \in \Sigma^k$,

$$\mathcal{F}(D, \boldsymbol{a})(\boldsymbol{z}) = D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})\theta_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}). \tag{9}$$

Then the non-uniform Fourier coefficient of D at $\boldsymbol{a}$ is simply the uniform Fourier coefficient of $\hat{\mathcal{F}}(D, \boldsymbol{a})$ at $\boldsymbol{a}$:

$$\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = \hat{\mathcal{F}}(D, \boldsymbol{a})(\boldsymbol{a}).$$

The idea of defining $D'_{\mathrm{supp}(\boldsymbol{a})}$ is that, if D is non-uniform k-wise independent, then $D'_{\mathrm{supp}(\boldsymbol{a})}$ will be a uniform distribution over the index set $\mathrm{supp}(\boldsymbol{a})$. Indeed, our main result in this section is to show a connection between the non-uniform Fourier coefficients of D and the property that distribution D is non-uniform k-wise independent. In particular we have the following simple characterization of non-uniform k-wise independence.

Theorem 4.3. *A distribution D over Σ^n is k-wise independent if and only if, for every non-zero vector $\boldsymbol{a} \in \Sigma^n$ with $\mathrm{wt}(\boldsymbol{a}) \leq k$, $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = 0$.*

The proof of Theorem 4.3 may be found in Appendix B. It is straightforward to show that if D is a non-uniform k-wise independent distribution, then all the non-zero non-uniform Fourier coefficients of degree at most k are zero. However, the proof of the converse is more involved. The key observation is that, if we write the non-uniform Fourier transform as a linear transformation, the non-uniform Fourier transform matrix, like the uniform Fourier transform matrix, can be expressed as a tensor product of a set of heterogeneous DFT (discrete Fourier transform) matrices (as opposed to homogeneous DFT matrices in the uniform case). This enables us to show that the non-uniform Fourier transform is invertible. When combined with the condition that all non-zero non-uniform Fourier coefficients are zero, this invertibility property implies that D must be a non-uniform k-wise independent distribution.

4.2 Zeroing-out Non-uniform Fourier Coefficients

Given a distribution D which is not k-wise independent, what is its distance to non-uniform k-wise independence? In the following, we will follow the same approach that is used in the uniform case and try to find a set of small-weight distributions to mix with D to zero-out all the non-uniform Fourier coefficients at vectors of weight at most k. Moreover, we can bound the total weight added to the original distribution in this zeroing-out process in terms of the non-uniform Fourier coefficients of distribution D. This will show the robustness of characterization of non-uniform k-wise independence given in Theorem 4.3.

A careful inspection of Theorem 3.17 and its proof shows that, if we focus on the weights added to correct any fixed prime vector and its siblings, we actually prove the following.

Theorem 4.4. *Let E' be a non-negative function[8] defined over Σ^n, $\boldsymbol{a}$ be a prime vector of weight at most k and $\hat{E}'(\boldsymbol{a}), \hat{E}'(2\boldsymbol{a}), \ldots, \hat{E}'((q-1)\boldsymbol{a})$ be the Fourier coefficients at $\boldsymbol{a}$ and its sibling vectors. Then there exist a set of non-negative real numbers $w_j, j = 0, 1, \ldots, q-1$, such that the (small-weight) distribution[9] $\mathscr{W}_{E',\boldsymbol{a}} \overset{\mathrm{def}}{=} \sum_{j=0}^{q-1} w_j U_{\boldsymbol{a},j}$ has the following properties. The Fourier coefficients of $E' + \mathscr{W}_{E',\boldsymbol{a}}$ at $\boldsymbol{a}, 2\boldsymbol{a}, \ldots, (q-1)\boldsymbol{a}$ all equal zero and $\hat{\mathscr{W}}_{E',\boldsymbol{a}}(\boldsymbol{b}) = 0$ for all non-zero vectors that are not siblings of $\boldsymbol{a}$. Moreover, the total weight of $\mathscr{W}_{E',\boldsymbol{a}}$ satisfies $\sum_{j=0}^{q-1} w_j \leq \sum_{\ell=1}^{q-1} \left|\hat{E}'(\ell\boldsymbol{a})\right|$.*

[8] In Theorem 3.17 we only prove this for the case when E' is a distribution. However it is easy to see that the result generalizes to non-negative functions as well.

[9] Recall that $U_{\boldsymbol{a},j}$ is the uniform distribution over all strings $x \in \mathbb{Z}_q^n$ with $\boldsymbol{a} \cdot \boldsymbol{x} \equiv j \pmod{\mathrm{q}}$.

Applying Theorem 4.4 with E' equal to $D'_{\mathrm{supp}(\boldsymbol{a})}$ gives rise to a small-weight distribution $\mathscr{W}_{D'_{\mathrm{supp}(\boldsymbol{a})},\boldsymbol{a}}$ which, by abuse of notation, we denote by $\mathscr{W}_{\boldsymbol{a}}$. When we add $\mathscr{W}_{\boldsymbol{a}}$ to $D'_{\mathrm{supp}(\boldsymbol{a})}$, the resulting non-negative function has zero Fourier coefficients at $\boldsymbol{a}$ and all its sibling vectors. That is,

$$\hat{\mathscr{W}}_{\boldsymbol{a}}(\ell\boldsymbol{a}) = -\hat{D}'_{\mathrm{supp}(\boldsymbol{a})}(\ell\boldsymbol{a}), \qquad \text{for every } 1 \le \ell \le q-1. \tag{10}$$

$$= -\hat{D}^{\mathrm{non}}(\ell'\boldsymbol{a}), \qquad \text{for every } \ell' \text{ such that } \mathrm{Supp}(\ell'\boldsymbol{a}) = \mathrm{supp}(\boldsymbol{a}). \tag{10$'$}$$

and for any $\boldsymbol{b}$ which is not a sibling vector of $\boldsymbol{a}$,

$$\hat{\mathscr{W}}_{\boldsymbol{a}}(\boldsymbol{b}) = 0. \tag{11}$$

However, this small-weight distribution only works for $D'_{\mathrm{supp}(\boldsymbol{a})}$ but what we are looking for is a small-weight distribution that corrects the non-uniform Fourier coefficients of D at $\boldsymbol{a}$. To this end, we apply the reversed compressing/stretching factor to $\mathscr{U}_{\boldsymbol{a}}$ to get $\tilde{\mathscr{W}}_{\boldsymbol{a}}$,

$$\tilde{\mathscr{W}}_{\boldsymbol{a}}(\boldsymbol{x}) \stackrel{\mathrm{def}}{=} \frac{\mathscr{W}_{\boldsymbol{a}}(\boldsymbol{x})}{\theta_{[n]}(\boldsymbol{x})}. \tag{12}$$

The following Lemma shows that mixing D with $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ results in a distribution whose non-uniform Fourier coefficients at $\boldsymbol{a}$ as well as its sibling vectors are zero[10]. In addition, the mixing only adds a relative small amount of weight and may increase the magnitudes of the non-uniform Fourier coefficients only at vectors whose supports are completely contained in the support of $\boldsymbol{a}$.

Lemma 4.5. *Let D be a distribution over Σ^n and $\boldsymbol{a}$ be a prime vector of weight at most k. Let* $\mathrm{supp}(\boldsymbol{a})$ *be the support set of $\boldsymbol{a}$ and $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ be as defined in* (12)*. Let the maximum multiplier over all possible compressing/stretching factors be denoted as $\gamma_k \stackrel{\mathrm{def}}{=} \max_{S,\boldsymbol{z}} \frac{1}{\theta_S(\boldsymbol{z})}$, where S is a subset of $[n]$ of size at most k and $\boldsymbol{z} \in \Sigma^{|S|}$. Then $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ satisfies the following properties:*

1. *The non-uniform Fourier coefficients of $D + \tilde{\mathscr{W}}_{\boldsymbol{a}}$ at $\boldsymbol{a}$ as well as at the sibling vectors of $\boldsymbol{a}$ whose support sets are also* $\mathrm{supp}(\boldsymbol{a})$ *are all zero.*[11] *Moreover, $\hat{\tilde{\mathscr{W}}}^{\mathrm{non}}_{\mathbf{a}}(\boldsymbol{a}') = 0$ for every vector $\mathbf{a}'$ whose support set is* $\mathrm{supp}(\boldsymbol{a})$ *but is not a sibling vector of* $\mathbf{a}$.

2. *For any vector $\mathbf{b}$ with* $\mathrm{supp}(\boldsymbol{b}) \not\subseteq \mathrm{supp}(\boldsymbol{a})$, $\hat{\tilde{\mathscr{W}}}^{\mathrm{non}}_{\mathbf{a}}(\boldsymbol{b}) = 0$.

3. *The total weight of $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ is at most $\gamma_k \sum_{\boldsymbol{x}\in\Sigma^n} \mathscr{W}_{\boldsymbol{a}}(\boldsymbol{x}) \le \gamma_k \sum_{j=1}^{q-1} \left|\hat{D}^{\mathrm{non}}(j\boldsymbol{a})\right|$.*

4. *For any non-zero vector $\mathbf{c}$ with* $\mathrm{supp}(\boldsymbol{c}) \subset \mathrm{supp}(\boldsymbol{a})$, $\hat{\tilde{\mathscr{W}}}^{\mathrm{non}}_{\mathbf{a}}(\boldsymbol{c}) \le \gamma_k \sum_{j=1}^{q-1} \left|\hat{D}^{\mathrm{non}}(j\boldsymbol{a})\right|$.

Proof. For simplicity, we assume that $\mathrm{supp}(\boldsymbol{a}) = [k]$. Recall that $\mathscr{W}_{\boldsymbol{a}} = \sum_{j=0}^{q-1} w_j U_{\boldsymbol{a},j}$ and $U_{\boldsymbol{a},j}$ is the uniform distribution over the strings $\boldsymbol{x} \in \mathbb{Z}_q^n$ such that $\sum_{i=1}^n a_i x_i \equiv j \pmod{\mathrm{q}}$. A key observation is the following: Since the support of $\boldsymbol{a}$ is $[k]$, if $x_1 \cdots x_k$ satisfies the constraint $\sum_{i=1}^k a_i x_i \equiv j \pmod{\mathrm{q}}$, then for any $y_{k+1} \cdots y_n \in \Sigma^{n-k}$, $x_1 \cdots x_k y_{k+1} \cdots y_n$ will satisfy the constraint and thus is in the support of the distribution.

[10] In fact, this only guarantees to zero-out the Fourier coefficients at $\boldsymbol{a}$ and its siblings whose support sets are the same as that of $\boldsymbol{a}$. But that suffices for our correcting purposes because we will proceed to vectors with smaller support sets in later stages.

[11] It worth noting that, if $\boldsymbol{a}$ is a prime vector and $\boldsymbol{a}'$ is a sibling vector of $\boldsymbol{a}$, then $\mathrm{supp}(\boldsymbol{a}') \subseteq \mathrm{supp}(\boldsymbol{a})$.

