A Characterization of Constant-Sample Testable Properties

Eric Blais*
David R. Cheriton School of Computer Science
University of Waterloo
eric.blais@uwaterloo.ca

Yuichi Yoshida†
National Institute of Informatics
yyoshida@nii.ac.jp

June 12, 2018

Abstract

We characterize the set of properties of Boolean-valued functions $f\colon \mathcal{X} \to \{0,1\}$ on a finite domain $\mathcal{X}$ that are testable with a constant number of samples $(x, f(x))$ with x drawn uniformly at random from $\mathcal{X}$. Specifically, we show that a property $\mathcal{P}$ is testable with a constant number of samples if and only if it is (essentially) a k-part symmetric property for some constant k, where a property is *k-part symmetric* if there is a partition $X_1, \dots, X_k$ of $\mathcal{X}$ such that whether $f\colon \mathcal{X} \to \{0,1\}$ satisfies the property is determined solely by the densities of f on $X_1, \dots, X_k$.

We use this characterization to show that symmetric properties are essentially the only graph properties and affine-invariant properties that are testable with a constant number of samples and that for every constant $d \geq 1$, monotonicity of functions on the d-dimensional hypergrid is testable with a constant number of samples.

1 Introduction

Property testing [17, 22] is concerned with the general question: for which properties of mathematical objects can we efficiently distinguish the objects that have the property from those that are "far" from having the same property? This question is formalized as follows. For a finite set $\mathcal{X}$, let $\{0,1\}^{\mathcal{X}}$ denote the set of Boolean-valued functions on $\mathcal{X}$ endowed with the normalized Hamming distance metric $\mathrm{d}(f,g) := |\{x \in \mathcal{X} : f(x) \neq g(x)\}|/|\mathcal{X}|$. A *property* of functions mapping $\mathcal{X}$ to $\{0,1\}$ is a subset $\mathcal{P} \subseteq \{0,1\}^{\mathcal{X}}$. A function is *$\epsilon$-close* to $\mathcal{P}$ if it is in the set $\mathcal{P}_\epsilon := \{f \in \{0,1\}^{\mathcal{X}} : \exists g \in \mathcal{P} \text{ s.t. } \mathrm{d}(f,g) \leq \epsilon\}$; otherwise, it is *$\epsilon$-far* from $\mathcal{P}$. An *ϵ-tester* for a property $\mathcal{P}$ is a randomized algorithm with error at most $\frac{1}{3}$ that accepts functions in $\mathcal{P}$ and rejects those that are ϵ-far from $\mathcal{P}$.

Most research in property testing has focused on the *query-based model*, where the tester is able to query the value of the function on any inputs of its choosing. In this work, however, we focus on the *sample-based model* of property testing that was also introduced by Goldreich, Goldwasser, and Ron in their seminal work [17]. In this model, the testing algorithm observes pairs $(x, f(x))$ where x is drawn uniformly at random from $\mathcal{X}$. The *sample complexity* of an ϵ-tester in the sample model is the maximum number of pairs $(x, f(x))$ that it observes before accepting or rejecting. When $\mathcal{P}$

*Supported by an NSERC Discovery Grant
†Supported by JSPS KAKENHI Grant Number JP17H04676 and JST ERATO Grant Number JPMJER1201

is a property where for every $\epsilon > 0$ there exists an ϵ-tester whose sample complexity is independent of the domain size $|\mathcal{X}|$, we say that $\mathcal{P}$ is *constant-sample testable*.

The goal of the present work is to determine which properties are constant-sample testable. It appears to be a widespread belief within the property testing community that "essentially no interesting properties" are constant-sample testable. This belief is supported by some results on individual properties. It is known, for instance, that monotonicity of Boolean functions [16], linearity [3, 18], linear threshold functions [4], and k-colorability of graphs [18] are all properties that are not constant-sample testable.

It is also known, however, that *some* non-trivial properties are constant-sample testable. All symmetric properties are easily seen to be constant-sample testable (see for example the discussion in [20]). And it is also folklore knowledge that the function identity testing properties—properties $\mathcal{P} = \{g\}$ that contain a single function—are constant-sample testable as well. More generally, every property that corresponds to a class of functions which can be learned with $O(1)$ samples is constant-sample testable [17]. Intuition may suggest that these two classes of properties are essentially the only ones that are constant-sample testable; i.e., that a property $\mathcal{P}$ is constant-sample testable if and only if it is close to some combination (e.g., the union, intersection, or symmetric difference) of a symmetric property $\mathcal{P}'$ and a property $\mathcal{P}''$ that corresponds to a class of functions that can be learned with $O(1)$ samples. This intuition, however, is wrong. Goldreich et al. [17] already showed that any characterization of constant-sample testability must also include other properties that are not of this form. Kearns and Ron [21] further showed that there are natural properties of functions over low-dimensional domains—namely unions of intervals and decision trees—that can also be tested with a constant number of samples. And, most recently, Berman, Murzabulatov, and Raskhodnikova [5] showed that convexity and halfspaces are two fundamental properties of images—which correspond to functions over the domain $\mathcal{X} = [n] \times [n]$—that are both constant-sample testable.

This collection of results shows that the class of constant-sample testable properties may be richer than is generally believed and that unifying the positive results on constant-sample testability requires new understanding of the structure of these properties.

