THEORY OF COMPUTING, Volume 9 (15), 2013, pp. 559–577
www.theoryofcomputing.org

SPECIAL ISSUE: APPROX-RANDOM 2012

Almost k-Wise vs. k-Wise Independent Permutations, and Uniformity for General Group Actions*

Noga Alon† Shachar Lovett‡

Received August 19, 2012; Revised January 14, 2013; Published May 30, 2013

Abstract: A family of permutations in S_n is k-wise independent if a uniform permutation chosen from the family maps any sequence of k distinct elements to any sequence of k distinct elements with equal probability. Efficient constructions of k-wise independent permutations are known for $k = 2$ and $k = 3$ based on multiply transitive permutation groups but are unknown for $k \geq 4$. In fact, it is known that there are no nontrivial subgroups of S_n for $n \geq 25$ which are 4-wise independent ("4-transitive"). Faced with this obstacle, research has turned towards constructing almost k-wise independent families, where small errors are allowed. Constructions of almost k-wise independent families of permutations, with optimal size up to polynomial factors, have been achieved by several authors.

Motivated by this problem, we give several results relating almost k-wise and k-wise distributions over permutations.

*An earlier version of this paper appeared in the Proceedings of the 16th International Workshop on Randomization and Computation (RANDOM '12), pages 350–361, 2012.

†Supported in part by an ERC advanced grant and by NSF grant DMS-0835373.
‡Supported by NSF grant DMS-0835373.

ACM Classification: G.3

AMS Classification: 68W20,68Q25

Key words and phrases: local independence, permutations, groups, representation theory

 DOI: 10.4086/toc.2013.v009a015

1. Any almost k-wise independent distribution, with small enough error, is close in statistical distance to a k-wise independent distribution.
2. A uniformly random set of $n^{O(k)}$ permutations supports, with high probability, a distribution which is k-wise independent.
3. Derandomizing this, we show that any family which is almost $2k$-wise independent, with small enough error, supports a distribution which is k-wise independent.

These results allow for simplified analysis of randomized algorithms. For example, our results show that one can analyze an algorithm assuming access to k-wise independent permutation families, but then use it with only almost k-wise independent families, with a provable correctness guarantee.

In fact, we prove all of these results in the general setting of a group actions. Let G be a group acting on a set X. The case of k-wise permutations corresponds to $G = S_n$ and X the set of sequences of k distinct elements. A subset S of G is X-uniform if for any $x, y \in X$, the probability over a uniform $g \in S$ that $g(x) = y$ is the same as when g is chosen uniformly from G. It is approximately X-uniform if these probabilities are close. We prove all the above results in this general setting, relating almost X-uniform and X-uniform distributions. Our proof is based on basic tools from the representation theory of finite groups.

1 Introduction

Small probability spaces of limited independence are widely used in many applications. Specifically, if the analysis of a randomized algorithm depends only on the assumption that the random bits used are only k-wise independent, one can replace the random tape by a tape selected from a k-wise independent distribution. One application of this is a derandomization of the algorithm by enumerating over all possible random strings. Another application is when the random string needs to be saved, for example in data structures, where using k-wise independence allows one to maintain a succinct data structure.

The case of k-wise independent distributions over $\{0,1\}^n$ has been widely studied, and there are optimal constructions of k-wise independent probability spaces of size $n^{O(k)}$ (see e. g., [4]). Moreover, these constructions are *strongly explicit*: given an index of an element $i \in [n^{O(k)}]$ and an index of a bit $j \in [n]$, one can compute the j-th bit of the i-th string in time $O(k \log n)$. This is crucial for several applications, for example for streaming algorithms and cryptography, where operations need to be performed in poly-logarithmic time.

Another widely studied case is that of k-wise independent permutations of n elements. This problem is motivated by cryptographic applications, as k-wise independent permutations allow perfect secrecy even if one allows k oracle queries to the encryption. For more details on the role of k-wise independent permutations in cryptography, see, e. g., [21, 25, 26, 27].

Here, the situation is much less understood. For $k = 2$ the group of invertible affine transformations $x \mapsto ax + b$ over a finite field $\mathbb{F}$ yields a 2-wise independent family; and for $k = 3$ the group of Möbius transformations $x \mapsto (ax+b)/(cx+d)$ with $ad - bc = 1$ over the projective line $\mathbb{F} \cup \{\infty\}$ yields a 3-wise independent family.

For $k \geq 4$ (and n large enough) the situation is dramatically different. Until recently, no k-wise independent families of permutation were known other than the full symmetric group S_n and the alternating group A_n. In fact, it is known (cf., e. g., [7], Theorem 5.2) that for $n \geq 25$ and $k \geq 4$ there are no other subgroups of S_n which form a k-wise independent family.[1] Recently, two works broke this barrier. The first is by Finucane, Peled and Yaari [9] who gave an explicit construction of a k-wise independent family of permutations of size k^{2n}. The second is by Kuperberg, Lovett and Peled [17] who proved the existence of k-wise independent families of permutations of size $n^{O(k)}$. Still, no explicit construction of k-wise independent families of permutations of size smaller than exponential in n is known.

Faced with this problem, research has turned towards constructing families of permutations which are *almost k-wise independent*, allowing for small errors. There has been much research towards constructing explicit almost k-wise independent families of minimal size. This was achieved, up to polynomial factors, by Kaplan, Naor and Reingold [12], who gave a construction of such a family of size $n^{O(k)}$. Alternatively, such a family can also be obtained from the construction of Kassabov [14] of a constant size expanding set of S_n by considering all words of length $O(k \log n)$. Both of these constructions are also strongly explicit: given an index of a permutation $i \in [n^{O(k)}]$ and an element $j \in [n]$, one can compute the image of the i-th permutation on j in time $O(k \log n)$. Again, this is crucial for applications such as streaming algorithms or cryptography.

For many applications, almost k-wise independent families are just as good as k-wise independent families. However, the analysis must take the error into account, which in some cases is not trivial. Our first result shows that by choosing the error small enough, one can analyze an algorithm using k-wise independent permutations, and then apply almost k-wise independent permutations to achieve almost the same results.

Theorem 1.1. *Let μ be a distribution taking values in S_n which is almost k-wise independent with error ε. Then there exists a distribution μ' over permutations which is k-wise independent, and such that the statistical distance between μ and μ' is at most $O(\varepsilon \cdot n^{4k})$.*

A similar result for k-wise independent hash functions was obtained by Alon, Goldreich and Mansour [5], and more generally over product spaces by Rubinfeld and Xie [20]. Our proof technique is similar in spirit, although technically more involved. This allows for an oblivious derandomization of randomized algorithms (with two-sided error) which "work" given any k-wise independent distribution over permutations: let f be a boolean function, and let A be a randomized algorithm such that

$$\Pr_{\pi \sim \mu}[A(x,\pi) = f(x)] \geq 2/3$$

for any k-wise independent distribution over permutations μ. Then A can be derandomized by letting π be chosen uniformly from an almost k-wise independent distribution with error $\varepsilon \leq O(n^{-4k})$. Since such distributions can be generated strongly explicitly, the overhead (in terms of the number of bits needed to sample from the distribution) is just $O(k \log n)$.

Another relaxation of the problem of constructing small families of k-wise independent permutations is that of considering weighted families, or equivalently distributions of small support, which are k-wise

[1] In the language of group theory, these are k-transitive groups. The currently known proof of this fact is hard, as it requires the classification of finite simple groups.

independent. The relaxation here is that the elements can have unequal weights. We note that such distributions are typically useless for derandomization purposes as the weights could require exponentially small precision, hence requiring polynomially many random bits to compute. Still, it is interesting to inspect what is known for distributions.