Remark on notation. *In the rest of this section, we always write $\boldsymbol{x}$ for an n-bit vector in Σ^n and write $\boldsymbol{z}$ for a k-bit vector in Σ^k.*

Note that we may decompose $\mathscr{W}_{\boldsymbol{a}}$ (or any non-negative function) into a sum of q^k weighted distributions as $\mathscr{W}_{\boldsymbol{a}} = \sum_{\boldsymbol{z}\in\Sigma^k} w_{\boldsymbol{z}}\mathscr{U}_{\boldsymbol{z}}$, such that each of the distribution $\mathscr{U}_{\boldsymbol{z}}$ is supported on the $|\Sigma|^{n-k}$ strings whose k-bit prefixes equal $\boldsymbol{z}$. That is,

$$w_{\boldsymbol{z}}\mathscr{U}_{\boldsymbol{z}}(\boldsymbol{x}) = \begin{cases} \mathscr{W}_{\boldsymbol{a}}(\boldsymbol{x}), & \text{if } \boldsymbol{x}_{[k]} = \boldsymbol{z}, \\ 0, & \text{otherwise.} \end{cases}$$

To make $\mathscr{U}_{\boldsymbol{z}}$ indeed a distribution, i.e., $\sum_{\boldsymbol{x}} \mathscr{U}_{\boldsymbol{z}}(\boldsymbol{x}) = 1$, we simply set

$$w_{\boldsymbol{z}} \stackrel{\text{def}}{=} (\mathscr{W}_{\boldsymbol{a}})_{[k]}(\boldsymbol{z}). \tag{13}$$

That is, $w_{\boldsymbol{z}}$ equals the mass of the projected distribution $\mathscr{W}_{\boldsymbol{a}}$ at $\boldsymbol{z}$. By Theorem 4.4 clearly we have

$$\sum_{\boldsymbol{z}\in\Sigma^k} w_{\boldsymbol{z}} \le \sum_{j=1}^{q-1} \left|\hat{D}^{\text{non}}(j\boldsymbol{a})\right|. \tag{14}$$

The aforementioned observation then implies that, for every $\boldsymbol{z} \in \Sigma^k$, $\mathscr{U}_{\boldsymbol{z}}$ is the uniform distribution over all $|\Sigma|^{n-k}$ strings whose k-bit prefixes equal $\boldsymbol{z}$. In other words, $\mathscr{U}_{\boldsymbol{z}}$ is uniform over the strings in its support. We will refer to these distributions as *atomic uniform distributions*. More explicitly,

$$\mathscr{U}_{\boldsymbol{z}}(\boldsymbol{x}) = \begin{cases} \frac{1}{q^{n-k}}, & \text{if } \boldsymbol{x}_{[k]} = \boldsymbol{z}, \\ 0, & \text{otherwise.} \end{cases}$$

After applying the compressing/stretching factor, $\mathscr{U}_{\boldsymbol{z}}$ is transformed into $\tilde{\mathscr{U}}_{\boldsymbol{z}}$:

$$\tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x}) = \begin{cases} \frac{1}{q^{n-k}\theta_{[n]}(\boldsymbol{x})}, & \text{if } \boldsymbol{x}_{[k]} = \boldsymbol{z}, \\ 0, & \text{otherwise.} \end{cases}$$

We will call $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ a *transformed atomic uniform distribution*. Clearly we have

$$\tilde{\mathscr{W}}_{\boldsymbol{a}} = \sum_{\boldsymbol{z}\in\Sigma^k} w_{\boldsymbol{z}}\tilde{\mathscr{U}}_{\boldsymbol{z}}.$$

We remark that both atomic uniform distributions and transformed atomic uniform distributions are introduced only for the sake of analysis; they play no role in the testing algorithm.

Our plan is to show the following: On one hand, $\{w_{\boldsymbol{z}}\tilde{\mathscr{U}}_{\boldsymbol{z}}\}_{\boldsymbol{z}}$, the weighted transformed atomic uniform distributions, *collectively* zero-out the non-uniform Fourier coefficients of D at $\boldsymbol{a}$ and all the sibling vectors of $\boldsymbol{a}$ whose supports are the same as $\boldsymbol{a}$. On the other hand, *individually*, each transformed atomic uniform distribution $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ has zero Fourier coefficient at any vector whose support is not a subset of $\text{supp}(\boldsymbol{a})$. Then by linearity of Fourier transform, $\tilde{\mathscr{W}}_{\boldsymbol{a}}$ also has zero Fourier coefficients at these vectors.

We first show that, if we project $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ to index set $[k]$ to obtain distribution $\left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[k]}$, then $\left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[k]}$ is supported only on a single string (namely $\boldsymbol{z}$) and has total weight $\frac{1}{\theta_{[k]}(\boldsymbol{z})}$, which is independent of compressing/stretching factors applied to the last $n-k$ coordinates.

Remark on notation. *To simplify calculation, we will use Kronecker's delta function, $\delta(\boldsymbol{u}, \boldsymbol{v})$, in the following. By definition, $\delta(\boldsymbol{u}, \boldsymbol{v})$ equals 1 if $\boldsymbol{u} = \boldsymbol{v}$ and 0 otherwise. An important property of δ-function is $\sum_{\boldsymbol{u}'} f(\boldsymbol{u}')\delta(\boldsymbol{u}, \boldsymbol{u}') = f(\boldsymbol{u})$, where f is an arbitrary function.*

Claim 4.6. *We have*

$$\left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[k]}(\boldsymbol{z}') = \frac{\delta(\boldsymbol{z}', \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})}, \tag{15}$$

and consequently

$$\sum_{\boldsymbol{x} \in \Sigma^n} \tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x}) = \frac{1}{\theta_{[k]}(\boldsymbol{z})}. \tag{15$'$}$$

Proof. Note that $\tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x})$ can be written as

$$\tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x}) = \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \frac{1}{q^{n-k}\theta_{[k+1,n]}(\boldsymbol{x}_{[k+1,n]})} = \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \frac{1}{q^{n-k}\theta_{k+1}(x_{k+1})\cdots\theta_n(x_n)}.$$

Then by simple calculation,

$$\begin{aligned}
\left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[k]}(\boldsymbol{x}_{[k]}) = \sum_{x_{k+1},\ldots,x_n} \tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x}) &= \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \sum_{x_{k+1},\ldots,x_n} \frac{1}{q^{n-k}\theta_{k+1}(x_{k+1})\cdots\theta_n(x_n)} \\
&= \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \frac{1}{q^{n-k}} \sum_{x_{k+1},\ldots,x_n} q^{n-k} p_{k+1}(x_{k+1})\cdots p_n(x_n) \\
&= \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \left(\sum_{x_{k+1}} p_{k+1}(x_{k+1})\right)\cdots\left(\sum_{x_n} p_n(x_n)\right) \\
&= \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})}.
\end{aligned}$$

□

Note that (15) is exactly what we want, since to compute the non-uniform Fourier coefficient of $w_{\boldsymbol{z}}\tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{z}')$ at $\boldsymbol{a}$, we need to multiply the projected distribution by $\theta_{[k]}(\boldsymbol{z}')$. Specifically, denote $\mathcal{F}(\tilde{\mathscr{W}}_{\boldsymbol{a}}, \boldsymbol{a})$ (as defined in (9)) by $\mathscr{W}'$ and use (15), then we have, for every $\boldsymbol{z}' \in \Sigma^k$,

$$\begin{aligned}
\mathscr{W}'(\boldsymbol{z}') &= \left(\tilde{\mathscr{W}}_{\boldsymbol{a}}\right)_{[k]}(\boldsymbol{z}')\theta_{[k]}(\boldsymbol{z}') \\
&= \sum_{\boldsymbol{z}} w_{\boldsymbol{z}} \left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[k]}(\boldsymbol{z}')\theta_{[k]}(\boldsymbol{z}') \\
&= \sum_{\boldsymbol{z}} w_{\boldsymbol{z}} \frac{\delta(\boldsymbol{z}', \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})}\theta_{[k]}(\boldsymbol{z}') \\
&= w_{\boldsymbol{z}'}.
\end{aligned}$$

It follows that $\mathscr{W}' = \mathscr{W}_{\boldsymbol{a}}$ by (13). Therefore for any vector $\boldsymbol{b}$ whose support is $[k]$, we have $\hat{\tilde{\mathscr{W}}}{}^{\text{non}}_{\boldsymbol{a}}(\boldsymbol{b}) = \hat{\mathscr{W}}_{\boldsymbol{a}}(\boldsymbol{b})$. In particular, by (10$'$) and (11), $\hat{\tilde{\mathscr{W}}}{}^{\text{non}}_{\boldsymbol{a}}(\ell'\boldsymbol{a}) = -\hat{D}^{\text{non}}(\ell'\boldsymbol{a})$ for every vector $\ell'\boldsymbol{a}$ such that $\mathrm{Supp}(\ell'\boldsymbol{a}) = \mathrm{supp}(\boldsymbol{a})$ and $\hat{\tilde{\mathscr{W}}}{}^{\text{non}}_{\boldsymbol{a}}(\boldsymbol{b}) = 0$ for every vector $\boldsymbol{b}$ which is not a sibling of $\boldsymbol{a}$ and satisfies $\mathrm{supp}(\boldsymbol{b}) = \mathrm{supp}(\boldsymbol{a})$. This proves the first part of the Lemma.