1.1 Main result

We show that constant-sample testability is closely tied to a particular notion of symmetry—or invariance—of properties. Let $\mathcal{S}_{\mathcal{X}}$ denote the set of permutations on a finite set $\mathcal{X}$, and for any subset $S \subseteq \mathcal{X}$, let $\mathcal{S}_{\mathcal{X}}^{(S)}$ denote the set of permutations on $\mathcal{X}$ that preserves the elements in S. A permutation $\pi \in \mathcal{S}_{\mathcal{X}}$ acts on functions $f\colon \mathcal{X} \to \{0,1\}$ in the obvious way: πf is the function that satisfies $(\pi f)(x) = f(\pi x)$ for every $x \in \mathcal{X}$. The property $\mathcal{P} \subseteq \{0,1\}^{\mathcal{X}}$ is *invariant* under a permutation $\pi \in \mathcal{S}_{\mathcal{X}}$ if for every $f \in \mathcal{P}$, we also have $\pi f \in \mathcal{P}$. $\mathcal{P}$ is *(fully) symmetric* if it is invariant under all permutations in $\mathcal{S}_{\mathcal{X}}$. The following definition relaxes this condition to obtain a notion of "partial" symmetry.

Definition 1. For any $k \in \mathbb{N}$, the property $\mathcal{P} \subseteq \{0,1\}^{\mathcal{X}}$ is *k-part symmetric* if there is a partition of $\mathcal{X}$ into k parts $X_1, \ldots, X_k$ such that $\mathcal{P}$ is invariant under all the permutations in $\mathcal{S}_{\mathcal{X}}^{(X_1)} \cap \cdots \cap \mathcal{S}_{\mathcal{X}}^{(X_k)}$.

Equivalently, $\mathcal{P}$ is k-part symmetric if there exists a partition $X_1, \ldots, X_k$ of $\mathcal{X}$ such that the event $f \in \mathcal{P}$ is completely determined by the density $\frac{|f^{-1}(1) \cap X_i|}{|X_i|}$ of f in each of the sets $X_1, \ldots, X_k$. Our main result shows that $O(1)$-part symmetry is also essentially equivalent to constant-sample testability.

Theorem 1. *The property $\mathcal{P}$ of functions mapping $\mathcal{X}$ to $\{0,1\}$ is constant-sample testable if and only if for any $\epsilon > 0$, there exists a constant $k = k_{\mathcal{P}}(\epsilon)$ that is independent of $|\mathcal{X}|$ and a k-part symmetric property $\mathcal{P}'$ such that $\mathcal{P} \subseteq \mathcal{P}' \subseteq \mathcal{P}_\epsilon$.*

In words, Theorem 1 says that constant-sample testable properties are the properties $\mathcal{P}$ that can be covered by some $O(1)$-part symmetric property $\mathcal{P}'$ that does not include any function that is ϵ-far from $\mathcal{P}$. Note that this characterization cannot be replaced with the condition that $\mathcal{P}$ itself is k-part symmetric. To see this, consider the *function non-identity* property $\textsc{NotEq}(g)$ that includes every function except some non-constant function $g \colon \mathcal{X} \to \{0,1\}$. This property is not k-part symmetric for any $k = O(1)$, but the trivial algorithm that accepts every function is a valid ϵ-tester for $\textsc{NotEq}(g)$ for any constant $\epsilon > 0$.

Theorem 1 can easily be generalized to apply to properties of functions mapping $\mathcal{X}$ to any finite set $\mathcal{Y}$. We restrict our attention to the range $\mathcal{Y} = \{0,1\}$ for simplicity and clarity of presentation. The sample-based property testing model is naturally extended to non-uniform distributions over the input domain. It appears likely that Theorem 1 can be generalized to this more general setting as well, though we have not attempted to do so.

1.2 Applications

The characterization of constant-sample testability in Theorem 1 can be used to derive a number of different corollaries. We describe a few of these.

When $\mathcal{X}$ is identified with the $\binom{n}{2}$ pairs of vertices in V, the function $f \colon \mathcal{X} \to \{0,1\}$ represents a graph $G = (V,E)$ where $E = f^{-1}(1)$. A *graph property* is a property of these functions that is invariant under relabelling of the vertices. This definition corresponds to the *dense graph* model of property testing, and a number of basic graph properties are known to be testable with a constant number of queries in this model when those queries are selected by the algorithm (see [15, §8] and the references therein). By contrast, samples drawn at random from $\mathcal{X}$ appear to yield no useful information about the structure of a graph since the tester's observations will correspond to a collection of disjoint pairs of vertices that are either connected by an edge or not, so we may expect that no non-symmetric graph property is constant-sample testable. We can use our characterization to show that this is indeed the case: the only graph properties that are constant-sample testable are those that are (essentially) fully symmetric.

Corollary 1. *For every $\epsilon > 0$, if $\mathcal{P}$ is a graph property that is $\frac{\epsilon}{2}$-testable with a constant number of samples, then there is a symmetric property $\mathcal{P}^{\mathrm{sym}}$ such that $\mathcal{P} \subseteq \mathcal{P}^{\mathrm{sym}} \subseteq \mathcal{P}_\epsilon$.*

When $\mathcal{X}$ is identified with a finite field $\mathbb{F}_q^n$, a property $\mathcal{P} \subseteq \{0,1\}^{\mathcal{X}}$ is *affine-invariant* if it is invariant under any affine transformation over $\mathbb{F}_q^n$. Theorem 1 can be used to show that symmetric properties are essentially the only constant-sample testable affine-invariant properties.

Corollary 2. *If $\mathcal{P}$ is an affine-invariant property of functions $f \colon \mathbb{F}_q^n \to \{0,1\}$ that is $\frac{\epsilon}{2}$-testable with a constant number of samples, then there is a symmetric property $\mathcal{P}^{\mathrm{sym}}$ such that $\mathcal{P} \subseteq \mathcal{P}^{\mathrm{sym}} \subseteq \mathcal{P}_\epsilon$.*

Fix a constant $d \geq 1$. Two points $x, y \in [n]^d := \{1, 2, \ldots, n\}^d$ satisfy $x \preceq y$ when $x_1 \leq y_1$, $\ldots$, and $x_d \leq y_d$. The function $f \colon [n]^d \to \{0,1\}$ is *monotone* if for every $x \preceq y \in [n]^d$, we have $f(x) \leq f(y)$. When $d = 1$, it is folklore knowledge that monotonicity of Boolean-valued functions on the line is constant-sample testable. Using Theorem 1, we show that the same holds for every other constant dimension d.