Contrary to the case of families, it is simple to establish that there exist distributions of small support which are k-wise independent. First, note that given a family S of permutations, it is easy to decide if there exists a distribution μ supported on S which is k-wise independent using linear programming: for a permutation π define the matrix $M_k(\pi)$ to be the permutation on sequences of k distinct elements induced by π. It is an $(n)_k \times (n)_k$ permutation matrix, where $(n)_k := \prod_{i=0}^{k-1}(n-i)$. Let U denote the uniform matrix all of whose elements are $(n-k)!/n!$. Then there exists a k-wise independent distribution supported on S iff U belongs to the convex hull of $\{M_k(\pi) : \pi \in S\}$. The latter condition can be easily verified using linear programming. Now, starting with any set of permutations which support k-wise independent permutations (for example the set of all permutations), one can apply Carathéodory theorem [8] and deduce that U lies in the convex hull of at most n^{2k} permutations. That is, there exist k-wise independent distributions which are supported on at most n^{2k} permutations. Moreover, and somewhat surprisingly, one can algorithmically find a k-wise independent distribution with small support in a *weakly explicit* manner (i. e., in time $n^{O(k)}$) using the ideas of Karp and Papadimitriou [13] and Koller and Megiddo [15].[2]

We consider the problem of constructing small explicit sets which support k-wise independent distributions. First, we establish that most small sets support k-wise independent distributions.

Theorem 1.2. *Let S be a uniformly random subset of S_n of size cn^{6k} for an appropriately chosen absolute constant c. Then with high probability (w. h. p., for short) over the choice of S, there exists a distribution μ supported on S which is k-wise independent.*

A similar result for k-wise independent hash functions was obtained by Austrin and Håstad [6]. Our result implies a somewhat surprising consequence for search algorithms which "work" given any k-wise independent distribution over permutations, which allows us to transform weak guarantees to strong guarantees. Let f be a function and A an algorithm, such that for any k-wise independent distribution μ,

$$\Pr_{\pi\sim\mu}[A(x,\pi) = f(x)] > 0\,.$$

Then since almost all sets of size $n^{O(k)}$ support such a distribution, we must have that A has a noticeable fraction of witnesses in S_n,

$$\Pr_{\pi\in S_n}[A(x,\pi) = f(x)] \geq n^{-O(k)}\,.$$

We also show that almost $2k$-wise independent permutations give an explicit construction of a set which supports k-wise independence, thus derandomizing Theorem 1.2.

Theorem 1.3. *Let S be a subset of S_n such that S is almost $2k$-wise independent with error $\varepsilon \leq O(n^{-7k})$. Then there exists a distribution μ supported on S which is k-wise independent.*

[2] Essentially, the linear program for finding μ has $n!$ variables and $n^{O(k)}$ constraints. Its dual has $n^{O(k)}$ variables and $n!$ constraints. The dual problem can be solved efficiently using the ellipsoid method since it has an efficient separating-hyperplane oracle.

We are not aware of a similar result, even in the case of k-wise independent hash functions. This allows for an oblivious derandomization of search algorithms which "work" given any k-wise independent distribution over permutations: let f be a function, and let A be a randomized algorithm such that

$$\Pr_{\pi\sim\mu}[A(x,\pi)=f(x)]>0$$

for any k-wise independent distribution μ over permutations. Then taking S to be an almost $2k$-wise independent family of permutations with error $O(n^{-7k})$, we get that there exists $\pi\in S$ for which $A(x,\pi)=f(x)$, achieving an oblivious derandomization of A with overhead (measured in bits, as before) $O(k\log n)$.

Here is a toy example illustrating the way the last theorem and the discussion preceding it can be applied. Let $G=(V,E)$ be a graph on a set V of n vertices, and suppose that each vertex $v\in V$ has a real positive weight $w(v)$. Let $d(v)$ be the degree of v, and assume all degrees are bounded by k. We claim that G contains an independent set $U\subset V$ of total weight $W(U)=\sum_{u\in U}w(u)$ at least

$$\sum_{v\in V}\frac{w(v)}{d(v)+1}\,.$$

To prove it, let π be a random permutation of the set of vertices V, and let U consist of all vertices u so that $\pi(u)$ precedes $\pi(v)$ for every neighbor v of u. It is clear that U is an independent set, and for any vertex $u\in V$ the probability that $u\in U$ is exactly $1/(d(u)+1)$, as this is the probability that u precedes all its neighbors. By linearity of expectation, the expected value of the total weight of U is $\sum_{v\in V}w(v)/(d(v)+1)$ and hence there exists an independent set U of total weight at least as claimed.

The above proof clearly works even if π is only assumed to be $(k+1)$-wise independent.[3] Therefore, the discussion preceding Theorem 1.3 implies that if π is chosen uniformly at random, then the probability it provides a set U satisfying $W(U)\geq\sum_{v\in V}w(v)/(d(v)+1)$, is at least $n^{-O(k)}$. The theorem itself shows that the support of any set of almost $(2k+2)$-wise independent permutations with sufficiently small error must contain a permutation π that provides an independent set U as above.

A similar reasoning can be applied to other arrangement problems. Given a k-uniform hypergraph with a weight for each permutation of the vertices in each of its edges, one may want to find a permutation maximizing the total weight of all orders induced on the sets of vertices in the edges. Problems of this type are called k-CSP-rank problems, (see, e. g., [1]), and include Betweenness and Feedback Arc Set. In most of these problems, finding the precise optimum is NP-hard, and the reasoning above provides some insight about algorithms for the (much easier) problem of finding a permutation in which the total weight is at least as large as the expected weight in a uniformly random permutation.

1.1 Group action uniformity vs. almost uniformity

We actually prove all the aforementioned results in the general setting of *group actions*, of which k-wise independent permutations as well as k-wise independent random variables form specific instances. A group G *acts* on a set X if G acts as a group of permutations on X. That is, $g:X\to X$ is a permutation of X for all $g\in G$, and $(gh)(x)=g(h(x))$ for all $g,h\in G$ and $x\in X$. This gives a general framework: k-wise independent permutations correspond to the case of $G=S_n$ the group of permutations, and

[3] In fact, it suffices if π is $(k+1)$-minwise independent.

$X = [n]_k = \{i_1, \dots, i_k \in [n] \text{ distinct}\}$ is the set of sequences of k distinct elements, where the action of G on X is straightforward. The case of k-wise independent distributions over $\{0,1\}^n$ corresponds to $G = \mathbb{F}_2^n$ and $X = [n]_k \times \mathbb{F}_2^k$, where the action of $g = (g_1, \dots, g_n) \in \mathbb{F}_2^n$ on $x = ((i_1, \dots, i_k), (b_1, \dots, b_k)) \in [n]_k \times \mathbb{F}_2^k$ is given by $g(x) = ((i_1, \dots, i_k), (b_1 + g_{i_1}, \dots, b_k + g_{i_k}))$. Similarly, one can obtain in this way distributions supporting k-wise independent random variables, even when each variable is distributed over a different domain.

We now introduce some definitions. If G acts on X, a distribution μ over G is *X-uniform* if

$$\Pr_{g \sim \mu}[g(x) = y] = \Pr_{g \in G}[g(x) = y]$$

for all $x, y \in X$; and is *almost X-uniform with error ε* if

$$\left| \Pr_{g \sim \mu}[g(x) = y] - \Pr_{g \in G}[g(x) = y] \right| \leq \varepsilon$$

for all $x, y \in X$. These definitions coincide with k-wise independence and almost k-wise independence for permutations when $G = S_n$ and $X = [n]_k$. Theorem 1.1, Theorem 1.2 and Theorem 1.3 are immediate corollaries of the following general theorems, when applied to $G = S_n$ and $X = [n]_k$.