Next we consider the non-uniform Fourier coefficient of $\tilde{\mathscr{U}}_{\boldsymbol{a}}$ at $\boldsymbol{b}$, where $\mathrm{supp}(\boldsymbol{b}) \not\subseteq [k]$. Without loss of generality, assume that $\mathrm{supp}(\boldsymbol{b}) = \{\ell+1, \ldots, k, k+1, \ldots, k+m\}$, where $\ell \leq k-1$ and $m \geq 1$. Now consider the non-uniform Fourier coefficient of any atomic uniform distribution $\tilde{\mathscr{U}}_{\boldsymbol{z}}$. By the form of $\tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x})$ we just used,

$$
\begin{aligned}
&\left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{\mathrm{supp}(\boldsymbol{b})}(x_{\ell+1}, \ldots, x_{k+m}) = \left(\tilde{\mathscr{U}}_{\boldsymbol{z}}\right)_{[\ell+1,k+m]}(x_{\ell+1}, \ldots, x_{k+m}) \\
&= \sum_{x_1,\ldots,x_\ell} \sum_{x_{k+m+1},\ldots,x_n} \tilde{\mathscr{U}}_{\boldsymbol{z}}(\boldsymbol{x}) \\
&= \frac{1}{q^{n-k}} \sum_{x_1,\ldots,x_\ell} \frac{\delta(\boldsymbol{x}_{[k]}, \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})} \sum_{x_{k+m+1},\ldots,x_n} \frac{1}{\theta_{k+1}(x_{k+1})\cdots\theta_{k+m}(x_{k+m})\theta_{k+m+1}(x_{k+m+1})\cdots\theta_n(x_n)} \\
&= \frac{1}{q^{n-k}\theta_{[k]}(\boldsymbol{z})} \delta(\boldsymbol{x}_{[\ell+1,k]}, \boldsymbol{z}_{[\ell+1,k]}) \frac{q^{n-k-m}}{\theta_{k+1}(x_{k+1})\cdots\theta_{k+m}(x_{k+m})} \left(\sum_{x_{k+m+1}} p_{k+m+1}(x_{k+m+1})\right) \cdots \left(\sum_{x_n} p_n(x_n)\right) \\
&= \frac{1}{q^m \theta_{[k]}(\boldsymbol{z})\theta_{k+1}(x_{k+1})\cdots\theta_{k+m}(x_{k+m})} \delta(\boldsymbol{x}_{[\ell+1,k]}, \boldsymbol{z}_{[\ell+1,k]}).
\end{aligned}
$$

Therefore, after applying the compressing/stretching transformation, $\tilde{\mathscr{U}}_{\boldsymbol{z}}$ is uniform over $[k+1, k+m]$. Consequently, its non-uniform Fourier coefficient at $\boldsymbol{b}$ is

$$
\begin{aligned}
\hat{\tilde{\mathscr{U}}}_{\boldsymbol{z}}^{\mathrm{non}}(\boldsymbol{b}) &= \sum_{x_{\ell+1},\ldots,x_{k+m}} \frac{\delta(\boldsymbol{x}_{[\ell+1,k]}, \boldsymbol{z}_{[\ell+1,k]})\theta_{\ell+1}(x_{\ell+1})\cdots\theta_{k+m}(x_{k+m})}{q^m \theta_{[k]}(\boldsymbol{z})\theta_{k+1}(x_{k+1})\cdots\theta_{k+m}(x_{k+m})} e^{\frac{2\pi i}{q}(b_{\ell+1}x_{\ell+1}+\cdots+b_{k+m}x_{k+m})} \\
&= \frac{e^{\frac{2\pi i}{q}(b_{\ell+1}z_{\ell+1}+\cdots+b_k z_k)}}{q^m \theta_1(z_1)\cdots\theta_\ell(z_\ell)} \sum_{x_{k+1},\ldots,x_{k+m}} e^{\frac{2\pi i}{q}(b_{k+1}x_{k+1}+\cdots+b_{k+m}x_{k+m})} \\
&= \frac{e^{\frac{2\pi i}{q}(b_{\ell+1}z_{\ell+1}+\cdots+b_k z_k)}}{q^m \theta_1(z_1)\cdots\theta_\ell(z_\ell)} \sum_{x_{k+2},\ldots,x_{k+m}} e^{\frac{2\pi i}{q}(b_{k+2}x_{k+2}+\cdots+b_{k+m}x_{k+m})} \sum_{x_{k+1}} e^{\frac{2\pi i}{q}(b_{k+1}x_{k+1})} \\
&= 0,
\end{aligned}
$$

where the last step follows from Fact 2.1 since b_{k+1} is non-zero. This proves the second part of the Lemma.

By (15′) in Claim 4.6 the total weight added by a transformed atomic uniform distribution is $\frac{w_{\boldsymbol{z}}}{\theta_{[k]}(\boldsymbol{z})} \leq \gamma_k w_{\boldsymbol{z}}$. Adding all the atomic uniform distributions together and using (14) proves the third part of the Lemma.

For the last part, assume $\mathrm{supp}(\boldsymbol{c}) = T \subset [k]$. Now consider the contribution of a transformed atomic uniform distribution $w_{\boldsymbol{z}}\tilde{\mathscr{U}}_{\boldsymbol{z}}$ to the non-uniform Fourier coefficient at $\boldsymbol{c}$:

$$
\begin{aligned}
\mathcal{F}(w_{\boldsymbol{z}}\tilde{\mathscr{U}}_{\boldsymbol{z}}, \boldsymbol{c})(\boldsymbol{z}'_T) &= w_{\boldsymbol{z}} \left(\frac{\delta(\boldsymbol{z}', \boldsymbol{z})}{\theta_{[k]}(\boldsymbol{z})}\right)_T \theta_T(\boldsymbol{z}'_T) \\
&= w_{\boldsymbol{z}} \frac{\theta_T(\boldsymbol{z}'_T)}{\theta_{[k]}(\boldsymbol{z})} \delta(\boldsymbol{z}'_T, \boldsymbol{z}_T).
\end{aligned}
$$

We can upper bound the non-uniform Fourier coefficient at $\boldsymbol{c}$ by

$$\begin{aligned}
\left|\hat{\mathcal{F}}(w_{\boldsymbol{z}}\tilde{\mathscr{U}}_{\boldsymbol{z}}, \boldsymbol{c})(\boldsymbol{c})\right| &\leq \left|\sum_{\boldsymbol{z}'_T} \mathcal{F}(w_{\boldsymbol{z}}\tilde{\mathscr{U}}_{\boldsymbol{z}}, \boldsymbol{c})(\boldsymbol{z}'_T)\right| \\
&= w_{\boldsymbol{z}} \frac{\theta_T(\boldsymbol{z}'_T)}{\theta_{[k]}(\boldsymbol{z})} && \text{(since } \mathcal{F}(w_{\boldsymbol{z}}\tilde{\mathscr{U}}_{\boldsymbol{z}}, \boldsymbol{c}) \text{ is non-negative)} \\
&\leq w_{\boldsymbol{z}} \frac{1}{\theta_{[k]}(\boldsymbol{z})} && \text{(since } \theta_T(\boldsymbol{z}'_T) \leq 1\text{)} \\
&\leq \gamma_k w_{\boldsymbol{z}}.
\end{aligned}$$

Finally we add up the weight of all transformed atomic uniform distributions in $\tilde{\mathscr{W}}$ and apply (14) to complete the last part of the Lemma and thus finish the proof of Lemma 4.5. □

Now we can, for any prime vector $\boldsymbol{a}$ whose support set is of size k, mix D with $\tilde{\mathscr{U}}_{\boldsymbol{a}}$ to zero-out the non-uniform Fourier coefficient at $\boldsymbol{a}$ and all its sibling vectors whose supports are the same as $\boldsymbol{a}$. By Lemma 4.5 the added small-weight distribution can only increase the magnitudes of non-uniform Fourier coefficients at vectors whose supports are strict subsets of the support of $\boldsymbol{a}$. Keep doing this for all the prime vectors at level k, the at the end, we obtain a distribution whose non-uniform Fourier coefficients at level k are all zero. We then recompute the non-uniform Fourier coefficients of the new distribution and repeat this process for vectors whose support sets are of size $k-1$. We keep doing this until we zero-out all the non-uniform Fourier coefficients at vectors of weight 1, at which point we finally obtain a non-uniform k-wise independent distribution.

Theorem 4.7. *Let D be a distribution over Σ^n, then*

$$\Delta(D, \mathcal{D}_{\text{kwi}}) \leq O\left(n^{\frac{k^2-k+2}{2}} q^{k(k+1)}\right) \max_{\boldsymbol{a}:0<\text{wt}(\boldsymbol{a})\leq k} \left|\hat{D}^{\text{non}}(\boldsymbol{a})\right|.$$

Proof. First observe that, for every $1 \leq i \leq n$ and every $z \in \Sigma$, $\frac{1}{\theta_i(z)} = qp_i(z) < q$, so $\gamma_j < q^j$, for every $1 \leq j \leq k$. We consider the zeroing-out processes in $k+1$ stages. At Stage 0 we have the initial distribution. At stage 1, we zero-out all the level k non-uniform Fourier coefficients. Doing this may increase the non-uniform Fourier coefficients at lower levels. Therefore we re-compute all the non-uniform Fourier coefficients at level below k and then proceed to zero-out the (new) non-uniform Fourier coefficients at level $k-1$. Finally at stage k, we zero-out all the level 1 non-uniform Fourier coefficients and obtain a non-uniform k-wise independent distribution.

Let $f_{\max} = \max_{0<\text{wt}(\boldsymbol{a})\leq k} \left|\hat{D}^{\text{non}}(\boldsymbol{a})\right|$. For simplicity, we will normalize by $f_{\max}$ every bound on the magnitudes of the non-uniform Fourier coefficients as well as every bound on the total weight added in each stage. That is, we divide all the quantities by $f_{\max}$ and work with the ratios.

Let $f^{(j)}$ denote the maximum magnitude, divided by $f_{\max}$, of any non-uniform Fourier coefficient that have not been corrected at stage j. That is, non-uniform Fourier coefficients at level i for $1 \leq i \leq k-j$. Clearly $f^{(0)} = 1$.

Now we consider the zeroing-out process at stage 1. There are $\binom{n}{k}(q-1)^k$ vectors at level k, and by part(3) of Lemma 4.5, correcting the non-uniform Fourier coefficient at each vector adds a weight at most $\gamma_k(q-1)f^{(0)}$. Therefore, the total weight added at stage 1 is at most $\binom{n}{k}(q-1)^k\gamma_k(q-1)f^{(0)} = O(n^kq^{2k+1})$. Next we calculate $f^{(1)}$, that is, the maximal magnitude of the remaining non-uniform Fourier coefficients.

For any vector $\boldsymbol{c}$ at level i, $1 \le i \le k-1$, there are $\binom{n-i}{k-i}(q-1)^{k-i}$ vectors at level k whose support sets are supersets of the support of $\boldsymbol{c}$. By part(4) of Lemma 4.5, zeroing-out the non-uniform Fourier coefficient at each such vector may increase $\left|\hat{D}^{\mathrm{non}}(\boldsymbol{c})\right|$ by $\gamma_k(q-1)f^{(0)}$. Therefore the magnitude of the non-uniform Fourier coefficient at $\boldsymbol{c}$ is at most

$$f^{(0)} + \binom{n-i}{k-i}(q-1)^{k-i}\gamma_k(q-1)f^{(0)} = O\left(n^{k-i}q^{2k-i+1}\right).$$

Clearly the worst case happens when $i=1$ and we thus have $f^{(1)} \le O\left(n^{k-1}q^{2k}\right)$.