Corollary 3. *For every constant $d \geq 1$ and constant $\epsilon > 0$, we can ϵ-test monotonicity of functions $f\colon [n]^d \to \{0,1\}$ on the d-dimensional hypergrid with a constant number of samples.*

Chen, Servedio, and Tan [9] showed that the number of queries (and thus also of samples) required to test monotonicity of $f\colon [n]^d \to \{0,1\}$ must depend on d. (The same result for the case where $n = 2$ was first established by Fischer et al. [13].) Combined with the result above, this shows that monotonicity of Boolean-valued functions on the hypergrid is constant-sample testable if and only if the number of dimensions of the hypergrid is constant.

1.3 Techniques

The proof of Theorem 1 follows the general outline of previous characterizations of the properties testable in the query-based model (e.g., [2, 7, 27]). As with those results, the most interesting part of the proof is the direction showing that constant-sample testability implies coverage by an $O(1)$-part symmetric property, and this proof is established with a regularity lemma. Our proof departs from previous results in both the type of regularity lemma that we use and in how we use it, as discussed below. The full proof of Theorem 1 is presented in Section 2.

Symmetry implies testability. The proof of this direction of the theorem is straightforward and is obtained by generalizing the following folklore proof that symmetric properties can be tested with a constant number of samples. Let $\mathcal{P}$ be any symmetric property. A tester can estimate the density $\mathbb{E}_{x \in \mathcal{X}}[f(x)]$ up to additive accuracy γ for any small $\gamma > 0$ with a constant number of samples. This estimated density can be used to accept or reject the function based on how close it is to the density of the functions in $\mathcal{P}$. The validity of this tester is established by showing that a function can be ϵ-far from $\mathcal{P}$ only when its density is far from the density of every function in $\mathcal{P}$.

Consider now a property $\mathcal{P}$ that is k-part symmetric for some constant k. Let $X_1, \ldots, X_k$ be a partition of $\mathcal{X}$ associated with $\mathcal{P}$. We show that a tester which estimates the densities $\mu_{X_i}(f) := \mathbb{E}_{x \in X_i}[f(x)]$ for each $i = 1, \ldots, k$ and then uses these densities to accept or reject is a valid tester for $\mathcal{P}$. We do so by showing that any function that is ϵ-far from $\mathcal{P}$ must have a density vector that is far from those of every function in $\mathcal{P}$.

Testability implies symmetry. To establish the second part of the theorem, we want to show that the existence of a constant-sample tester $\mathcal{T}$ for a property $\mathcal{P}$ implies that there is a partition of $\mathcal{X}$ into a constant number of parts for which $\mathcal{P}$ is nearly determined by the density of functions within those parts. We do so by using a variant of the Frieze–Kannan weak regularity lemma [14] for hypergraphs. An *s-uniform weighted hypergraph* is a hypergraph $G = (V, \xi)$ on $|V|$ vertices where $\xi\colon V^s \to [0,1]$ denotes the weight associated with each hyperedge. Given a subpartition $V_1, \ldots, V_k$ of V and a multi-index $I = (I_1, \ldots, I_s)$ with $I_1, \ldots, I_s \in [k]$, the *weight of hyperedges of G in V_I* is the average weight of the hyperedges in G with one vertex in each of the parts $V_{I_1}, \ldots, V_{I_s}$:

$$w_G(V_I) = w_G(V_{I_1}, \ldots, V_{I_s}) = \frac{\sum_{v_1 \in V_{I_1}, \ldots, v_s \in V_{I_s}} \xi(v_1, \ldots, v_s)}{\prod_{j \in [s]} |V_{I_j}|}.$$

For any subset $S \subseteq V$, we define $S \cap V_I = (S \cap V_{I_1}, \ldots, S \cap V_{I_s})$.

Lemma 1 (Weak regularity lemma). *For every $\epsilon > 0$ and every s-uniform weighted hypergraph $G = (V, \xi)$ with $\xi \colon V \to [0,1]$, there is a partition $V_1, \dots, V_k$ of V with $k = 2^{O(\log(\frac{1}{\epsilon})/\epsilon^2)}$ parts such that for every subset $S \subseteq V$,*

$$\sum_{I \in [k]^s} \frac{\prod_{j \in [s]} |S \cap V_{I_j}|}{|V|^s} \big| w_G(S \cap V_I) - w_G(V_I) \big| \le \epsilon.$$

This specific formulation of the weak regularity lemma seems not to have appeared previously in the literature, but its proof is essentially the same as that of usual formulations of the weak regularity lemma. For completeness, we provide a proof of Lemma 1 in Section 4.

Lemma 1 is best described informally when we consider the special case of unweighted graphs. In this setting, the weak regularity lemma says that for every graph G, there is a partition of the vertices of G into $k = O(1)$ parts $V_1, \dots, V_k$ such that for every subset S of vertices, the density of edges between $S \cap V_i$ and $S \cap V_j$ is close to the density between V_i and V_j *on average* over the choice of V_i and V_j. Regularity lemmas where this density-closeness condition is satisfied for *almost all* pairs of parts V_i and V_j are known as "strong" regularity lemmas. To the best of our knowledge, all previous characterization results in property testing that relied on regularity lemmas (e.g., [2, 7, 27]) used strong regularity lemmas. This approach unavoidably introduces tower-type dependencies between the query complexity and the characterization parameters. By using a weak regularity lemma instead, we get a much better (though still triply-exponential) dependence between the sample complexity and the partial symmetry parameter.