First, we show that distributions over G which are almost X-uniform with small enough error, are close in statistical distance to distributions which are X-uniform.

Theorem 1.4. *Let μ be a distribution on G which is almost X-uniform with error ε. Then there exists a distribution μ' on G which is X-uniform, and such that the statistical distance between μ and μ' is at most $\varepsilon \cdot 3|X|^4$.*

Second, we show that a small random subset of G supports w. h. p. an X-uniform distribution.

Theorem 1.5. *Let $S \subset G$ be a uniformly random set of size $c|X|^6$ for an appropriately chosen absolute constant c. Then with high probability over the choice of S, there exists a distribution μ supported on S which is X-uniform.*

Finally, we derandomize Theorem 1.5. Recall that if G acts on X, then G also acts on $X \times X$ in the obvious manner, i. e., $g((x_1, x_2)) = (g(x_1), g(x_2))$. We show that if a distribution over G is almost $X \times X$-uniform with a small enough error, then it must support an X-uniform distribution.

Theorem 1.6. *Let μ be a distribution supported on a set $S \subset G$ which is almost $(X \times X)$-uniform with error $\varepsilon < 0.5|X|^{-7}$. Then there exists a distribution μ' supported on S which is X-uniform.*

The proof of Theorem 1.5 is by a counting argument using the symmetry of the group action. The proofs of Theorem 1.4 and Theorem 1.6 rely on representation theory of finite groups. In the language of the Fourier analysis literature, we prove results regarding quadrature rules for the representations appearing in the action of G on X. Technically, our arguments involving representation theory are quite basic, and as such are similar in spirit to several known results in the Fourier analysis literature. In particular, Theorem 1.5 is similar to theorems established in [16, 2]. However, our proof is arguably simpler, as it applies the Carathéodory theorem instead of a more involved second moment argument. Also, some technical parts used in the proof of Theorem 1.6 are related to known results in the Fourier analysis literature, e. g., in [18, 19].

Paper organization We present preliminary definitions in Section 2. Theorem 1.4 is proved in Section 3, Theorem 1.5 in Section 4 and Theorem 1.6 in Section 5. We conclude with some open problems in Section 6. Throughout the paper we do not attempt to optimize constants.

2 Preliminaries

For an integer $t \geq 1$ we denote $[t] := \{1, 2, \ldots, t\}$. The statistical (or total variation) distance between two distributions μ, μ' is given by $\mathrm{dist}(\mu, \mu') = \sum_x |\mu(x) - \mu'(x)|$. The expression $\delta_{i,j}$ is equal to 1 if $i = j$ and is equal to 0 otherwise.

Group action and uniformity A group G *acts* on a set X if there is a homomorphism from G to the permutation group on X. That is, each $g \in G$ is a permutation on X, and $(gh)(x) = g(h(x))$ for all $g, h \in G, x \in X$. We denote by U_G the uniform distribution over G. For a distribution μ over G we denote by $g \sim \mu$ a random element chosen according to μ. We abbreviate $\Pr_{g \in G}[\cdot] = \Pr_{g \sim U_G}[\cdot]$.

We recall some definitions from the introduction: a distribution μ over G is *X-uniform* if

$$\Pr_{g \sim \mu}[g(x) = y] = \Pr_{g \in G}[g(x) = y]$$

for all $x, y \in X$; and a distribution μ is *almost X-uniform with error ε* if

$$\left| \Pr_{g \sim \mu}[g(x) = y] - \Pr_{g \in G}[g(x) = y] \right| \leq \varepsilon$$

for all $x, y \in X$. A family $S \subset G$ is *X-uniform* (*almost X-uniform*, accordingly) if the uniform distribution over S is such. We will need some basic facts in linear algebra, geometry and representation theory, which are presented below.

Linear algebra Let $A = (a_{i,j})$ be a complex matrix. The L_∞ norm of A is $\|A\|_\infty = \max |a_{i,j}|$ (not to be confused with the operator norm of A on L_∞, which is not used in this paper). The Frobenius norm of A is $\|A\|_{\mathrm{Fr}} = \sqrt{\sum |a_{i,j}|^2}$. For every A, $\|A\|_\infty \leq \|A\|_{\mathrm{Fr}}$. A matrix A is unitary if $A\overline{A^t} = I$. The Frobenius norm of a matrix is invariant under any unitary basis change. That is, if U is unitary then $\|U^{-1}AU\|_{\mathrm{Fr}} = \|A\|_{\mathrm{Fr}}$. The tensor product of a $d_1 \times d_1$ matrix A_1 and a $d_2 \times d_2$ matrix A_2, denoted $A_1 \otimes A_2$, is a $(d_1 d_2) \times (d_1 d_2)$ matrix, whose entries are given by $(A_1 \otimes A_2)_{(i,i'),(j,j')} = (A_1)_{i,j}(A_2)_{i',j'}$. The tensor product of unitary matrices is also unitary.

Geometry Let $X = \{x_1, \ldots, x_N\}$ be a set of points in $\mathbb{R}^d$. The *convex hull* of X is the set of points contained in the minimal convex set containing X; equivalently, it is the set of all points

$$\left\{ \sum \lambda_i x_i : \lambda_1, \ldots, \lambda_N \geq 0, \sum \lambda_i = 1 \right\}.$$

Fact 2.1 (Carathéodory theorem [8]). *Let X be a finite set of points in $\mathbb{R}^d$, and let y be a point in the convex hull of X. Then there exists a subset $Y \subset X$ of size $|Y| \leq d+1$ such that y is in the convex hull of Y.*

Any hyperplane H partitions a set X of points into two sets: if $H = \{x : \langle a,x\rangle = b\}$ then the sets are $\{x \in X : \langle a,x\rangle \geq b\}$ and $\{x \in X : \langle a,x\rangle < b\}$. We need the following bound on the maximal number of ways a set can be partitioned by hyperplanes.[4]

Fact 2.2 (Harding [11])**.** *Let X be a set of N points in $\mathbb{R}^d$. The number of different ways to partition X into two sets by a hyperplane is at most $\sum_{i=0}^{d} \binom{N-1}{i} \leq N^d$.*

Representation theory Let G be a finite group. A representation of G (over $\mathbb{C}$) is a homomorphism $R : G \to \mathrm{GL}(d,\mathbb{C})$. That is, $R(g)$ for $g \in G$ is a $d \times d$ nonsingular complex matrix, and $R(gh) = R(g)R(h)$ for every $g,h \in G$. The dimension of the representation R is d. Two representations R,R' of G of dimension d are *equivalent* if there exists an invertible matrix A such that $R'(g) = A^{-1}R(g)A$ for all $g \in G$. This is denoted as $R \equiv R'$.

A representation R is *unitary* if $R(g)$ is unitary for all $g \in G$.

Fact 2.3. *Any representation of G is equivalent to a unitary representation.*

We will restrict our attention only to unitary representations in this paper. We note that if R,R' are unitary and equivalent, then there exists a unitary matrix A such that $R'(g) = A^{-1}R(g)A$.

Let G be a group which acts on a set X, that is, $g : X \to X$ is a permutation of X for all $g \in G$, and $(gh)(x) = g(h(x))$ for all $g,h \in G$ and $x \in X$. The associated representation R_X maps each $g \in G$ to the permutation matrix it induces on the set X. That is, $R_X(g)$ is an $|X| \times |X|$ matrix, indexed by $x,y \in X$, defined as $(R_X(g))_{x,y} = 1$ if $g(x) = y$ and $(R_X(g))_{x,y} = 0$ otherwise. Note that R_X is always a unitary representation.