In general it is easy to see that, at every stage, the maximum magnitude increases of the non-uniform Fourier coefficients always occur at level 1. Then at stage j, we are zeroing-out the non-uniform Fourier coefficients at level $k-j+1$. For every vector at level 1, there are $\binom{n-1}{k-j}(q-1)^{k-j}$ vectors at level $k-j+1$ whose support sets are supersets of that of the level-1 vector, and the increase in non-uniform Fourier coefficient magnitude cause by each such level-$(k-j+1)$ vector is at most $\gamma_{k-j+1}(q-1)f^{(j-1)}$, we thus have

$$f^{(j)} \le \binom{n-1}{k-j}(q-1)^{k-j}\gamma_{k-j+1}(q-1)f^{(j-1)} \le O\left(n^{k-j}q^{2(k-j+1)}\right)f^{(j-1)}, \qquad \text{for } 1 \le j \le k-1.$$

This in turn gives

$$f^{(j)} \le O\left(n^{\frac{j(2k-j-1)}{2}}q^{j(2k-j+1)}\right), \qquad \text{for } 1 \le j \le k-1.$$

It is easy to check that the weight added at stage k dominates the weight added at all previous stages, therefore the total weight added during all $k+1$ stages is at most

$$O\left(\binom{n}{1}(q-1)\gamma_1\right)f^{(k-1)} \le O\left(n^{\frac{k^2-k+2}{2}}q^{k(k+1)}\right). \qquad \square$$

4.3 Testing Algorithm and its Analysis

In Fig. 3, we give an outline of the algorithm for testing non-uniform k-wise independence when all the marginal probabilities $p_i(z)$ are assumed to be known.[12] The analysis of the testing algorithm is very much the same[13] as that presented in Section 3.2.4, we leave the details to interested readers.

Acknowledgments

We would like to thank Tali Kaufman for useful discussions and suggestions, Elchanan Mossel for his enthusiasm in this problem and a helpful conversation and Per Austrin for correspondence on constructing orthogonal real functions. We are grateful to anonymous referees for pointing out a critical error in an earlier version of this paper and suggesting the approach presented in Appendix A.

[12] If we assume that these probabilities are all bounded away from 0 or 1, then they can also be estimated from a small number of samples drawn independently from the distribution.

[13] One major difference is that we need to use the following simple fact to show completeness of the testing algorithm: if $\Delta(D, D_{\mathrm{kwi}}) \le \delta$, then $|\hat{D}^{\mathrm{non}}(\boldsymbol{a})| \le q\gamma_k\delta$ for all non-zero vectors $\boldsymbol{a}$ of weight at most k.

```
Testing Non-uniform KWI$(D, k, q, \epsilon)$
  1. Sample $D$ uniformly and independently $M$ times
  2. Use the samples to estimate, for each non-zero vector $\boldsymbol{a}$ of weight at most $k$ and each
     $\boldsymbol{z} \in \Sigma^{\mathrm{supp}(\boldsymbol{a})}$, $D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$, where $\mathrm{supp}(\boldsymbol{a})$ is the support set of $\boldsymbol{a}$
       • Compute $D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) = \theta_S(\boldsymbol{z}) D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$
       • Compute $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) \overset{\mathrm{def}}{=} \hat{D}'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{a}) = \sum_{\boldsymbol{z}} D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) e^{\frac{2\pi i}{q} \boldsymbol{a} \cdot \boldsymbol{z}}$
  3. If $\max_{\boldsymbol{a}} \left| \hat{D}^{\mathrm{non}}(\boldsymbol{a}) \right| \leq \delta$ return "Yes"; else return "No"
```

Figure 3: Algorithm for testing non-uniform k-wise independence.

References

[1] N. Alon, A. Andoni, T. Kaufman, K. Matulef, R. Rubinfeld, and N. Xie. Testing k-wise and almost k-wise independence. In *Proc. 39th Annual ACM Symposium on the Theory of Computing*, pages 496–505, 2007.

[2] N. Alon, L. Babai, and A. Itai. A fast and simple randomized algorithm for the maximal independent set problem. *Journal 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. *Information Processing Letters*, 88:107–110, 2003.

[5] P. Austrin. *Conditional inapproximability and limited independence*. PhD thesis, KTH - Royal Institute of Technology, 2008. Available at http://www.csc.kth.se/ austrin/papers/thesis.pdf.

[6] Y. Azar, J. Naor, and R. Motwani. Approximating probability distributions using small sample spaces. *Combinatorica*, 18:151–171, 1998.

[7] T. Batu, E. Fischer, 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] C. Bertram-Kretzberg and H. Lefmann. MOD_p-tests, almost independence and small probability spaces. *Random Structures and Algorithms*, 16(4):293–313, 2000.

[11] E. Blais. Testing juntas nearly optimally. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 151–158, 2009.

[12] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1993. Earlier version in STOC'90.

[13] B. Chor and O. Goldreich. On the power of two-point based sampling. *Journal of Complexity*, 5(1):96–106, 1989.

[14] A. Czumaj and C. Sohler. Sublinear-time algorithms. *Bulletin of the European Association for Theoretical Computer Science*, 89:23–47, 2006.

[15] I. Diakonikolas, H. Lee, K. Matulef, K. Onak, R. Rubinfeld, R. Servedio, and A. Wan. Testing for concise representations. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 549–558, 2007.

[16] K. Efremenko. 3-query locally decodable codes of subexponential length. In *Proc. 41st Annual ACM Symposium on the Theory of Computing*, pages 39–44, 2009.

[17] G. Even, O. Goldreich, M. Luby, N. Nisan, and B. Velickovic. Efficient approximation of product distributions. *Random Structures and Algorithms*, 13(1):1–16, 1998. Earlier version in STOC'92.

[18] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the European Association for Theoretical Computer Science*, 75, 2001.

[19] O. Goldreich, S. Goldwaser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45:653–750, 1998.

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

[21] V. Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit ramsey graphs. *Combinatorica*, 20(1):71–86, 2000.

[22] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 5th edition, 1980.

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

[24] H. Karloff and Y. Mansour. On construction of k-wise independent random variables. In *Proc. 26th Annual ACM Symposium on the Theory of Computing*, pages 564–573, 1994.

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

[26] D. Koller and N. Megiddo. Constructing small sample spaces satisfying given constraints. In *Proc. 25th Annual ACM Symposium on the Theory of Computing*, pages 268–277, 1993.

[27] R. Kumar and R. Rubinfeld. Sublinear time algorithms. *SIGACT News*, 34:57–67, 2003.

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

[29] M. Luby. Removing randomness in parallel computation without a processor penalty. In *Proc. 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 162–173, 1988.

[30] E. Mossel. Gaussian bounds for noise correlation of functions and tight analysis of long codes. In *Proc. 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 156–165, 2008.

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

[32] C. P. Neuman and D. I. Schonbach. Discrete (Legendre) orthogonal polynomials - A survey. *International Journal for Numerical Methods in Engineering*, 8:743–770, 1974.

[33] A. F. Nikiforov, S. K. Suslov, and V. B. Uvarov. *Classical Orthogonal Polynomials of a Discrete Variable*. Springer-Verlag, 1991.

[34] S. Raskhodnikova, D. Ron, A. Shpilka, and A. Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 559–569, 2007.

[35] D. Ron. Property testing (a tutorial). In P.M. Pardalos, S. Rajasekaran, J. Reif, and J.D.P. Rolim, editors, *Handbook of Randomized Computing*, pages 597–649. Kluwer Academic Publishers, 2001.

[36] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25:252–271, 1996.

[37] A. Sahai and S. Vadhan. A complete problem for statistical zero knowledge. *Journal of the ACM*, 50(2):1–54, 2003.

[38] J.R. Silvester. Determinants of block matrices. *Maths Gazette*, 84:460–467, 2000.

[39] H. J. S. Smith. On systems of linear indeterminate equations and congruences. *Phil. Trans. Royal Soc. London*, A151:293–326, 1861.

[40] D. Štefankovič. Fourier transform in computer science. Master's thesis, University of Chicago, 2000.

[41] A. Terras. *Fourier Analysis on Finite Groups and Applications*. Cambridge University Press, 1999.

[42] S. Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM*, 55(1):1–16, 2008.

A An alternative proof of Theorem 3.17

In this Appendix we give an alternative and conceptually simple proof of Theorem 3.17. The bound we prove here is somewhat weaker that stated in Theorem 3.17. The basic idea is to apply the "cut in the Fourier space and then mend in the function space" approach in [1] to Fourier expansions with discrete orthogonal real polynomials as the basis functions.

A.1 Generalized Fourier series

The discrete Fourier transform reviewed in Section 2 can be generalized to decompositions over any orthonormal basis of an inner product space. In particular, for the discrete function space $\mathbb{R}^{\{0,\ldots,q-1\}}$, any orthonormal basis of real functions $\{g_0(x), \ldots, g_{q-1}(x)\}$ with $g_0(x) = 1$ for every x (the identity function) can be used in place of the standard Fourier basis $\{1, e^{\frac{2\pi i x}{q}}, \ldots, e^{\frac{2\pi i(q-1)x}{q}}\}$. In general, such a basis of functions may be constructed by the Gram-Schmidt process. For concreteness, we present an explicit construction of one such function basis by means of *discrete Legendre orthogonal polynomials* [32], a special case of Hahn polynomials. An extensive treatment of discrete orthogonal polynomials may be found in [33]. We remark that our proof works for any set of complete orthonormal basis of real functions as long as one of the basis functions is the identity function.

For $n \geq 0$, we write $(x)_n := x(x-1)\cdots(x-n+1)$ for the falling factorial. For any integer $q \geq 2$, the discrete Legendre orthogonal polynomials, $\{P_a(x;q)\}_{a=0}^{q-1}$, are defined as

$$\begin{aligned} P_a(x;q) &= \sum_{j=0}^{a} (-1)^j \binom{a}{j} \binom{a+j}{j} \frac{(x)_j}{(q-1)_j}, \\ P_a(0;q) &= 1, \text{ for all } a = 0, 1, \ldots, q-1. \end{aligned}$$

These polynomials satisfy the following orthogonal properties (see, e.g., [32]):

$$\sum_{x=0}^{q-1} P_a(x;q) P_b(x;q) = \begin{cases} 0, & \text{if } a \neq b, \\ \frac{1}{2a+1} \frac{(q+a)_{a+1}}{(q-1)_a}, & \text{if } a = b. \end{cases}$$

Now we define [14] a complete set of orthonormal functions $\{\chi_a^{\mathrm{OF}}(x)\}_{a=0}^{q-1}$ by

$$\chi_a^{\mathrm{OF}}(x) = \sqrt{\frac{(2a+1)(q)_{a+1}}{(q+a)_{a+1}}} P_a(x;q),$$

then they form a complete basis for the real functions space over $\{0, 1, \ldots, q-1\}$ and satisfy the orthogonality relation

$$\sum_{x=0}^{q-1} \chi_a^{\mathrm{OF}}(x) \chi_b^{\mathrm{OF}}(x) = \begin{cases} 0, & \text{if } a \neq b, \\ q, & \text{if } a = b. \end{cases}$$

Any real function $f : \{0, 1, \ldots, q-1\} \to \mathbb{R}$ can thus be expanded in terms of these basis functions

$$f(x) = \frac{1}{q} \sum_{a=0}^{q-1} \hat{f}^{\mathrm{OF}}(a) \chi_a^{\mathrm{OF}}(x),$$

with the inversion formula

$$\hat{f}^{\mathrm{OF}}(a) = \sum_{x=0}^{q-1} f(x) \chi_a^{\mathrm{OF}}(x).$$

[14] We add the superscript OF (denoting *orthogonal functions*) to distinguish them from the standard Fourier basis functions over $\{0,1\}^n$.