The second point of departure of our proof from previous characterizations is in how we use the regularity lemma. In prior work, the regularity lemma was applied to the tested object itself (e.g., the dense graphs being tested in [2]) and the testability of the property was used to show that the objects with the given property could be described by some combinatorial characteristics related to the regular partition whose existence is promised by the regularity lemma. Instead, in our proof of Theorem 1, we apply Lemma 1 to a hypergraph associated with the *tester* itself, not with the tested object.

Specifically, let $\mathcal{T}$ be an s-sample ϵ-tester for some property $\mathcal{P} \subseteq \{0,1\}^{\mathcal{X}}$. We associate $\mathcal{T}$ with an s-uniform weighted hypergraph $G_{\mathcal{T}}$ on the set of vertices $\mathcal{X} \times \{0,1\}$. The weight of each s-hyperedge of $G_{\mathcal{T}}$ is the acceptance probability of $\mathcal{T}$ when its s observations correspond to the s vertices covered by the hyperedge. By associating each function $f \colon \mathcal{X} \to \{0,1\}$ with the subset $S \subseteq \mathcal{X} \times \{0,1\}$ that includes all $2^{\mathcal{X}}$ vertices of the form $(x, f(x))$, we see that the probability that $\mathcal{T}$ accepts f is the expected value of a hyperedge whose s vertices are drawn uniformly and independently at random from the set S. We can use Lemma 1 to show that there is a partition of V into a constant number of parts such that for each function f with associated set S, this probability is well approximated by some function of the density of S in each of the parts. We then use this promised partition of V to partition the original input domain $\mathcal{X}$ into a constant number of parts where membership in $\mathcal{P}$ is essentially determined by the density of a function in each of these parts, as required.

1.4 Related work

Sample-based property testing. The first general result regarding constant-sample testable properties goes back to the original work of Goldreich, Goldwasser, and Ron [17]. They showed that every property with constant VC dimension (and, more generally, every property that corresponds to a class of functions that can be properly learned with a constant number of samples) is constant-sample testable. As they also show, this condition is not necessary for constant-sample testability—in fact, there are even properties that are testable with a constant number of samples whose corresponding class require a *linear* number of samples to learn [17, Prop. 3.3.1].

More general results on sample-based testers were obtained by Balcan *et al.* [4]. In particular, they defined a notion of *testing dimension* of a property $\mathcal{P}$ in terms of the total variation distance between the distributions on the tester's observations when a function is drawn from distributions π_{yes} and π_{no} essentially supported on $\mathcal{P}$ and $\overline{\mathcal{P}_\epsilon}$, respectively. They show that this testing dimension captures the sample complexity of $\mathcal{P}$ up to constant factors, and observe that it can be interpreted as an "average VC dimension"-type of complexity measure. It would be interesting to see whether the combinatorial characterization in Theorem 1 could be combined with these results to offer new insights into the connections between invariance and VC dimension-like complexity measures.

Finally, Goldreich and Ron [18] and Fischer *et al.* [11, 12] established connections between the query- and sample-based models of property testing giving sufficient conditions for *sublinear*-sample testability of properties. The exact bounds between sample complexity and partial symmetry in the proof of Theorem 1 yield another sufficient condition for sublinear-sample testability: every property $\mathcal{P}$ that can be ϵ-covered by an $o(\log\log|\mathcal{X}|)$-part symmetric function $\mathcal{P}'$ has sublinear sample complexity $o(|\mathcal{X}|)$. As far as we can tell, these two characterizations are incomparable.

Symmetry and testability. The present work was heavily influenced by the systematic exploration of connections between the invariances of properties and their testability initiated by Kaufman and Sudan [20]. (See also [24].) In that work, the authors showed that such connections yield new insights into the testability of algebraic properties in the query-based property testing model, and advocated for further study of the invariance of properties as a means to better understand their testability. Invariances and symmetry have also played a key role in the study of the query complexity for testing other properties as well, including for example in the study of graph properties [19] and properties of functions over finite fields [6]. Theorem 1 provides evidence that this approach is a critical tool in the study of sample-based property testing model as well.

The notion of partial symmetry and its connections to computational efficiency has a long history—it goes back at least to the pioneering work of Shannon [23]. Partial symmetry also appeared previously in a property testing context in the authors' joint work with Amit Weinstein on characterizing the set of functions for which isomorphism testing is constant-query testable [8]. However, it should be noted that the notion of partial symmetry considered in [8] does *not* correspond to the notion of k-part symmetry studied here. In fact, as mentioned in the conference version of that paper, there are 2-part symmetric functions for which isomorphism testing is not constant-query testable, so the two characterizations inherently require different notions of partial symmetry.

1.5 Organization

The proof of Theorem 1 is presented in Section 2. The proofs of the application results are in Section 3. Finally, since the weak regularity lemma that we use in the proof of Theorem 1 is not completely standard, we include its proof in Section 4 for completeness.

2 Proof of Theorem 1

We prove the two parts (sufficiency and necessity) of Theorem 1 in Sections 2.1 and 2.2, respectively. In the proofs, the *density* of a function $f\colon \mathcal{X} \to \{0,1\}$ in a set $S \subseteq \mathcal{X}$ is $\mu_S(f) = \mathbb{E}_{x\in S}[f(x)]$, the expected value of $f(x)$ when x is drawn uniformly at random from S.

2.1 Symmetry implies testability

We begin the proof of Theorem 1 with the easy direction.