The sum of two representations $R_1 : G \to \mathrm{GL}(d_1,\mathbb{C})$ and $R_2 : G \to \mathrm{GL}(d_2,\mathbb{C})$ is a representation $R : G \to \mathrm{GL}(d_1+d_2,\mathbb{C})$ where $R(g)$ is defined as a block diagonal matrix with two blocks, given by $R_1(g)$ and $R_2(g)$. For $e \geq 1$ let $eR := R + \cdots + R$ where the sum is over e copies of R. A representation R is *reducible* if it is equivalent to the sum of two representations. Otherwise, the representation R is *irreducible*. We summarize a few basic properties of representations below. For details we refer the reader to any standard book on representation theory, e. g., [10].

Fact 2.4 (Maschke's theorem)**.** *Any representation R of G is equivalent to a sum of irreducible representations $R \equiv e_1R_1 + \cdots + e_tR_t$, where $R_1,\ldots,R_t$ are nonequivalent irreducible representations, and $e_i \geq 1$ is the multiplicity of R_i. We have $\sum e_i \dim(R_i) = \dim(R)$.*

Fact 2.5 (Schur's lemma)**.** *Let R be a unitary irreducible representation of G of dimension d. Then for any $i,j,k,\ell \in [d]$,*

$$\frac{1}{|G|}\sum_{g\in G} R(g)_{i,j}\overline{R(g)_{k,\ell}} = \frac{1}{d}\delta_{i,k}\delta_{j,\ell}\,.$$

Let R',R'' be two non-equivalent unitary irreducible representations of G of dimensions d',d''. Then for any $i,j \in [d']$ and $k,\ell \in [d'']$,

$$\frac{1}{|G|}\sum_{g\in G} R'(g)_{i,j}\overline{R''(g)_{k,\ell}} = 0\,.$$

[4] A quick way to prove a slightly weaker estimate is as follows: the VC-dimension [24] of halfspaces is $d+1$. Hence by the VC-dimension theorem [24, 22, 23] the number of partitions is at most $\sum_{i=0}^{d+1}\binom{N}{i} \leq N^{d+1}$.

The trivial representation is given by $\mathbf{1}(g) = 1$ for all $g \in G$. An immediate corollary of Schur's lemma is that for every representation R which is irreducible and nontrivial, we have $\sum_{g\in G} R(g) = 0$.

The group algebra $\mathbb{C}[G]$ is the linear space of functions $\mu : G \to \mathbb{C}$. It is often written as $\mu = \sum_{g\in G} \mu(g)\cdot g$. Note that the distributions over G form a subset of $\mathbb{C}[G]$. For $\mu \in \mathbb{C}[G]$ and a representation R of G, let $R(\mu) := \sum_{g\in G} \mu(g)R(g)$. If μ is a distribution, this is equivalent to $R(\mu) = \mathbb{E}_{g\sim\mu}[R(g)]$. Note that in this notation, the statistical distance between two distributions μ, μ' on G is $\mathrm{dist}(\mu, \mu') = \|\mu - \mu'\|_1$.

3 Almost X-uniform distributions are statistically close to X-uniform distributions

We prove in this section Theorem 1.4, which states that almost X-uniform distributions with small enough error are statistically close to X-uniform distributions. We recall the statement of the theorem.

Theorem 1.4 *Let μ be a distribution on G which is almost X-uniform with error ε. Then there exists a distribution μ' on G which is X-uniform, and such that the statistical distance between μ and μ' is at most $\varepsilon \cdot 3|X|^4$.*

We first rephrase the conditions for a distribution to be X-uniform, or almost X-uniform, in terms of representations. Let R_X be the representation of the action of G on X, i. e., $R_X(g)_{x,y} = 1_{g(x)=y}$. Let U_G denote the uniform distribution over G.

Proposition 3.1. *Let μ be a distribution on G. Then*

1. *μ is X-uniform iff $R_X(\mu) = R_X(U_G)$;*
2. *μ is almost X-uniform with error ε iff $\|R_X(\mu) - R_X(U_G)\|_\infty \le \varepsilon$.*

Proof. The claim is immediate from the definitions of X-uniform and almost X-uniform distributions, since $R_X(\mu)_{x,y} = \Pr_{g\sim\mu}[g(x) = y]$ and $R_X(U_G)_{x,y} = \Pr_{g\in G}[g(x) = y]$. □

The first step is to decompose R_X into its irreducible representations. Let $R_X \equiv e_0\mathbf{1} + e_1R_1 + \cdots + e_tR_t$, where $R_1, \ldots, R_t$ are unitary nonequivalent non-trivial irreducible representations, and e_i is the multiplicity of R_i in R_X. We next transform the conditions of Proposition 3.1 to the basis of the irreducible representations.

Proposition 3.2. *Let μ be a distribution on G. Then*

1. *μ is X-uniform iff $R_i(\mu) = 0$ for all $i \in [t]$;*
2. *if μ is almost X-uniform with error ε then $\|R_i(\mu)\|_\infty \le \varepsilon|X|$ for all $i \in [t]$.*

Proof. As μ is a distribution, then $\mathbf{1}(\mu) = \sum_{g\in G} \mu(g) = 1$, and the same holds for U_G. Hence always $\mathbf{1}(\mu) = \mathbf{1}(U_G)$. Thus, $R_X(\mu) = R_X(U_G)$ iff $R_i(\mu) = R_i(U_G)$ for all $i \in [t]$. The first item follows since $R_i(U_G) = 0$ for all $i \in [t]$. To see that, note that by Schur's lemma

$$R_i(U_G)_{j,k} = \frac{1}{|G|}\sum_{g\in G} R_i(g)_{j,k} = \frac{1}{|G|}\sum_{g\in G} R_i(g)_{j,k}\overline{\mathbf{1}(g)} = 0$$

since R_i and $\mathbf{1}$ are nonequivalent unitary irreducible representations. To prove the second item, let μ be an almost X-uniform distribution with error ε. By Proposition 3.1 this is equivalent to $\|R_X(\mu)-R_X(U_G)\|_\infty \le \varepsilon$. The L_∞ norm is not convenient for the basis change required to switch to the basis of the irreducible representations. We thus switch to the Frobenius norm, which is trivially bounded by

$$\|R_X(\mu)-R_X(U_G)\|_{\mathrm{Fr}} \le \varepsilon|X|\,.$$

Note that the Frobenius norm is invariant under unitary change of basis, and as both R_X and $R_1,\ldots,R_t$ are unitary, the basis change can also be assumed to be unitary. We thus have

$$\sqrt{\sum_{i=1}^{t} e_i\|R_i(\mu)-R_i(U_G)\|_{\mathrm{Fr}}^2} = \|R_X(\mu)-R_X(U_G)\|_{\mathrm{Fr}} \le \varepsilon|X|\,,$$

which combined with the fact that $R_i(U_G)=0$ implies that $\|R_i(\mu)\|_\infty \le \varepsilon|X|$. □

The main idea in the proof of Theorem 1.4 is to "correct" each element of $R_i(\mu)$ to be zero by making a small statistical change in μ, and without affecting the other elements of R_i or in any other $R_{i'}$. This is analogous to the proof idea of [5] for almost k-wise independent bits (see also [3]). Performing all these local changes sequentially over all elements of R_i, $i\in[t]$, will shift μ into an X-uniform distribution. Actually, as a first step we will get an element of $\mathbb{C}[G]$ which we then fix to be a distribution.

Let R_i be one of the irreducible representations, and let $d_i=\dim(R_i)$ be its dimension. For $j,k\in[d_i]$ we define $\Delta_{i,j,k}\in\mathbb{C}[G]$ as

$$\Delta_{i,j,k}(g)=\frac{d_i}{|G|}\overline{R_i(g)_{j,k}}\,.$$

We consider how shifting μ by a small multiple of $\Delta_{i,j,k}$ affects the entries of $R_1,\ldots,R_t$.