The expansion coefficients $\{\hat{f}^{\rm OF}(a)\}$ are called the *generalized Fouier coefficients*.

Generalizing this expansion to real functions over higher dimensional spaces is straightforward. Let $n \geq 1$ be an integer and let $f : \{0, 1, \ldots, q-1\}^n \to \mathbb{R}$. The generalized Fourier expansion of f is simply

$$f(\boldsymbol{x}) = \frac{1}{q^n} \sum_{\boldsymbol{a}} \hat{f}^{\rm OF}(\boldsymbol{a}) \chi_{\boldsymbol{a}}^{\rm OF}(\boldsymbol{x}),$$

with the inversion formula

$$\hat{f}^{\rm OF}(\boldsymbol{a}) = \sum_{\boldsymbol{x}} f(\boldsymbol{x}) \chi_{\boldsymbol{a}}^{\rm OF}(\boldsymbol{x}),$$

where $\chi_{\boldsymbol{a}}^{\rm OF}(\boldsymbol{x}) \overset{\rm def}{=} \prod_{i=1}^n \chi_{a_i}^{\rm OF}(x_i)$ and satisfy the orthogonality relation $\sum_{\boldsymbol{x}} \chi_{\boldsymbol{a}}^{\rm OF}(\boldsymbol{x}) \chi_{\boldsymbol{b}}^{\rm OF}(\boldsymbol{x}) = \begin{cases} 0, & \text{if } \boldsymbol{a} \neq \boldsymbol{b}, \\ q^n, & \text{if } \boldsymbol{a} = \boldsymbol{b}. \end{cases}$

A direct consequence of the orthogonality of the basis functions $\{\chi_{\boldsymbol{a}}^{\rm OF}(\boldsymbol{x})\}$ is the Parseval equality

$$\sum_{\boldsymbol{x}} f^2(\boldsymbol{x}) = \frac{1}{q^n} \sum_{\boldsymbol{a}} \hat{f}^{\rm OF}(\boldsymbol{a})^2.$$

It is easy to check that the following characterizations of uniform distributions and k-wise independent distributions over $\{0, 1, \ldots, q-1\}^n$ in terms of the generalized Fourier coefficients. The proofs follow directly from the orthogonality of $\{\chi_{\boldsymbol{a}}^{\rm OF}(\boldsymbol{x})\}$ and definition of k-wise independence, therefore we omit them here.

Proposition A.1. *Let D be a distribution over $\{0, 1, \ldots, q-1\}^n$. Then D is the uniform distribution if and only if for any non-zero vector $\boldsymbol{a} \in \{0, 1, \ldots, q-1\}^n$, $\hat{D}^{\rm OF}(\boldsymbol{a}) = 0$.*

Corollary A.2. *A distribution D over $\{0, 1, \ldots, q-1\}^n$ is k-wise independent if and only if, for all non-zero vectors $\boldsymbol{a}$ of weight at most k, $\hat{D}^{\rm OF}(\boldsymbol{a}) = 0$.*

A.2 Another Proof of Theorem 3.17

The basic idea of [1] is the following. Given a distribution D, we first operate in the Fourier space to construct a "pseudo-distribution" D_1 by setting all the first k-level generalized Fourier coefficients (except for the trivial Fourier coefficient) to zero. All other generalized Fourier Fourier coefficients of D_1 are the same as D. Generally speaking, D_1 is not going to be a distribution because it may assume negative values at some points. We then correct all these negative points by mixing D_1 with the uniform distribution of some appropriate weight. That is, we set $D' = \frac{1}{1+w} D_1 + \frac{w}{1+w} U$, where U is the uniform distribution and $w > 0$ is the weight of the uniform distribution. After such an operation, since the uniform distribution clearly has all its first k-level generalized Fourier coefficients equal to zero, we maintain that all the first k-level generalized Fourier coefficients of D' are still zero; on the other hand, we increase the weights at negative points so that they now assume non-negative values in D'. Bounding the total statistical distance between D and D' then offers an upper bound on the distance between D and k-wise independence.

Let $D : \{0, 1, \ldots, q-1\}^n \to \mathbb{R}^{\geq 0}$ be a distribution, that is, $D(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$ and $\sum_{\boldsymbol{x}} D(\boldsymbol{x}) = 1$. First we define a real function $D_1 : \{0, 1, \ldots, q-1\}^n \to \mathbb{R}$ by explicitly specifying all its generalized Fourier coefficients:

$$\hat{D}_1^{\rm OF}(\boldsymbol{a}) = \begin{cases} 0, & \text{if } 0 < \mathrm{wt}(\boldsymbol{a}) \leq k \\ \hat{D}^{\rm OF}(\boldsymbol{a}), & \text{otherwise.} \end{cases}$$

We call D_1 a "pseudo-distribution" because D_1 may assume negative values at some points in the domain, which are called the *holes* in D_1. Note that, since $\hat{D_1}^{\mathrm{OF}}(\mathbf{0}) = \hat{D}^{\mathrm{OF}}(\mathbf{0}) = 1$, we have $\sum_{\boldsymbol{x}} D_1(\boldsymbol{x}) = 1$. So the only difference between D_1 and a distribution is these holes. The following lemma bounds the maximum depth of the holes in D_1.

Lemma A.3. *Let h be the maximum depth of the holes in D_1, then*

$$h \leq \frac{q^{k/2}}{q^n} \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|.$$

Proof. First observe that $|\chi_a^{\mathrm{OF}}(x)| \leq \sqrt{q}$ for every $x \in \{0,1,\ldots,q-1\}$ and every $0 \leq a \leq q-1$, due to the fact that $\sum_{x=0}^{q-1} |\chi_a^{\mathrm{OF}}(x)|^2 = q$. It follows that $|\chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x})| \leq q^{k/2}$ if the weight of $\boldsymbol{a}$ is at most k. Now notice that, since $D(\boldsymbol{x}) \geq 0$ for every $\boldsymbol{x}$ in the domain and D_1 is obtained by cutting off all the first k level generalized Fourier coefficients of D, by linearity of the generalized Fourier expansion, for all $\boldsymbol{x}$ with $D_1(\boldsymbol{x}) < 0$,

$$\begin{aligned} |D_1(\boldsymbol{x})| &= \left| \frac{1}{q^n} \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} \hat{D}^{\mathrm{OF}}(\boldsymbol{a}) \chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x}) - D(\boldsymbol{x}) \right| \\ &\leq \left| \frac{1}{q^n} \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} \hat{D}^{\mathrm{OF}}(\boldsymbol{a}) \chi_{\boldsymbol{a}}^{\mathrm{OF}}(\boldsymbol{x}) \right| \\ &\leq \frac{q^{k/2}}{q^n} \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|. \end{aligned}$$

□

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

Lemma A.4 ([1]). *Let f be a real function defined over $\{0,1,\ldots,q-1\}^n$ such that $\sum_{\boldsymbol{x}} f(\boldsymbol{x}) = 1$. Let $D_1, \ldots, D_\ell$ be distributions over the same domain and suppose there exist non-negative real numbers $w_1, \ldots, w_\ell$ such that $D' \stackrel{\mathrm{def}}{=} \frac{1}{1+\sum_{i=1}^{\ell} w_i}(f + \sum_{i=1}^{\ell} w_i D_i)$ is non-negative for all $\boldsymbol{x} \in \{0,1,\ldots,q-1\}^n$. Then $\sum_{\boldsymbol{x}} |f(\boldsymbol{x}) - D'(\boldsymbol{x})| \leq 2\sum_{i=1}^{\ell} w_i$.*

Now we can mix D_1 with a uniform distribution over $\{0,1,\ldots,q-1\}^n$ of weight $q^n h$ to obtain a distribution D', that is,

$$D' \stackrel{\mathrm{def}}{=} \frac{1}{1+q^n h} D_1 + \frac{q^n h}{1+q^n h} U,$$

where U is the uniform distribution over $\{0,1,\ldots,q-1\}^n$. Then D' is non-negative at every point in the domain and D' has all its first k-level generalized Fourier coefficients equal to zero. Thus D' is a k-wise independent distribution by Corollary A.2. Furthermore, by Lemma A.4,

$$\sum_{\boldsymbol{x}} |D_1(\boldsymbol{x}) - D'(\boldsymbol{x})| \leq 2q^n h \leq q^{k/2} \sum_{0<\mathrm{wt}(\boldsymbol{a})\leq k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|.$$

By Parseval equality, $\sum_{\boldsymbol{x}} |D(\boldsymbol{x}) - D_1(\boldsymbol{x})|^2 = \frac{1}{q^n} \sum_{0<\mathrm{wt}(\boldsymbol{a}) \le k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|^2$. Combining this with Cauchy-Schwarz inequality yields

$$\sum_{\boldsymbol{x}} |D(\boldsymbol{x}) - D_1(\boldsymbol{x})| \le \sqrt{\sum_{0<\mathrm{wt}(\boldsymbol{a}) \le k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|^2}.$$

Now by triangle inequality,

$$\begin{aligned}
\Delta(D, D') &= \frac{1}{2} \sum_{\boldsymbol{x}} |D(\boldsymbol{x}) - D'(\boldsymbol{x})| \\
&\le \frac{1}{2} \sum_{\boldsymbol{x}} |D(\boldsymbol{x}) - D_1(\boldsymbol{x})| + \frac{1}{2} \sum_{\boldsymbol{x}} |D_1(\boldsymbol{x}) - D'(\boldsymbol{x})| \\
&\le \frac{1}{2} \sqrt{\sum_{0<\mathrm{wt}(\boldsymbol{a}) \le k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|^2} + q^{k/2} \sum_{0<\mathrm{wt}(\boldsymbol{a}) \le k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})| \\
&= O(q^{k/2}) \sum_{0<\mathrm{wt}(\boldsymbol{a}) \le k} |\hat{D}^{\mathrm{OF}}(\boldsymbol{a})|.
\end{aligned}$$

We thus prove the following theorem

Theorem A.5. *Let D be a distribution over $\{0, 1, \ldots, q-1\}^n$, then*

$$\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \le O(q^{k/2}) \sum_{0<\mathrm{wt}(\boldsymbol{a}) \le k} \left| \hat{D}^{\mathrm{OF}}(\boldsymbol{a}) \right|. \tag{16}$$

In particular,

$$\Delta(D, \mathcal{D}_{\mathrm{kwi}}) \le O(q^{k/2}) M(n, k, q) \max_{0<\mathrm{wt}(\boldsymbol{a}) \le k} \left| \hat{D}^{\mathrm{OF}}(\boldsymbol{a}) \right|.$$

B New characterization of non-uniform k-wise independence

Recall that we define a set of non-uniform Fourier coefficients on all non-zero vectors of weight at most k as follows. Given a distribution D over Σ^n, let $p_i(z) = \Pr_D[X_i = z]$ be the marginal distribution at the i^{th} coordinate. Define $\theta_i(z) = \frac{1}{q p_i(z)}$ as the compressing/stretching factor, and if $S = \{i_1, \ldots, i_\ell\}$ and $\boldsymbol{z} = z_{i_1} \cdots z_{i_\ell}$, we write $\theta_S(\boldsymbol{z}) = \theta_{i_1}(z_{i_1}) \cdots \theta_{i_\ell}(z_{i_\ell})$. For any non-zero $\boldsymbol{a}$ with $\mathrm{wt}(\boldsymbol{a}) \le k$ and let $\mathrm{supp}(\boldsymbol{a})$ be its support set, then for any $\boldsymbol{z} \in \Sigma^{|\mathrm{supp}(\boldsymbol{a})|}$ define

$$D'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) = \theta_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z}) D_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{z})$$

to be the transformed "distribution" of the projection distribution $D_{\mathrm{supp}(\boldsymbol{a})}$. Then the non-uniform Fourier coefficient at $\boldsymbol{a}$ is

$$\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = \hat{D}'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{a}).$$

Our new characterization of non-uniform k-wise independent distributions is the following.