Lemma 2. *Let $\mathcal{P} \subseteq \{0,1\}^{\mathcal{X}}$ be a property where for every $\epsilon > 0$, there exists a constant $k = k_{\mathcal{P}}(\epsilon)$ that is independent of $|\mathcal{X}|$ and a k-part symmetric property $\mathcal{P}'$ such that $\mathcal{P} \subseteq \mathcal{P}' \subseteq \mathcal{P}_\epsilon$. Then $\mathcal{P}$ is constant-sample testable.*

Proof. Fix $\epsilon > 0$. We show that we can distinguish functions in $\mathcal{P}$ from functions that are ϵ-far from $\mathcal{P}$ with a constant number of samples. From the premise of the lemma, there exists a k-part symmetric property $\mathcal{P}'$ with $\mathcal{P} \subseteq \mathcal{P}' \subseteq \mathcal{P}_{\epsilon/2}$ for some $k = k_{\mathcal{P}}(\epsilon/2)$. Let $X_1, \ldots, X_k$ be a partition of $\mathcal{X}$ such that whether a function $f\colon \mathcal{X} \to \{0,1\}$ satisfies $\mathcal{P}'$ is determined by $\mu_{X_1}(f), \ldots, \mu_{X_k}(f)$. For a set $S \subseteq \mathcal{X}$, let $c_S(f) = \mu_S(f)|S|$ be the number of $x \in S$ with $f(x) = 1$.

Our algorithm for testing $\mathcal{P}$ is as follows. For each $i \in [k]$, we draw $q := O(k^2 \log k/\epsilon^2)$ samples $x_1, \ldots, x_q$ and compute the estimates $\widetilde{c}_{X_i}(f) := \frac{|\mathcal{X}|}{q} \sum_{j\in[q]: x_j \in X_i} f(x_j)$ for each $i \in [k]$. We accept if there exists $g \in \mathcal{P}'$ such that

$$\sum_{i\in[k]} |\widetilde{c}_{X_i}(f) - c_{X_i}(g)| < \frac{\epsilon}{4}|\mathcal{X}|,$$

and reject otherwise.

Let us now establish the correctness of the algorithm. By Hoeffding's bound, for each $i \in [k]$, we have $|c_{X_i}(f) - \widetilde{c}_{X_i}(f)| < \frac{\epsilon}{4k}|\mathcal{X}|$ with probability at least $1 - \frac{1}{3k}$ by choosing the hidden constant in the definition of q sufficiently large. By union bound, with probability at least $2/3$, we have $|c_{X_i}(f) - \widetilde{c}_{X_i}(f)| < \frac{\epsilon}{4k}|\mathcal{X}|$ for every $i \in [k]$. In what follows, we assume this inequality holds.

If $f \in \mathcal{P}$, then the algorithm accepts f because $\sum_{i\in[k]} |\widetilde{c}_{X_i}(f) - c_{X_i}(f)| < \frac{\epsilon}{4}|\mathcal{X}|$ and $f \in \mathcal{P} \subseteq \mathcal{P}'$.

If f is ϵ-far from satisfying $\mathcal{P}$, then for any $g \in \mathcal{P}'$, the triangle inequality and the fact that $\mathcal{P}' \subseteq \mathcal{P}_{\epsilon/2}$ imply that

$$\sum_{i\in[k]} |\widetilde{c}_{X_i}(f) - c_{X_i}(g)| \geq \sum_{i\in[k]} \Big(|c_{X_i}(f) - c_{X_i}(g)| - |\widetilde{c}_{X_i}(f) - c_{X_i}(f)|\Big) > \frac{\epsilon}{2}|\mathcal{X}| - \frac{\epsilon}{4}|\mathcal{X}| = \frac{\epsilon}{4}|\mathcal{X}|$$

and, therefore, the algorithm rejects f. □

2.2 Testability implies symmetry

Suppose that a property $\mathcal{P}$ is testable by a tester $\mathcal{T}$ with sample complexity s. We want to show that for any $\epsilon > 0$, there exists a k-part symmetric property $\mathcal{P}'$ for $k = k(\epsilon)$ such that $\mathcal{P} \subseteq \mathcal{P}' \subseteq \mathcal{P}_\epsilon$

For any $\mathbf{x} = (x_1, \ldots, x_s) \in \mathcal{X}^s$, we define $f(\mathbf{x}) = \big(f(x_1), \ldots, f(x_s)\big)$ and we let $T(\mathbf{x}, f(\mathbf{x})) \in [0,1]$ denote the acceptance probability of the tester $\mathcal{T}$ of f when the samples drawn are $\mathbf{x}$. The overall acceptance probability of f by $\mathcal{T}$ is

$$p_{\mathcal{T}}(f) = \mathop{\mathbb{E}}_{\mathbf{x}}[T(\mathbf{x}, f(\mathbf{x}))].$$

We show that there is a family $\mathcal{S}$ of a constant number of subsets of $\mathcal{X}$ such that, for every function $f\colon \mathcal{X} \to \{0,1\}$, the acceptance probability $p_{\mathcal{T}}(f)$ is almost completely determined by the density of f on the subsets in $\mathcal{S}$.