Proposition 3.3. *Let $i\in[t]$, $j,k\in[d_i]$ and $i'\in[t]$, $j',k'\in[d_{i'}]$. For any $\delta\in\mathbb{R}$ we have*

$$R_{i'}(\mu+\delta\Delta_{i,j,k})_{j',k'} = R_{i'}(\mu)_{j',k'}+\delta\cdot 1_{(i,j,k)=(i',j',k')}\,.$$

Proof. First, by additivity

$$R_{i'}(\mu+\delta\Delta_{i,j,k})_{j',k'} = R_{i'}(\mu)_{j',k'}+\delta\cdot R_{i'}(\Delta_{i,j,k})_{j',k'}.$$

The claim follows from the orthogonality of the entries of the irreducible representations. By Schur's Lemma,

$$R_{i'}(\Delta_{i,j,k})_{j',k'} = \frac{d_i}{|G|}\sum_{g\in G} R_{i'}(g)_{j',k'}\overline{R_i(g)_{j,k}} = 1_{(i,j,k)=(i',j',k')}\,.$$ □

We will also need the following proposition, which asserts that $\mathbf{1}(\Delta_{i,j,k})=0$ and that $\|\Delta_{i,j,k}\|_\infty$ is bounded.

Proposition 3.4. *Let $i\in[t]$, $j,k\in[d_i]$. Then*

1. $\mathbf{1}(\Delta_{i,j,k})=0$*;*

2. $\|\Delta_{i,j,k}\|_\infty \le |X|/|G|$.

Proof. The first item follows because $\sum_{g\in G} R_i(g)_{j,k} = 0$ by Schur's lemma, since R_i is a nontrivial irreducible representation. The second item follows because $d_i \le |X|$ and because $|R_i(g)_{j,k}| \le 1$ since $R_i(g)$ is a unitary matrix. □

Applying Proposition 3.3 and Proposition 3.4 iteratively over all elements of $R_1,\dots,R_t$, we obtain the following corollary.

Corollary 3.5. *Let μ be a distribution over G which is almost X-uniform with error ε. Define $\Delta \in \mathbb{C}[G]$ by*

$$\Delta(g) = -\sum_{i\in[t]} \sum_{j,k\in[d_i]} R_i(\mu)_{j,k} \cdot \Delta_{i,j,k}(g)\,.$$

Then

1. $R_X(\mu+\Delta) = R_X(U_G)$;
2. $\|\Delta\|_\infty \le \varepsilon|X|^4/|G|$.

Proof. The first item holds since $R_i(\mu+\Delta)_{j,k} = R_i(U_G)_{j,k}$ for all $i\in[t]$ and $j,k\in d_i$ by Proposition 3.3, and since $\mathbf{1}(\mu+\Delta) = \mathbf{1}(U_G) = 1$ by the first item in Proposition 3.4. The second item holds since $\sum d_i^2 \le |X|^2$ as $\dim(R_X) = |X|$, $|R_i(\mu)_{j,k}| \le \varepsilon|X|$ by Proposition 3.2, and $|\Delta_{i,j,k}(g)| \le |X|/|G|$ by the second item in Proposition 3.4. □

We are nearly done. The only problem is that $\mu+\Delta$ may not be a distribution: it may be complex, or have negative values. This can be fixed, without increasing the statistical distance too much. This concludes the proof of Theorem 1.4.

Proof of Theorem 1.4. Let $\lambda = |G|\cdot\|\Delta\|_\infty \le \varepsilon|X|^4$. Define

$$\mu' = (1-\lambda)\left(\mu + \frac{\Delta+\overline{\Delta}}{2}\right) + \lambda U_G\,.$$

We claim that μ' is a distribution which is X-uniform. This will establish the result as

$$\|\mu-\mu'\|_1 \le \lambda\|\mu\|_1 + (1-\lambda)\|\Delta\|_1 + \lambda\|U_G\|_1 \le 3\lambda = 3\varepsilon|X|^4\,.$$

First let us show that $R_X(\mu') = R_X(U_G)$. We already know by Corollary 3.5 that $R_X(\mu+\Delta) = R_X(U_G)$. Conjugating this equality, since R_X is a real representation (i. e., all elements in $R_X(g)$ are real), and since $\mu, U_G \in \mathbb{R}[G]$ are also real, we obtain that also

$$R_X(\mu+\overline{\Delta}) = R_X(U_G)\,.$$

Thus $R_X(\mu') = R_X(U_G)$ since $R_X(\mu')$ is a convex combination of $R_X(\mu+\Delta), R_X(\mu+\overline{\Delta})$ and $R_X(U_G)$.

To conclude we need to show that μ' is a distribution, i. e., it is real, nonnegative and sums to one. By definition of μ' it is real, and since $R_X(\mu') = R_X(U_G)$ we have $\sum_{g\in G}\mu'(g) = \mathbf{1}(\mu') = \mathbf{1}(U_G) = 1$. The bound $\mu'(g) \ge 0$ for all $g\in G$ follows by elementary calculations from $\mu(g)\ge 0$, $|\Delta(g)| \le \lambda/|G|$ and $U_G(g) = 1/|G|$. □

4 Uniformly random sets support X-uniform distributions

We establish Theorem 1.5 in this section, which states that w. h. p. a uniformly random set of size $|X|^{O(1)}$ supports an X-uniform distribution. We recall the statement of the theorem.

Theorem 1.5 *Let $S \subset G$ be a uniformly random set of size k for $k = O(|X|^6)$. Then with high probability over the choice of S, there exists a distribution μ supported on S which is X-uniform.*

Recall that a distribution μ is X-uniform if $\Pr_{g \sim \mu}[g(x) = y] = \Pr_{g \in G}[g(x) = y]$ for all $x, y \in X$. We say a set S *supports X-uniformity* if there exists a distribution supported on S which is X-uniform. We first establish that this is a purely geometric property of S.

Let R_X be the representation of the action of G on X, that is,

$$R_X(g)_{x,y} = 1_{g(x)=y}\,.$$

Let $U = R_X(U_G) = \mathbb{E}_{g \in G}[R_X(g)]$ denote the matrix which corresponds to the action on X of the uniform distribution over G. We consider these matrices as points in $\mathbb{R}^d$ for $d = |X|^2$.

Proposition 4.1. *A set $S \subset G$ supports X-uniformity iff the convex hull of the matrices $\{R_X(g) : g \in S\}$ contains the matrix U.*

Proof. A point in the convex hull is given by $M = \sum_{g \in S} \mu(g) \cdot R_X(g)$ where $\mu(g) \geq 0$ and $\sum_{g \in S} \mu(g) = 1$. Thus, each point in the convex hull corresponds to a distribution μ over S, and vice versa. Note that $M_{x,y} = \Pr_{g \sim \mu}[g(x) = y]$, hence an X-uniform distribution corresponds to the matrix U. $\square$

Let $S \subset G$ be a uniformly random set. By Proposition 4.1 it is enough to show that the matrix U lies in the convex hull of $\{R_X(g) : g \in S\}$. Suppose this is not the case; then there must exist a hyperplane H in $\mathbb{R}^d$ which passes through U and such that all matrices $\{R_X(g) : g \in S\}$ lie on one side of H. We first show that any hyperplane which passes through U has a noticeable fraction of the matrices $\{R_X(g) : g \in G\}$ on both sides.