Theorem 4.3. *A distribution D over Σ^n is k-wise independent if and only if for every non-zero vector $\boldsymbol{a} \in \Sigma^k$ with $\mathrm{wt}(\boldsymbol{a}) \le k$, $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = 0$.*

Proof. Suppose D is a non-uniform k-wise independent distribution. Then it is easy to see that for any non-empty $T \subset [n]$ of size at most k (not just for subsets of sizes exactly k),

$$D_T(z_i : i \in T) = \prod_{i \in T} p_i(z_i).$$

Indeed, if $|T| = k$ then it follows directly from the definition of non-uniform k-wise independent distributions. If $|T| < k$, let $S \supset T$ be any index set of size k, then

$$\begin{aligned} D_T(z_i : i \in T) &= \sum_{z_j : j \in S \setminus T} D_S(z_\ell : \ell \in S) \\ &= \sum_{z_j : j \in S \setminus T} \prod_{\ell \in S} p_\ell(z_\ell) \\ &= \prod_{i \in T} p_i(z_i) \sum_{z_j : j \in S \setminus T} \prod_{j \in S \setminus T} p_j(z_j) \\ &= \prod_{i \in T} p_i(z_i) \prod_{j \in S \setminus T} \left(\sum_{z_j} p_j(z_j) \right) \\ &= \prod_{i \in T} p_i(z_i). \end{aligned}$$

Let $\boldsymbol{a}$ be any non-zero vector of weight $\ell \leq k$ and $\mathrm{supp}(\boldsymbol{a})$ be its support set. Now we show that $D'_{\mathrm{supp}(\boldsymbol{a})}$ is a uniform distribution and consequently all the non-uniform Fourier coefficients whose support set is $\mathrm{supp}(\boldsymbol{a})$ must be zero. By definition,

$$D'_{\mathrm{supp}(\boldsymbol{a})}(z_i : i \in \mathrm{supp}(\boldsymbol{a})) = D_{\mathrm{supp}(\boldsymbol{a})}(z_i : i \in \mathrm{supp}(\boldsymbol{a})) \prod_{i \in \mathrm{supp}(\boldsymbol{a})} \theta_i(z_i) = \prod_{i \in \mathrm{supp}(\boldsymbol{a})} p_i(z_i) \prod_{i \in \mathrm{supp}(\boldsymbol{a})} \frac{1}{q p_i(z_i)} = \frac{1}{q^\ell}$$

for every $(z_i : i \in \mathrm{supp}(\boldsymbol{a})) \in \{0, 1, \ldots, q-1\}^\ell$. Hence $\hat{D}^{\mathrm{non}}(\boldsymbol{a}) = \hat{D}'_{\mathrm{supp}(\boldsymbol{a})}(\boldsymbol{a}) = 0$ by Corollary 2.4.

The converse direction will follow directly from the following Lemma B.1 by setting $E = D_S$ in the statement. □

Lemma B.1. *Let $E : \Sigma^k \to \mathbb{R}^{\geq 0}$ be a distribution. For any index set $T \subseteq [k]$, let $E_T(\boldsymbol{z})$, $E'_T(\boldsymbol{z})$ and $\hat{E}^{\mathrm{non}}(\boldsymbol{a})$ be defined analogously to those for $D_T(\boldsymbol{z})$, $D'_T(\boldsymbol{z})$ and $\hat{D}^{\mathrm{non}}(\boldsymbol{a})$, respectively. If $\hat{E}^{\mathrm{non}}(\boldsymbol{a}) = 0$ for every non-zero vector $\boldsymbol{a}$, then E is a non-uniform independent distribution; in other words, $E'_{[k]}$ is the uniform distribution and consequently E is a product distribution.*

One may think of Lemma B.1 as the non-uniform version of Proposition 2.3.

Proof. For convenience we write $S = [k]$. Let T be a subset of S of size $k-1$, and without loss of generality, we assume that $T = \{1, \ldots, k-1\}$. We first observe the following relation between $E'_S(\boldsymbol{z})$ and $E'_T(\boldsymbol{z'})$.

$$\begin{aligned} E'_T(z_1, \ldots, z_{k-1}) &= E_T(z_1, \ldots, z_{k-1}) \theta_1(z_1) \cdots \theta_{k-1}(z_{k-1}) \\ &= \sum_{z_k} E_S(z_1, \ldots, z_{k-1}, z_k) \theta_1(z_1) \cdots \theta_{k-1}(z_{k-1}) \\ &= \sum_{z_k} \frac{1}{\theta_k(z_k)} E'_S(z_1, \ldots, z_k) = \sum_{z_k} q p_k(z_k) E'_S(z_1, \ldots, z_k). \end{aligned}$$

By induction, we have in general, for any $T \subset S$,

$$E'_T(z_i : i \in T) = \sum_{z_j : j \in S \setminus T} E'_S(z_1, \dots, z_k) \prod_{j \in S \setminus T} (qp_j(z_j)). \tag{17}$$

Next we use (17) to eliminate the intermediate projection distributions E'_T, therefore writing the non-uniform Fourier transform of distribution E'_S as a linear transform of $E'_S(\boldsymbol{z})$'s. Let $\boldsymbol{a}$ be a vector with T being its support set, then

$$\begin{aligned}
\hat{E}^{\text{non}}(\boldsymbol{a}) &= \hat{E}'_T(\boldsymbol{a}) \\
&= \sum_{z_i : i \in T} E'_T(z_i : i \in T) e^{\frac{2\pi i}{q} \sum_{i \in T} a_i z_i} \\
&= \sum_{z_i : i \in T} \sum_{z_j : j \in S \setminus T} E'_S(\boldsymbol{z}) e^{\frac{2\pi i}{q} \sum_{i \in T} a_i z_i} \prod_{j \in S \setminus T} (qp_j(z_j)) \\
&= \sum_{\boldsymbol{z} \in \Sigma^S} E'_S(\boldsymbol{z}) \prod_{i \in T} e^{\frac{2\pi i}{q} a_i z_i} \prod_{j \in S \setminus T} (qp_j(z_j)) \\
&= \sum_{\boldsymbol{z} \in \Sigma^S} E'_S(\boldsymbol{z}) \prod_{i \in \text{supp}(\boldsymbol{a})} e^{\frac{2\pi i}{q} a_i z_i} \prod_{j \in S \setminus \text{supp}(\boldsymbol{a})} (qp_j(z_j)). \qquad (18)
\end{aligned}$$

Define a q^k-dimensional column vector $\mathbf{E}'$ with entries $E'_S(\boldsymbol{z})$ (we will specify the ordering of the entries later). Similarly define the non-uniform Fourier coefficient column vector $\hat{\mathbf{E}}^{\text{non}}$ with entries $\hat{E}^{\text{non}}(\boldsymbol{a})$. Then we may write (18) more compactly as

$$\hat{\mathbf{E}}^{\text{non}} = \tilde{\mathbf{F}} \mathbf{E}'. \tag{19}$$

In what follows, we will show that $\tilde{\mathbf{F}}$ can be written nicely as a tensor product of k matrices. This in turn enables us to show the matrix is non-singular.

Let $\omega = e^{\frac{2\pi i}{q}}$ be a primitive q^{th} root of unity. The q-point discrete Fourier transform (DFT) matrix is given by

$$\mathbf{F} = \begin{bmatrix}
1 & 1 & 1 & 1 & \cdots & 1 \\
1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{q-1} \\
1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(q-1)} \\
1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(q-1)} \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
1 & \omega^{q-1} & \omega^{2(q-1)} & \omega^{3(q-1)} & \cdots & \omega^{(q-1)(q-1)}
\end{bmatrix}$$

Note that a DFT matrix is also a Vandermonde matrix and therefore $\det(\mathbf{F}) \neq 0$.

Definition B.2 (Tensor product of vectors and matrices)**.** *If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is a $p \times q$ matrix, then the tensor product (a.k.a. Kronecker product) $\mathbf{A} \otimes \mathbf{B}$ is the $mp \times nq$ block matrix*

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix}
a_{00}\mathbf{B} & \cdots & a_{0,n-1}\mathbf{B} \\
\vdots & \ddots & \vdots \\
a_{m-1,0}\mathbf{B} & \cdots & a_{m-1,n-1}\mathbf{B}
\end{bmatrix}$$

$$= \begin{bmatrix} a_{00}b_{00} & \cdots & a_{00}b_{0,q-1} & \cdots & \cdots & a_{0,n-1}b_{00} & \cdots & a_{0,n-1}b_{0,q-1} \\ \vdots & \ddots & \vdots & & & \vdots & \ddots & \vdots \\ a_{00}b_{p-1,0} & \cdots & a_{00}b_{p-1,q-1} & \cdots & \cdots & a_{0,n-1}b_{p-1,0} & \cdots & a_{0,n-1}b_{p-1,q-1} \\ \vdots & & \vdots & \ddots & & \vdots & & \vdots \\ \vdots & & \vdots & & \ddots & \vdots & & \vdots \\ a_{m-1,0}b_{00} & \cdots & a_{m-1,0}b_{0,q-1} & \cdots & \cdots & a_{m-1,n-1}b_{00} & \cdots & a_{m-1,n-1}b_{p-1,q-1} \\ \vdots & \ddots & \vdots & & & \vdots & \ddots & \vdots \\ a_{m-1,0}b_{p-1,0} & \cdots & a_{m-1,0}b_{p-1,q-1} & \cdots & \cdots & a_{m-1,n-1}b_{p-1,0} & \cdots & a_{m-1,n-1}b_{p-1,q-1} \end{bmatrix}$$

Let $\boldsymbol{a}$ be an m-dimensional column vector in $\mathbb{R}^m$ and $\boldsymbol{b}$ be a p-dimensional column vector. Then the tensor product $\boldsymbol{a} \otimes \boldsymbol{b}$ is an mp-dimensional column vector in $\mathbb{R}^{mp}$ and its entries are given by

$$\boldsymbol{a} \otimes \boldsymbol{b} = \begin{bmatrix} a_0 \\ \vdots \\ a_{m-1} \end{bmatrix} \otimes \begin{bmatrix} b_0 \\ \vdots \\ b_{p-1} \end{bmatrix} = \begin{bmatrix} a_0 b_0 \\ \vdots \\ a_0 b_{p-1} \\ \vdots \\ \vdots \\ a_{m-1} b_0 \\ \vdots \\ a_{m-1} b_{p-1} \end{bmatrix}$$

Let $q \geq 2$ be an integer. The *q-ary representation* of a natural number r is an ordered tuple $(b_k, \ldots, b_1, b_0)$ such that $0 \leq b_i \leq q-1$ for every $0 \leq i \leq k$ and $r = b_0 + b_1 \cdot q + \cdots b_k \cdot q^k$. The following simple while useful fact about the tensor product of matrices can be proved easily by induction on the number of matrices in the product.