Lemma 3. *For any $\gamma > 0$ and any s-sample tester $\mathcal{T}$, there is a family $\mathcal{S} = \{S_1, \ldots, S_m\}$ of $m \le 2^{O(2^{2s}/\gamma^2)}$ subsets of $\mathcal{X}$ and a function $\varphi_{\mathcal{T}}\colon [0,1]^m \to [0,1]$ such that for every $f\colon \mathcal{X} \to \{0,1\}$,*

$$\left|p_{\mathcal{T}}(f) - \varphi_{\mathcal{T}}\big(\mu_{S_1}(f), \ldots, \mu_{S_m}(f)\big)\right| \le \gamma.$$

Proof. Consider the weighted hypergraph $G = (V, \xi)$ defined by setting $V = \mathcal{X} \times \{0,1\}$ and letting ξ be constructed by adding a hyperedge $((x_1, y_1), \ldots, (x_s, y_s))$ of weight $T(\mathbf{x}, \mathbf{y})$ for each $\mathbf{x} = (x_1, \ldots, x_s) \in \mathcal{X}^s$ and $\mathbf{y} = (y_1, \ldots, y_s) \in \{0,1\}^s$. Let $V_1, \ldots, V_k$ be the partition of V guaranteed to exist by Lemma 1 with the approximation parameter $\epsilon = \gamma/2^s$.

A function $f\colon \mathcal{X} \to \{0,1\}$ corresponds to the subset $U \subseteq V$ defined by $U := \{(x, f(x)) \mid x \in \mathcal{X}\}$. The probability that $\mathcal{T}$ accepts f is

$$p_{\mathcal{T}}(f) = \mathop{\mathbb{E}}_{\mathbf{x} \in \mathcal{X}^s}[T(\mathbf{x}, f(\mathbf{x}))] = \mathop{\mathbb{E}}_{\mathbf{v} \in V^s}[\xi(\mathbf{v}) \mid \mathbf{v} \in U^s] = \mathop{\mathbb{E}}_{\mathbf{v} \in U^s}[\xi(\mathbf{v})]$$

where the last expectation is over the uniform distribution of $\mathbf{v}$ in U^s. Since $|U| = |V|/2$, we observe that

$$\mathop{\mathbb{E}}_{\mathbf{v} \in U^s}[\xi(\mathbf{v})] = \sum_{I \in [k]^s} \frac{\prod_{i \in I} |V_i \cap U|}{|U|^s} w_G(U \cap V_I) = 2^s \sum_{I \in [k]^s} \frac{\prod_{i \in I} |V_i \cap U|}{|V|^s} w_G(U \cap V_I).$$

Using the conclusion of Lemma 1, we then obtain

$$\left| p_{\mathcal{T}}(f) - 2^s \sum_{I \in [k]^s} \frac{\prod_{i \in I} |V_i \cap U|}{|V|^s} w_G(V_I) \right| \le 2^s \sum_{I \in [k]^s} \frac{\prod_{i \in I} |V_i \cap U|}{|V|^s} \left| w_G(U \cap V_I) - w_G(V_I) \right| \le \gamma.$$

For every part V_i, let $V_i^1 = \{x \in \mathcal{X} : (x,1) \in V_i\}$ and let $V_i^0 = \{x \in \mathcal{X} : (x,0) \in V_i\}$. Then the value of $2^s \sum_{I \in [k]^s} \frac{\prod_{i \in I} |V_i \cap U|}{|V|^s} w_G(V_I)$ is completely determined by the density of f on $V_1^1, V_1^0, V_2^1, V_2^0, \ldots, V_k^1, V_k^0$. □

We are now ready to complete the second part of the proof of Theorem 1.

Lemma 4. *Fix any property $\mathcal{P} \subseteq \{0,1\}^{\mathcal{X}}$ and any $\epsilon > 0$. If there is an ϵ-tester for $\mathcal{P}$ with sample complexity $s \ge 1$, then there exists a value $k = 2^{2^{2^{O(s)}}}$ and a k-part symmetric property $\mathcal{P}'$ such that $\mathcal{P} \subseteq \mathcal{P}' \subseteq \mathcal{P}_\epsilon$.*

Proof. Let $\mathcal{T}$ be an ϵ-tester for $\mathcal{P}$ with sample complexity s and let γ be any constant that is less than $\frac{1}{6}$. By Lemma 3 applied with the parameter γ, there is a family $\mathcal{S} = \{S_1, \ldots, S_m\}$ with $m = 2^{O(2^{2s}/\gamma^2)}$ sets such that for every $f\colon \mathcal{X} \to \{0,1\}$,

$$|p_{\mathcal{T}}(f) - \varphi_{\mathcal{T}}(\mu_{S_1}(f), \ldots, \mu_{S_k}(f))| \le \gamma. \tag{1}$$

Define

$$\mathcal{P}' = \{f\colon \mathcal{X} \to \{0,1\} : \exists g \in \mathcal{P} \text{ s.t. } (\mu_{S_1}(f), \ldots, \mu_{S_k}(f)) = (\mu_{S_1}(g), \ldots, \mu_{S_k}(g))\}.$$

This construction trivially guarantees that $\mathcal{P}' \supseteq \mathcal{P}$. Furthermore, the inequality (1) guarantees that for every $f \in \mathcal{P}'$, if we let $g \in \mathcal{P}$ be one of the elements with the same density profile as f,

$$p_{\mathcal{T}}(f) \geq \varphi_{\mathcal{T}}(\mu_{S_1}(f), \ldots, \mu_{S_k}(f)) - \gamma = \varphi_{\mathcal{T}}(\mu_{S_1}(g), \ldots, \mu_{S_k}(g)) - \gamma \geq p_{\mathcal{T}}(g) - 2\gamma.$$

Since $\mathcal{T}$ is an ϵ-tester for $\mathcal{P}$ and $g \in \mathcal{P}$, we must have $p_{\mathcal{T}}(g) \geq \frac{2}{3}$. By our choice of $\gamma < 1/6$, this means that we also have $p_{\mathcal{T}}(f) > \frac{2}{3} - 2 \cdot \frac{1}{6} = \frac{1}{3}$ and so, again using the fact that $\mathcal{T}$ is an ϵ-tester for $\mathcal{P}$, we must have that $f \in \mathcal{P}_\epsilon$.