Proposition 4.2. *Let H be a hyperplane which passes through U. The number of matrices $\{R_X(g) : g \in G\}$ on any side of H is at least $|G|/(|X|^2 + 1)$.*

Proof. Let H^+ denote a halfspace defined by H, and let $G^+ = \{g \in G : R_X(g) \in H^+\}$ denote the set of permutations whose corresponding matrices lie in H^+. The matrix U can be written by Carathéodory theorem as the convex combination of $d+1$ matrices $R_X(g_0), \ldots, R_X(g_d)$. We claim that for any $h \in G$, the matrix U also belongs to the convex hull of $R_X(g_0 h), \ldots, R_X(g_d h)$. This follows since $R_X(g_i h) = R_X(g_i) R_X(h)$ and $U R_X(h) = U$. Thus, at least one of $g_0 h, \ldots, g_d h$ must lie in G^+, for any choice of $h \in G$. This concludes the proof since for a randomly chosen h,

$$1 = \Pr_{h \in G}[\exists i, g_i h \in G^+] \leq \sum_{i=0}^{d} \Pr_{h \in G}[g_i h \in G^+] = (d+1) \cdot \frac{|G^+|}{|G|}\,. \qquad \square$$

We now establish Theorem 1.5.

Proof of Theorem 1.5. Let $S \subset G$ be a uniformly random set of N elements, chosen with repetitions. Let $K \lhd G$ be the normal subgroup of G which acts trivially on X, i. e., $K = \{g \in G : g(x) = x\ \forall x \in X\}$. Observe that the quotient group G/K also acts on X, and that $\{R_X(g) : g \in G\} = \{R_X(g) : g \in G/K\}$. Thus the number of distinct matrices $R_X(g)$ is bounded by $|G/K| \le |X|!$, and by Fact 2.2 the number of ways to partition this set of matrices by any hyperplane, and in particular one which passes through U, is bounded by $(|X|!)^d$. Fix such a partition. The number of matrices $\{R_x(g) : g \in G\}$ which lies on each side of the partition is at least $|G|/(d+1)$ by Proposition 4.2. Hence, the probability that S is contained in one side of the partition is bounded by $2(1-1/(d+1))^N$. Thus, by the union bound, the probability that there exists a hyperplane passing through U, such that S is contained in one side of it, is at most

$$|(X!)|^d \cdot 2\left(1-\frac{1}{d+1}\right)^N \le 2\exp(-N/(d+1)+d\log(|X|!)),$$

which is at most 0.01 (say) for $N = O(d^2\log(|X|!)) \le O(|X|^6)$. □

5 Almost X-uniform distributions support X-uniform distributions

We prove in this section Theorem 1.6, which states that if μ is an almost $X \times X$-uniform distribution with small enough error, then there exists an X-uniform distribution μ' supported on the support of μ. We recall the statement of the theorem.

Theorem 1.6 *Let μ be a distribution supported on a set $S \subset G$ which is almost $(X \times X)$-uniform with error $\varepsilon < 0.5|X|^{-7}$. Then there exists a distribution μ' supported on S which is X-uniform.*

Fix such a distribution μ, and let S denote its support, $S = \{g : \mu(g) > 0\}$. Let R_X be the representation of G acting on X. By Proposition 4.1, S supports an X-uniform distribution iff $R_X(U_G) = \mathbb{E}_{g\in G}[R_X(g)]$ lies in the convex hull of $\{R_X(g) : g \in S\}$. Assume this is not the case; then there exists a hyperplane H which passes through $R_X(U_G)$ and such that all $\{R_X(g) : g \in S\}$ lie on one side of H.

We first project H into a hyperplane with a simpler representation. Let $R_X \equiv e_0\mathbf{1} + e_1R_1 + \cdots + e_tR_t$ denote the decomposition of R_X into unitary nonequivalent irreducible representation, and let $d_i = \dim(R_i)$ denote the dimension of each irreducible representation. Essentially, we will project H to "use" only one copy from each nontrivial irreducible representation. That is, we will show that H can be projected to a hyperplane separating 0 from $\{R_1(g) \times \cdots \times R_t(g) : g \in S\}$.

Proposition 5.1. *There exists a map $L : G \to \mathbb{R}$ given by*

$$L(g) := \sum_{i\in[t]} \sum_{j,k\in[d_i]} \lambda_{i,j,k} \cdot R_i(g)_{j,k}$$

for some coefficients $\{\lambda_{i,j,k} \in \mathbb{C} : i \in [t], j,k \in [d_i]\}$ such that

1. $\mathbb{E}_{g\in G}[L(g)] = 0$;

2. *for all $g \in S$, $L(g) > 0$.*

Proof. The existence of a hyperplane H separating $R_X(U_G)$ from $\{R_X(g) : g \in S\}$ implies that there exists a map $L' : G \to \mathbb{R}$ defined as $L'(g) = \sum_{x,y \in X} \alpha_{x,y} R_X(g)_{x,y}$ for some real coefficients $\{\alpha_{x,y} \in \mathbb{R} : x, y \in X\}$ such that

$$L'(g) > \mathbb{E}_{h \in G}[L'(h)]$$

for all $g \in S$. Applying the linear transformation mapping R_X into the basis of irreducible representations, we get that $L'(g)$ can be expressed as

$$L'(g) = \sum_{\ell \in [e_0]} \beta_{0,\ell} \mathbf{1}(g) + \sum_{i \in [t]} \sum_{j,k \in [d_i]} \sum_{\ell \in [e_i]} \beta_{i,j,k,\ell} R_i(g)_{j,k}\,,$$

where $\beta_{0,\ell}, \beta_{i,j,k,\ell} \in \mathbb{C}$ are obtained by a linear transformation (over $\mathbb{C}$) of $\alpha_{x,y}$. Observe that $\mathbb{E}_{g \in G}[L'(g)] = \sum_{\ell \in [e_0]} \beta_{0,\ell}$ by Schur's lemma, and define $L(g) := L'(g) - \mathbb{E}[L'(g)]$. Note that $L : G \to \mathbb{R}$ is real since $L' : G \to \mathbb{R}$ was real, that $\mathbb{E}[L] = 0$ and that $L(g) > 0$ for all $g \in S$. The coefficient $\lambda_{i,j,k}$ is given by the sum of all $\beta_{i,j,k,\ell}$ over $\ell \in [e_i]$. $\square$

We may assume without loss of generality that $\mathbb{E}_{g \in G}[L^2(g)] = 1$ by multiplying all coefficients $\lambda_{i,j,k}$ by an appropriate factor. The main idea is to show that if μ is almost $X \times X$ uniform, then $\mathbb{E}_{g \sim \mu}[L^2(g)] \approx \mathbb{E}_{g \in G}[L^2(g)] = 1$ while $\mathbb{E}_{g \sim \mu}[L(g)] \approx \mathbb{E}_{g \in G}[L(g)] = 0$. Combining this with a bound on $\|L\|_\infty$ a simple calculation shows that it cannot be the case that $L(g) > 0$ for all g in the support of μ.

The first step is to show that the coefficients $\lambda_{i,j,k}$ cannot be very large.

Proposition 5.2.

$$\sum_{i \in [t]} \sum_{j,k \in [d_i]} \frac{|\lambda_{i,j,k}|^2}{d_i} = 1\,.$$

In particular, $|\lambda_{i,j,k}| \le |X|^{1/2}$ *for all* i, j, k.