Fact B.3. *Let $F^{(1)}, \ldots, F^{(k)}$ be a set of k $q \times q$ matrices where the $(i,j)^{th}$ entry of $F^{(\ell)}$ is denoted by $F^{(\ell)}_{i,j}$, $0 \leq i, j \leq q-1$. Let $G = F^{(1)} \otimes \cdots \otimes F^{(k)}$. For $0 \leq I, J \leq q^k - 1$, let their q-ary representations be $I = (i_1, \ldots, i_k)$ and $J = (j_1, \ldots, j_k)$, respectively. Then*

$$G_{I,J} = F^{(1)}_{i_1,j_1} \cdots F^{(k)}_{i_k,j_k}.$$

Let $\mathbf{E}$ be the column vector with entries being the values of distribution E at all points. Similarly let $\hat{\mathbf{E}}$ be the column vector of E's Fourier transform. If we arrange the entries of $\mathbf{E}$ and $\hat{\mathbf{E}}$ in increasing order, then the one-dimensional (uniform) Fourier transform can be written in the matrix multiplication form as

$$\hat{\mathbf{E}} = \begin{bmatrix} \hat{E}(0) \\ \vdots \\ \hat{E}(q-1) \end{bmatrix} = \mathbf{F} \begin{bmatrix} E(0) \\ \vdots \\ E(q-1) \end{bmatrix} = \mathbf{F}\mathbf{E}. \tag{20}$$

We may view the k-dimensional point $(x_1, \ldots, x_k)$ in $E(x_1, \ldots, x_k)$ as the representation of a natural number X in q-ary representation: $X = x_1 \cdot q^{k-1} + \cdots + x_{k-1} \cdot q + x_k$. Then this provides a *natural ordering* of the entries in the column vector of $\mathbf{E}$ in a k-dimension space. Using tensor product and arranging the entries in $\mathbf{E}$ and $\hat{\mathbf{E}}$ in the natural ordering we just defined, the k-dimensional Fourier transform can be written as

$$\hat{\mathbf{E}} = \begin{bmatrix} \hat{E}(0,0,\ldots,0) \\ \vdots \\ \hat{E}(q-1,q-1,\ldots,q-1) \end{bmatrix} = \underbrace{\mathbf{F} \otimes \cdots \otimes \mathbf{F}}_{k \text{ times}} \begin{bmatrix} E(0,0,\ldots,0) \\ \vdots \\ E(q-1,q-1,\ldots,q-1) \end{bmatrix} = \left(\underbrace{\mathbf{F} \otimes \cdots \otimes \mathbf{F}}_{k \text{ times}} \right) \mathbf{E}. \tag{21}$$

Definition B.4 (Non-uniform DFT matrices)**.** *For every $1 \le i \le k$, define (recall that $p_i(z)$'s are the marginal probabilities of E at coordinate i) the* non-uniform DFT matrix *at coordinate i to be*

$$\tilde{\mathbf{F}}_{\mathbf{i}} = \begin{bmatrix} qp_i(0) & qp_i(1) & qp_i(2) & qp_i(3) & \cdots & qp_i(q-1) \\ 1 & \omega & \omega^2 & \omega^3 & \cdots & \omega^{q-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \cdots & \omega^{2(q-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \cdots & \omega^{3(q-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{q-1} & \omega^{2(q-1)} & \omega^{3(q-1)} & \cdots & \omega^{(q-1)(q-1)} \end{bmatrix}$$

The following Lemma follows directly from Fact B.3 and (18).

Lemma B.5. *If we arrange the entries in $\mathbf{E}'$ and $\hat{\mathbf{E}}^{\text{non}}$ in the natural ordering, then the $q^k \times q^k$ matrix $\tilde{\mathbf{F}}$ in* (19) *is the tensor product of k non-uniform DFT matrices, i.e.,*

$$\tilde{\mathbf{F}} = \tilde{\mathbf{F}}_{\mathbf{1}} \otimes \cdots \otimes \tilde{\mathbf{F}}_{\mathbf{k}},$$

and consequently

$$\hat{\mathbf{E}}^{\text{non}} = (\tilde{\mathbf{F}}_{\mathbf{1}} \otimes \cdots \otimes \tilde{\mathbf{F}}_{\mathbf{k}})\mathbf{E}'.$$

The following is a well-known fact on the determinants of tensor product matrices, see, e.g., [38] for an elementary proof.

Fact B.6. *If $\mathbf{A}$ is an $m \times m$ square matrix and $\mathbf{B}$ is an $n \times n$ square matrix, then*

$$\det(\mathbf{A} \otimes \mathbf{B}) = (\det(\mathbf{A}))^n (\det(\mathbf{B}))^m.$$

Proposition B.7. *The non-uniform DFT matrix is non-singular for every $1 \le i \le k$. In particular,*

$$\det(\tilde{\mathbf{F}}_{\mathbf{i}}) = q\,(p_i(0) + \cdots + p_i(q-1))\,(-1)^{q-1} \prod_{1 \le \ell < m \le q-1} (\omega^m - \omega^\ell) = (-1)^{q-1} q \prod_{1 \le \ell < m \le q-1} (\omega^m - \omega^\ell) \neq 0.$$

Proof. Using Laplace expansion along the first row, we have

$$\det(\tilde{\mathbf{F}}_{\mathbf{i}}) = \sum_{j=0}^{q-1} (-1)^j qp_i(j) \det(\mathbf{M}_{1j}). \tag{22}$$

The determinant of the minor $\mathbf{M}_{1j}$ is

$$\det(\mathbf{M}_{1j}) = \begin{vmatrix} 1 & \omega & \cdots & \omega^{j-1} & \omega^{j+1} & \cdots & \omega^{q-1} \\ 1 & \omega^2 & \cdots & \omega^{2(j-1)} & \omega^{2(j+1)} & \cdots & \omega^{2(q-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{q-1} & \cdots & \omega^{(j-1)(q-1)} & \omega^{(j+1)(q-1)} & \cdots & \omega^{(q-1)(q-1)} \end{vmatrix}$$

$$= \left(\prod_{\ell=0, \ell \neq j}^{q-1} \omega^{\ell} \right) \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 & \cdots & 1 \\ 1 & \omega & \cdots & \omega^{j-1} & \omega^{j+1} & \cdots & \omega^{q-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{q-2} & \cdots & \omega^{(j-1)(q-2)} & \omega^{(j+1)(q-2)} & \cdots & \omega^{(q-1)(q-2)} \end{vmatrix}$$

$$= \prod_{\ell=0, \ell \neq j}^{q-1} \omega^{\ell} \prod_{\substack{0 \le \ell < m \le q-1 \\ \ell, m \neq j}} (\omega^m - \omega^{\ell})$$

$$= \frac{\prod_{\ell=0,\ell\neq j}^{q-1} \omega^{\ell} \prod_{0 \le \ell < m \le q-1} (\omega^m - \omega^{\ell})}{\prod_{\ell=0}^{j-1} (\omega^j - \omega^{\ell}) \prod_{\ell=j+1}^{q-1} (\omega^{\ell} - \omega^j)},$$

since the matrix in the second step is a Vandermonde matrix.

Using the fact that $\omega^q = 1$, the denominator may be simplified as

$$\begin{aligned}
&\prod_{\ell=0}^{j-1} (\omega^j - \omega^{\ell}) \prod_{\ell=j+1}^{q-1} (\omega^{\ell} - \omega^j) \\
=&(-1)^j \prod_{\ell=0}^{j-1} \omega^{\ell} \prod_{\ell=1}^{j} (1 - \omega^{\ell}) \prod_{\ell=j+1}^{q-1} (\omega^{\ell} - \omega^j) \\
=&(-1)^j \prod_{\ell=0}^{j-1} \omega^{\ell} \prod_{\ell=1}^{j} (1 - \omega^{\ell}) \prod_{\ell=j+1}^{q-1} \omega^{\ell-q} (\omega^q - \omega^{q+j-\ell}) \\
=&(-1)^j \prod_{\ell=0}^{j-1} \omega^{\ell} \prod_{\ell=1}^{j} (1 - \omega^{\ell}) \prod_{\ell=j+1}^{q-1} \omega^{\ell} \prod_{\ell=j+1}^{q-1} (1 - \omega^{q+j-\ell}) \\
=&(-1)^j \prod_{\ell=0,\ell\neq j}^{q-1} \omega^{\ell} \prod_{\ell=1}^{q-1} (1 - \omega^{\ell}).
\end{aligned}$$

Therefore we have

$$\det(\mathbf{M}_{1j}) = (-1)^j (-1)^{q-1} \prod_{1 \le \ell < m \le q-1} (\omega^m - \omega^{\ell}).$$

Plugging $\det(\mathbf{M}_{1j})$ into (22) completes the proof. □

Combining Fact B.6 and Proposition B.7 gives

Lemma B.8. *We have that*

$$\det(\tilde{\mathbf{F}}) = \det(\tilde{\mathbf{F}}_{\mathbf{1}} \otimes \cdots \otimes \tilde{\mathbf{F}}_{\mathbf{k}}) \neq 0.$$

Recall that we are assuming that all the non-zero Fourier coefficients $\hat{E}^{\text{non}}(\boldsymbol{a})$ are zero. Now to make the linear system of equations in (19) complete, we add another constraint that $\hat{E}^{\text{non}}(\mathbf{0}) = \sum_{\boldsymbol{z}} E'(\boldsymbol{z}) = cq^k$, where c is a constant which will be determined later. Since $\tilde{\mathbf{F}}$ is non-singular, there is a unique solution to this system of q^k linear equations. But we know the uniform distribution $E'(\boldsymbol{z}) = c$ for every $\boldsymbol{z} \in \Sigma^k$ is a solution (by the first part of Theorem 4.3 we have already shown), therefore this is the unique solution.