Let $S'_1, \ldots, S'_k$ be the family of sets obtained by taking intersections and complements of $S_1, \ldots, S_m$. Note that $S'_1, \ldots, S'_k$ forms a partition of $\mathcal{X}$ and $\mu_{S_1}, \ldots, \mu_{S_m}$ is completely determined by $\mu_{S'_1}, \ldots, \mu_{S'_k}$. Furthermore, $k = O(2^m)$. Hence, $\mathcal{P}'$ is a k-part symmetric property induced by the partition $S'_1, \ldots, S'_k$ with $\mathcal{P} \subseteq \mathcal{P}' \subseteq \mathcal{P}_\epsilon$. $\square$

Theorem 1 follows immediately from Lemmas 2 and 4.

3 Applications

Corollaries 1 and 2 regarding constant-sample testable graph properties and affine-invariant properties, respectively, both follow directly from a more general result concerning constant-sample testable properties that are invariant under "mixing" sets of permutations. We present the general result in Section 3.1 and the proofs of the two corollaries in Sections 3.2 and 3.3. The proof of Corollary 3 regarding monotonicity testing is presented in Section 3.4.

3.1 Properties invariant under mixing groups of permutations

Definition 2. A group Π of permutations on a finite set $\mathcal{X}$ is (γ, τ, δ)-*mixing* if for every set $S \subseteq \mathcal{X}$ of cardinality $|S| \geq \delta|\mathcal{X}|$ and every function $f\colon \mathcal{X} \to \{0,1\}$,

$$\Pr_{\pi \in \Pi}\left[|\mu_S(\pi f) - \mu_{\mathcal{X}}(f)| > \tau\right] < \gamma.$$

Theorem 2. *Fix $\epsilon > 0$ and $s \geq 1$, and let $k = 2^{2^{2^{O(s)}}}$ be the constant in Lemma 4. Let Π be a $(\frac{1}{2k}, \frac{\epsilon}{8}, \frac{\epsilon}{4k})$-mixing group of permutations over the set $\mathcal{X}$. Then for every Π-invariant property $\mathcal{P} \subseteq \{0,1\}^{\mathcal{X}}$ that is $\frac{\epsilon}{2}$-testable with s samples, there exists a symmetric property $\mathcal{P}^{\mathrm{sym}}$ that satisfies $\mathcal{P} \subseteq \mathcal{P}^{\mathrm{sym}} \subseteq \mathcal{P}_\epsilon$.*

Proof. By Lemma 4, there is a partition $S_1, \ldots, S_k$ of $\mathcal{X}$ with k parts and a property $\mathcal{P}'$ that is invariant under permutations of $S_1, \ldots, S_k$ such that $\mathcal{P} \subseteq \mathcal{P}' \subseteq \mathcal{P}_{\epsilon/2}$. Define $\mathcal{P}^{\mathrm{sym}} = \{g : \exists f \in \mathcal{P} \text{ s.t. } \mu_{\mathcal{X}}(f) = \mu_{\mathcal{X}}(g)\}$ to be the closure of $\mathcal{P}$ over all permutations of $\mathcal{X}$. By construction, this property is symmetric and contains $\mathcal{P}$. To complete the proof, we want to show that every $g \in \mathcal{P}^{\mathrm{sym}}$ is also contained in $\mathcal{P}_\epsilon$.

Fix any function $f \in \mathcal{P}$ and any function $g\colon \mathcal{X} \to \{0,1\}$ that satisfies $\mu_{\mathcal{X}}(f) = \mu_{\mathcal{X}}(g)$. For any part S_i of size $|S_i| \geq \frac{\epsilon}{4k}|\mathcal{X}|$, the mixing property of Π guarantees that

$$\Pr_{\pi \in \Pi}\left[|\mu_{S_i}(\pi f) - \mu_{S_i}(\pi g)| > \tfrac{\epsilon}{4}\right] \leq \Pr_{\pi \in \Pi}\left[|\mu_{S_i}(\pi f) - \mu_{\mathcal{X}}(f)| > \tfrac{\epsilon}{8}\right] + \Pr_{\pi \in \Pi}\left[|\mu_{S_i}(\pi g) - \mu_{\mathcal{X}}(g)| > \tfrac{\epsilon}{8}\right] < \frac{1}{k}.$$

By the union bound,

$$\Pr_{\pi\in\Pi}\left[\bigwedge_{i\le k\,:\,|S_i|\ge\frac{\epsilon}{4k}|\mathcal{X}|}\left|\mu_{S_i}(\pi f)-\mu_{S_i}(\pi g)\right|\le\frac{\epsilon}{4}\right]>1-k\cdot\frac{1}{k}=0$$

so there must exist a permutation $\pi^\star\in\Pi$ for which

$$\sum_{i=1}^{k}\frac{|S_i|}{|\mathcal{X}|}\left|\mu_{S_i}(\pi^\star f)-\mu_{S_i}(\pi^\star g)\right|\le\sum_{i\le k\colon|S_i|<\frac{\epsilon}{4k}|\mathcal{X}|}\frac{|S_i|}{|\mathcal{X}|}\cdot 1+\sum_{i\le k\colon|S_i|\ge\frac{\epsilon}{4k}|\mathcal{X}|}\frac{|S_i|}{|\mathcal{X}|}\cdot\frac{\epsilon}{4}\le k\cdot\frac{\epsilon}{4k}+\frac{\epsilon}{4}=\frac{\epsilon}{2}.$$