Proof. We assumed $\mathbb{E}[L^2] = 1$, which, since L is real, implies $\mathbb{E}[|L|^2] = 1$. Using Schur's lemma we get

$$\begin{aligned} 1 &= \mathbb{E}_{g \in G}\big[L(g) \cdot \overline{L(g)}\big] \\ &= \sum_{i,i' \in [t]} \sum_{j,k \in [d_i]} \sum_{j',k' \in [d_{i'}]} \lambda_{i,j,k} \overline{\lambda_{i',j',k'}} \mathbb{E}_{g \in G}\big[R_i(g)_{j,k} \overline{R_{i'}(g)}_{j',k'}\big] \\ &= \sum_{i \in [t]} \sum_{j,k \in [d_i]} \frac{|\lambda_{i,j,k}|^2}{d_i}\,, \end{aligned}$$

and in particular $|\lambda_{i,j,k}|^2 \le d_i \le |X|$. $\square$

An immediate corollary is that $L(g)$ can never be very large.

Corollary 5.3. $|L(g)| \le |X|^{2.5}$ *for all* $g \in G$.

Proof. We have $|R_i(g)_{j,k}| \le 1$ since R_i is unitary, hence $|L(g)| \le \sum_{i \in [t]} \sum_{j,k \in [d_i]} |\lambda_{i,j,k}| \le |X|^{2.5}$ since $\sum d_i^2 \le |X|^2$. $\square$

The bound on $|\lambda_{i,j,k}|$ together with the assumption that μ is almost $X \times X$-uniform, implies that the first and second moments of L are approximately the same under μ and under the uniform distribution over G.

Proposition 5.4. *Let μ be a distribution which is almost $X \times X$-uniform with error ε. Then*

1. $|\mathbb{E}_{g\sim\mu}[L(g)]| \leq \varepsilon |X|^{4.5}$;
2. $|\mathbb{E}_{g\sim\mu}[L^2(g)] - 1| \leq \varepsilon |X|^7$.

Proof. We have

$$|\mathbb{E}_{g\sim\mu}[L(g)]| \leq \sum_{i\in[t]} \sum_{j,k\in[d_i]} |\lambda_{i,j,k}| |\mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}]| \,.$$

The bound on the first moment follows since $\sum d_i^2 \leq |X|^2$; since $|\lambda_{i,j,k}| \leq |X|^{1/2}$ by Proposition 5.2; and since μ is in particular X-uniform with error $\varepsilon|X|$, we have by Proposition 3.2 that $|\mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}]| \leq \varepsilon |X|^2$. The bound on the second moment is proved in a similar way. Recall that we proved the identity $\sum |\lambda_{i,j,k}|^2/d_i = 1$ in Proposition 5.2. So

$$\begin{aligned}\mathbb{E}_{g\sim\mu}[|L(g)|^2 - 1] &= \sum_{i,j,k} |\lambda_{i,j,k}|^2 \cdot \left(\mathbb{E}_{g\sim\mu}[|R_i(g)_{j,k}|^2] - 1/d_i\right) \\ &+ \sum_{(i,j,k)\neq(i',j',k')} \lambda_{i,j,k}\overline{\lambda_{i',j',k'}} \cdot \mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}] \,.\end{aligned}$$

To conclude the proof we need to show that $\mathbb{E}_{g\sim\mu}[|R_i(g)_{j,k}|^2] \approx 1/d_i$ and that $\mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}R_{i'}(g)_{j',k'}] \approx 0$.

The condition that μ is almost $X \times X$-uniform with error ε is equivalent to

$$\|R_{X\times X}(\mu) - R_{X\times X}(U_G)\|_\infty \leq \varepsilon \,.$$

Switching to the Frobenius norm, this implies

$$\|R_{X\times X}(\mu) - R_{X\times X}(U_G)\|_{\text{Fr}} \leq \varepsilon |X|^2 \,.$$

We now decompose $R_{X\times X}$ to simpler representations, coming from the irreducible representations of R_X. We have that $R_{X\times X} = R_X \otimes R_X$, which since R_X is real, also gives $R_{X\times X} = R_X \otimes \overline{R_X}$. Now, if $R_X \equiv e_0\mathbf{1} + \sum_{i=1}^t e_i R_i$ is the decomposition of R_X into irreducible unitary nonequivalent representations, we have

$$R_{X\times X} \equiv e_0^2\mathbf{1} + \sum_{i=1}^t e_0 e_i (R_i + \overline{R_i}) + \sum_{i,i'=1}^t e_i e_{i'} (R_i \otimes \overline{R_{i'}}) \,.$$

Note that this is not the decomposition of $R_{X\times X}$ into irreducible representations, since $R_i \otimes \overline{R_{i'}}$ is reducible! Nevertheless, we observe that as the basis change for R_X was unitary, so is the basis change for $R_{X\times X}$ (since the tensor product of two unitary matrices is again unitary). In particular, we get that

$$\|(R_i \otimes \overline{R_{i'}})(\mu) - (R_i \otimes \overline{R_{i'}})(U_G)\|_{\text{Fr}} \leq \varepsilon |X|^2 \,,$$

which implies that

$$\|(R_i \otimes \overline{R_{i'}})(\mu) - (R_i \otimes \overline{R_{i'}})(U_G)\|_\infty \leq \varepsilon |X|^2 \,.$$

The matrix $R_i \otimes \overline{R_{i'}}$ is indexed by $((j,j'),(k,k'))$, where $(R_i \otimes \overline{R_{i'}})(g)_{(j,j'),(k,k')} = R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}$. We thus get that for any i,j,k,i',j',k' we have

$$\left|\mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}] - \mathbb{E}_{g\in G}[R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}]\right| \le \varepsilon |X|^2 .$$

To conclude, by Schur's lemma

$$\mathbb{E}_{g\in G}[R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}] = \frac{1}{d_i} 1_{(i,j,k)=(i',j',k')} .$$

The bound for the second moment now follows by elementary calculations analogous to the ones for the first moment. □

We are now ready to prove Theorem 1.6.

Proof of Theorem 1.6. Let μ be almost $X \times X$ uniform with error $\varepsilon \le 0.5|X|^{-7}$. Summarizing Corollary 5.3 and Proposition 5.4, we have

1. $\|L\|_\infty \le |X|^{2.5}$;
2. $E_{g\sim\mu}[L(g)] \le \varepsilon|X|^{4.5}$;
3. $E_{g\sim\mu}[L(g)^2] \ge 1 - \varepsilon|X|^7$.

However, since we assumed by contradiction that $L(g) > 0$ for all g in the support of μ, we have

$$\mathbb{E}_{g\sim\mu}[L(g)^2] \le \|L(g)\|_\infty \cdot \mathbb{E}_{g\sim\mu}[L(g)] \le |X|^{2.5} \cdot \varepsilon|X|^{4.5},$$

i. e., we have

$$1 - \varepsilon|X|^7 \le \varepsilon|X|^7,$$

which is false whenever $\varepsilon < 0.5|X|^{-7}$. □

6 Summary and open problems

We showed that almost X-uniform (or almost $X \times X$-uniform) distributions are close to X-uniform distributions in two ways: they are statistically close to some X-uniform distribution μ', and they support a X-uniform distribution μ''. It may be possible that both can be realized by the same X-uniform distribution, i. e., that $\mu' = \mu''$. We leave this as an open problem.

Another interesting combinatorial problem is to construct small families which are uniform. Even in the special case of k-wise independent permutations, this is known explicitly only for families of exponential size [9] or non-explicitly by a probabilistic argument [17]. It is intriguing whether the probabilistic argument can be adapted to efficiently construct such families.

Acknowledgements We thank Avi Wigderson for helpful discussions and reference to the work of Karp and Papadimitriou [13], and Laci Babai for helpful comments.