Now we have, for every $\boldsymbol{z} \in \Sigma^k$, $E(\boldsymbol{z})\theta_S(\boldsymbol{z}) = c$. Observe that $1/\theta_S(\boldsymbol{z}) = q^k p_1(z_1) \cdots p_k(z_k)$, and since $p_i(z)$ are marginal probabilities, $\sum_{z \in \Sigma} p_i(z) = 1$ for every i, it follows that

$$\sum_{\boldsymbol{z} \in \Sigma^k} \frac{1}{\theta_S(\boldsymbol{z})} = q^k \sum_{\boldsymbol{z} \in \Sigma^k} p_1(z_1) \cdots p_k(z_k) = q^k.$$

Using the fact that $\sum_{\boldsymbol{z} \in \Sigma^k} E(\boldsymbol{z}) = 1$, we arrive at

$$1 = \sum_{\boldsymbol{z} \in \Sigma^k} E(\boldsymbol{z}) = c \sum_{\boldsymbol{z} \in \Sigma^k} \frac{1}{\theta_S(\boldsymbol{z})} = q^k c,$$

and therefore $c = \frac{1}{q^k}$ and $E(\boldsymbol{z}) = \frac{1}{q^k \theta_S(\boldsymbol{z})} = p_1(z_1) \cdots p_k(z_k)$ as desired. This completes the proof of Lemma B.1. □

C Testing Almost k-wise Independence over Product Spaces

We will follow [1] and define almost k-wise independence in terms of max-norm.

Definition C.1 (Uniform Almost k-wise Independence)**.** *Let Σ be a finite set with $|\Sigma| = q$. A discrete probability distribution D over Σ^n is (uniform)* (ϵ, k)-wise independent *if for any set of k indexes $\{i_1, \ldots, i_k\}$ and for all $z_1, \ldots, z_k \in \Sigma$,*

$$\left| \Pr_{\boldsymbol{X} \sim D}[X_{i_1} \cdots X_{i_k} = z_1 \cdots z_k] - 1/q^k \right| \leq \epsilon.$$

Generalizing this definition to non-uniform almost k-wise independence over product spaces is straightforward.

Definition C.2 (Non-uniform Almost k-wise Independence over Product Spaces)**.** *Let $\Sigma_1, \ldots, \Sigma_n$ be finite sets. A discrete probability distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$ is (non-uniform)* (ϵ, k)-wise independent *if for any set of k indexes $\{i_1, \ldots, i_k\}$ and for all $z_{i_1} \in \Sigma_{i_1}, \ldots, z_{i_k} \in \Sigma_{i_k}$,*

$$\left| \Pr_{\boldsymbol{X} \sim D}[X_{i_1} \cdots X_{i_k} = z_{i_1} \cdots z_{i_k}] - \Pr_{\boldsymbol{X} \sim D}[X_{i_1} = z_{i_1}] \times \cdots \times \Pr_{\boldsymbol{X} \sim D}[X_{i_k} = z_{i_k}] \right| \leq \epsilon.$$

From now on we will work with the most general notion of almost k-wise independence, that is non-uniform almost k-wise independent distributions over product spaces. Let $\mathcal{D}_{(\epsilon,k)}$ denote the set of all (ϵ, k)-wise independent distributions. The distance between a distribution D and the set of (ϵ, k)-wise independent distributions is the minimum statistical distance between D and any distribution in $\mathcal{D}_{(\epsilon,k)}$, i.e., $\Delta(D, \mathcal{D}_{(\epsilon,k)}) = \min_{D' \in \mathcal{D}_{(\epsilon,k)}} \Delta(D, D')$. D is said to be *δ-far* from (ϵ, k)-wise independence if $\Delta(D, \mathcal{D}_{(\epsilon,k)}) > \delta$. We write $q_{\max}$ for $\max_{1 \leq i \leq n} |\Sigma_i|$. To simplify notation, we use vectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ of dimensions $|\Sigma_1|, \ldots, |\Sigma_n|$, respectively to denote the marginal probabilities at each coordinates. That is, for every $z_j \in \Sigma_i$, the j^{th} component of $\boldsymbol{p}_i$ satisfies $\boldsymbol{p}_i(z_j) = \Pr_{\boldsymbol{X} \sim D}[X_i = z_j]$. Clearly we have $\sum_{z_j \in \Sigma_i} \boldsymbol{p}_i(z_j) = 1$ for every $1 \leq i \leq n$.

In the property testing setting, when given a distribution D, we would like to distinguish between the case that D is in $\mathcal{D}_{(\epsilon,k)}$ from the case that D is δ-far from $\mathcal{D}_{(\epsilon,k)}$.

Theorem C.3. *Given a discrete distribution D over $\Sigma_1 \times \cdots \times \Sigma_n$, there is a testing algorithm with query complexity $O(\frac{k\log(nq_{max})}{\epsilon^2\delta^2})$ and time complexity $\tilde{O}(\frac{(nq_{max})^k}{\epsilon^2\delta^2})$ such that the following holds. If $D \in \mathcal{D}_{(\epsilon,k)}$, then the algorithm outputs "Accept" with probability at least $2/3$; if D is δ-far from $\mathcal{D}_{(\epsilon,k)}$, then the algorithm outputs "Reject" with probability at least $2/3$.*

To analyze the testing algorithm we will need the following Lemma which, roughly speaking, states that the distance parameter δ can be translated into the error parameter ϵ (up to a factor of ϵ) in the definition of almost k-wise independence.

Lemma C.4 ([1]). *Let D be a distribution over $\Sigma_1 \times \cdots \times \Sigma_n$. If $\Delta(D, \mathcal{D}_{(\epsilon,k)}) > \delta$, then $D \notin \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$. If $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \leq \delta$, then $D \in \mathcal{D}_{(\epsilon+\delta,k)}$.*

Proof. For the first part, suppose $D \in \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$. Let $U_{\boldsymbol{p}_1,\dots,\boldsymbol{p}_n}$ denote the distribution that, for every $z_1 \cdots z_n \in \Sigma_1 \times \cdots \times \Sigma_n$, $U_{\boldsymbol{p}_1,\dots,\boldsymbol{p}_n}(z_1 \cdots z_n) = \boldsymbol{p}_1(z_1) \cdots \boldsymbol{p}_n(z_n)$. It is easy to check that, since $\sum_{z_i} \boldsymbol{p}_i = 1$, $U_{\boldsymbol{p}_1,\dots,\boldsymbol{p}_n}$ is indeed a distribution. Now define a new distribution D' as $D' = (1-\delta)D + \delta U_{\boldsymbol{p}_1,\dots,\boldsymbol{p}_n}$, then one can easily verify that $D' \in \mathcal{D}_{(\epsilon,k)}$, therefore $\Delta(D, \mathcal{D}_{(\epsilon,k)}) \leq \delta$.

For the second part, recall that no randomized procedure can increase the statistical difference between two distributions [37], therefore projecting to any set of k coordinates and looking at the probability of finding any specific string of length k can not increase the statistical distance between D and any distribution in $\mathcal{D}_{(\epsilon,k)}$. It follows that, when restricted to any k coordinates, the max-norm of D is at most $\epsilon + \delta$. □

Proof of Theorem C.3. The testing algorithm is to sample $Q = O(\frac{k\log(nq_{\max})}{\epsilon^2\delta^2})$ times from the distribution D and estimate, for every k-subset $I = \{i_1, \dots, i_k\}$ of $[n]$ and every $z_{i_1} \cdots z_{i_k}$, $\Pr_{\boldsymbol{X}\sim D}[X_{i_1} \cdots X_{i_k} = z_{i_1} \cdots z_{i_k}]$. Write $\bar{p}_I(z_{i_1} \cdots z_{i_k})$ for the estimated probability from the samples, $p_I^D(z_{i_1} \cdots z_{i_k})$ for $\Pr_{\boldsymbol{X}\sim D}[X_{i_1} \cdots X_{i_k} = z_{i_1} \cdots z_{i_k}]$ and $p_I(z_{i_1} \cdots z_{i_k})$ for $\Pr_{\boldsymbol{X}\sim D}[X_{i_1} = z_{i_1}] \times \cdots \times \Pr_{\boldsymbol{X}\sim D}[X_{i_k} = z_{i_k}]$. If, there are some I and $z_{i_1} \in \Sigma_{i_1}, \dots, z_{i_k} \in \Sigma_{i_k}$ such that $|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})| > \epsilon + \epsilon\delta/2$, then output "Reject", otherwise output "Accept".

The query complexity and time complexity of the testing algorithm are straightforward to check. For the correctness of the algorithm, we observe that $\mathbf{E}\left[\bar{p}_I(z_{i_1} \cdots z_{i_k})\right] = p_I^D(z_{i_1} \cdots z_{i_k})$. Since $\bar{p}_I(z_{i_1} \cdots z_{i_k})$ is the average of Q independent $0/1$ random variables, Chernoff bound gives

$$\Pr[\left|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I^D(z_{i_1} \cdots z_{i_k})\right| \geq \epsilon\delta/2] \leq \exp(-\Omega(\epsilon^2\delta^2 Q)).$$

By setting $Q = C\frac{k\log(nq_{\max})}{\epsilon^2\delta^2}$ for large enough constant C and applying a union bound argument to all k-subsets and all possible strings of length k, we get that, with probability at least $2/3$, for every I and every $z_{i_1}, \dots, z_{i_k}$, $\left|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I^D(z_{i_1} \cdots z_{i_k})\right| < \epsilon\delta/2$.

Now if $D \in \mathcal{D}_{(\epsilon,k)}$, then with probability at least $2/3$, for all I and all $z_{i_1}, \dots, z_{i_k}$, $\left|p_I^D(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})\right| \leq \epsilon$, so by triangle inequality $|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})| \leq \epsilon + \epsilon\delta/2$. Therefore the algorithm outputs "Accept".

If D is δ-far from (ϵ, k)-wise independence, then by Lemma C.4, $D \notin \mathcal{D}_{(\epsilon+\epsilon\delta,k)}$. That is, there are some I and $z_{i_1}, \dots, z_{i_k}$ such that $\left|p_I^D(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})\right| > \epsilon + \epsilon\delta$. Then with probability at least $2/3$, $|\bar{p}_I(z_{i_1} \cdots z_{i_k}) - p_I(z_{i_1} \cdots z_{i_k})| > \epsilon + \epsilon\delta/2$. Therefore the algorithm outputs "Reject". □