References

[1] NIR AILON AND NOGA ALON: Hardness of fully dense problems. *Inform. and Comput.*, 205(8):1117–1129, 2007. [doi:10.1016/j.ic.2007.02.006] 563

[2] GORJAN ALAGIC AND ALEXANDER RUSSELL: Spectral concentration of positive functions on compact groups. *J. Fourier Anal. Appl.*, 17(3):355–373, 2011. [doi:10.1007/s00041-011-9174-5] 564

[3] NOGA ALON, ALEXANDR ANDONI, TALI KAUFMAN, KEVIN MATULEF, RONITT RUBINFELD, AND NING XIE: Testing k-wise and almost k-wise independence. In *Proc. 39th STOC*, pp. 496–505. ACM Press, 2007. [doi:10.1145/1250790.1250863] 568

[4] NOGA ALON, LÁSZLÓ BABAI, AND ALON ITAI: A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms*, 7(4):567–583, 1986. [doi:10.1016/0196-6774(86)90019-2] 560

[5] NOGA ALON, ODED GOLDREICH, AND YISHAY MANSOUR: Almost k-wise independence versus k-wise independence. *Inform. Process. Lett.*, 88(3):107–110, 2003. [doi:10.1016/S0020-0190(03)00359-4] 561, 568

[6] PER AUSTRIN AND JOHAN HÅSTAD: Randomly supported independence and resistance. *SIAM J. Comput.*, 40(1):1–27, 2011. Preliminary version in STOC'09. [doi:10.1137/100783534] 562

[7] PETER J. CAMERON: Permutation groups. In *Handbook of Combinatorics, Vol. 1*, pp. 611–645. Elsevier, Amsterdam, 1995. 561

[8] CONSTANTIN CARATHÉODORY: Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen. *Rendiconti del Circolo Matematico di Palermo*, 32(1):193–217, 1911. [doi:10.1007/BF03014795] 562, 565

[9] HILARY FINUCANE, RON PELED, AND YARIV YAARI: A recursive construction of t-wise uniform permutations. Technical report, 2012. *Random Structures and Algorithms*, to appear. [arXiv:1201.4960] 561, 574

[10] WILLIAM FULTON AND JOE HARRIS: *Representation Theory: A First Course*. Volume 129 of *Graduate Texts in Mathematics*. Springer, 1st edition, 1991. 566

[11] EDWARD FRANK HARDING: The number of partitions of a set of N points in k dimensions induced by hyperplanes. *Proc. Edinburgh Math. Soc. (2)*, 15(4):285–289, 1967. [doi:10.1017/S0013091500011925] 566

[12] EYAL KAPLAN, MONI NAOR, AND OMER REINGOLD: Derandomized constructions of k-wise (almost) independent permutations. *Algorithmica*, 55(1):113–133, 2009. Preliminary version in RANDOM'05. [doi:10.1007/s00453-008-9267-y] 561

[13] RICHARD M. KARP AND CHRISTOS H. PAPADIMITRIOU: On linear characterizations of combinatorial optimization problems. *SIAM J. Comput.*, 11(4):620–632, 1982. Preliminary version in FOCS'80. [doi:10.1137/0211053] 562, 574

[14] MARTIN KASSABOV: Symmetric groups and expander graphs. *Inventiones Mathematicae*, 170(2):327–354, 2007. [doi:10.1007/s00222-007-0065-y] 561

[15] DAPHNE KOLLER AND NIMROD MEGIDDO: Constructing small sample spaces satisfying given constraints. *SIAM J. Discrete Math.*, 7(2):260–274, 1994. Preliminary version in STOC'93. [doi:10.1137/S0895480192228140] 562

[16] KA-LAM KUEH, TIMOTHY OLSON, DANIEL ROCKMORE, AND KI-SENG TAN: Nonlinear approximation theory on compact groups. *J. Fourier Anal. Appl.*, 7(3):257–281, 2001. [doi:10.1007/BF02511813] 564

[17] GREG KUPERBERG, SHACHAR LOVETT, AND RON PELED: Probabilistic existence of rigid combinatorial structures. In *Proc. 44th STOC*, pp. 1091–1106. ACM Press, 2012. [doi:10.1145/2213977.2214075] 561, 574

[18] DAVID MASLEN: Efficient computation of Fourier transforms on compact groups. *J. Fourier Anal. Appl.*, 4(1):19–52, 1998. [doi:10.1007/BF02475926] 564

[19] AIDAN ROY AND ANDREW J. SCOTT: Unitary designs and codes. *Designs, Codes and Cryptography*, 53(1):13–31, 2009. [doi:10.1007/s10623-009-9290-2] 564

[20] RONITT RUBINFELD AND NING XIE: Robust characterizations of k-wise independence over product spaces and related testing results. *Random Structures & Algorithms*, 2012 (online). Preliminary version in ICALP'10. [doi:10.1002/rsa.20423] 561

[21] ALEXANDER RUSSELL AND HONG WANG: How to fool an unbounded adversary with a short key. *IEEE Trans. Inform. Theory*, 52(3):1130–1140, 2006. Preliminary version in EUROCRYPT'02. [doi:10.1109/TIT.2005.864438] 560

[22] NORBERT SAUER: On the density of families of sets. *J. Combin. Theory Ser. A*, 13(1):145–147, 1972. [doi:10.1016/0097-3165(72)90019-2] 566

[23] SAHARON SHELAH: A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41(1):247–261, 1972. Project Euclid. 566

[24] VLADIMIR N. VAPNIK AND ALEXEY YA. CHERVONENKIS: On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, 16(2):264–280, 1971. [doi:10.1137/1116025] 566

[25] SERGE VAUDENAY: Provable security for block ciphers by decorrelation. In *Proc. 15th Symp. Theoretical Aspects of Comp. Sci. (STACS'98)*, pp. 249–275. Springer, 1998. [doi:10.1007/BFb0028566] 560

[26] SERGE VAUDENAY: Adaptive-attack norm for decorrelation and super-pseudorandomness. In *Proc. 6th Ann. Internat. Workshop on Selected Areas in Cryptography (SAC'99)*, pp. 49–61. Springer, 1999. [doi:10.1007/3-540-46513-8_4] 560

[27] SERGE VAUDENAY: Decorrelation: A theory for block cipher security. *J. Cryptology*, 16(4):249–286, 2003. [doi:10.1007/s00145-003-0220-6] 560

AUTHORS

Noga Alon
Professor
Tel Aviv University, Israel
nogaa@tau.ac.il
http://www.tau.ac.il/~nogaa

Shachar Lovett
Assistant Professor
University of California, San Diego CA
slovett@ucsd.edu
http://cse.ucsd.edu/~slovett

ABOUT THE AUTHORS

NOGA ALON received his Ph. D. in Mathematics at the Hebrew University of Jerusalem under the supervision of Micha Perles. He is a Baumritter Professor of Mathematics and Computer Science at Tel Aviv University, and visits frequently the Institute for Advanced Study in Princeton. He works in Combinatorics, Graph Theory and their applications in Theoretical Computer Science, focusing on combinatorial algorithms, combinatorial geometry, combinatorial number theory, algebraic and probabilistic methods in Combinatorics, and has also been working on Circuit Complexity, Streaming algorithms, and topological methods in Combinatorics. He is a member of the Israel National Academy of Sciences and of Academia Europaea, and received several awards including the Pólya Prize, the Gödel Prize, the Israel Prize and the EMET Prize. He is married to Nurit and has three daughters. More details can be found at Noga Alon's Home Page.

SHACHAR LOVETT received his Ph. D. from the Weizmann Institute of Science in 2010 under the supervision of Omer Reingold and Ran Raz. He was a member in the Institute for Advanced Study School of Mathematics between 2010 and 2012. He is now an assistant professor in University of California San Diego School of Computer Science and Engineering. He works in Theoretical Computer Science with special emphasis on Computational Complexity, Coding theory, Randomness and Pseudo-randomness, Algebraic techniques and applications of Additive Combinatorics. He is married to Iris and has one daughter and one son. More details can be found at Shachar Lovett's Home Page.