For each $i\in[k]$, we can find a permutation σ_i on $\mathcal{X}$ that is the identity outside of S_i for which $\Pr_{x\in S_i}[\sigma_i\pi^\star f(x)\ne\pi^\star g(x)]=|\mu_{S_i}(\pi^\star f)-\mu_{S_i}(\pi^\star g)|$. Define $h=\sigma_k\sigma_{k-1}\cdots\sigma_1\pi^\star f$. Since $\mathcal{P}'$ is invariant under permutations of $S_1,\ldots,S_k$, and $\pi^\star f\in\mathcal{P}\subseteq\mathcal{P}'$, the function h is in $\mathcal{P}'$. Also, the definition of h guarantees that $\Pr_{x\in S_i}[h(x)\ne\pi^\star g(x)]=|\mu_{S_i}(\pi^\star f)-\mu_{S_i}(\pi^\star g)|$ for every $i\in[k]$, so

$$\begin{aligned}\mathrm{d}(\pi^\star g,h)=\Pr_{x\in\mathcal{X}}[\pi^\star g(x)\ne h(x)]&=\sum_{i=1}^{k}\Pr_{x\in\mathcal{X}}[x\in S_i]\cdot\Pr_{x\in S_i}[\pi^\star g(x)\ne h(x)]\\&=\sum_{i=1}^{k}\frac{|S_i|}{|\mathcal{X}|}\cdot\left|\mu_{S_i}(\pi^\star f)-\mu_{S_i}(\pi^\star g)\right|\le\frac{\epsilon}{2}\end{aligned}$$

and $\pi^\star g$ is $\frac{\epsilon}{2}$-close to $\mathcal{P}'$. Therefore, $\pi^\star g$ is ϵ-close to $\mathcal{P}$. The property $\mathcal{P}$ is Π-invariant, so $g\in\mathcal{P}_\epsilon$ as well. ☐

3.2 Graph properties

A permutation $\pi\colon V\to V$ on the set of vertices acts on a graph $G=(V,E)$ over the same set of vertices by having $\pi G=(V,\pi E)$ be the graph with edge set $\pi E=\{(\pi u,\pi v):(u,v)\in E\}$. A property $\mathcal{P}\subseteq\{0,1\}^{\binom{V}{2}}$ is a *graph property* if for every $G\in\mathcal{P}$ and every permutation $\pi\colon V\to V$ of the vertex set, $\pi G\in\mathcal{P}$ as well. The density of a graph $G=(V,E)$ in a set $S\subseteq\binom{V}{2}$ is denoted $\mu_S(G)=\frac{|E\cap S|}{|S|}$ and the overall density of G is $\mu(G):=\mu_{\binom{V}{2}}(G)=|E|/\binom{|V|}{2}$.

Corollary 1 follows immediately from the following result.

Corollary 4. *For every $\epsilon>0$ and any integer $s\ge 1$, there is a positive integer $n_0=n_0(\epsilon,s)$ such that for every $n\ge n_0$, if $\mathcal{P}$ is a graph property on graphs over $|V|\ge n_0$ vertices that is $\frac{\epsilon}{2}$-testable with s samples, then there is a symmetric property $\mathcal{P}^{\mathrm{sym}}$ such that $\mathcal{P}\subseteq\mathcal{P}^{\mathrm{sym}}\subseteq\mathcal{P}_\epsilon$.*

Proof. By Theorem 2, it suffices to show that the group Π of permutations on $\binom{V}{2}$ defined by the set of vertex permutations is $(\frac{1}{2k},\frac{\epsilon}{8},\frac{\epsilon}{4k})$-mixing. Set $n_0=\lceil\frac{1024k^2}{\epsilon^3}\rceil+2$, and fix any graph $G=(V,E)$ with at least n_0 vertices and any set $S\subseteq\binom{V}{2}$ of cardinality $|S|\ge\frac{\epsilon}{4k}\binom{|V|}{2}$. The expected density of πG in S is

$$\mathop{\mathbb{E}}_{\pi}[\mu_S(\pi G)]=|S|^{-1}\sum_{e\in E}\Pr_{\pi}[\pi e\in S]=|S|^{-1}\cdot|E|\cdot\frac{|S|}{\binom{|V|}{2}}=\mu(G).$$

The expected value of the square of the density of πG in S satisfies

$$\begin{aligned}\mathop{\mathbb{E}}_{\pi}[\mu_S(\pi G)^2] &= |S|^{-2} \sum_{e,e' \in E} \Pr_{\pi}[\pi e \in S \wedge \pi e' \in S] \\ &= |S|^{-2}\Big(\sum_{e \in E} \Pr_{\pi}[\pi e \in S] + \sum_{(u,v),(u,v') \in E:\ v \neq v'} \Pr_{\pi}[(\pi u, \pi v) \in S \wedge (\pi u, \pi v') \in S] \\ &\qquad + \sum_{(u,v),(u',v') \in E:\ u,u',v,v' \text{ all distinct}} \Pr_{\pi}[(\pi u, \pi v) \in S \wedge (\pi u', \pi v') \in S]\Big).\end{aligned}$$

For every edge $e \in E$, $\Pr_\pi[\pi e \in S] = |S|/\binom{|V|}{2}$. For every distinct $u, v, v' \in V$, the path $(\pi v, \pi u, \pi v')$ of length two is distributed uniformly among all $|V|(|V|-1)(|V|-2)$ paths of length two in $\binom{V}{2}$. Since there are at most $|S| \cdot (|V|-2)$ paths of length two in S,

$$\Pr_{\pi}[(\pi u, \pi v) \in S \wedge (\pi u, \pi v') \in S] \leq \frac{|S| \cdot (|V|-2)}{|V|(|V|-1)(|V|-2)} = \frac{|S|}{2\binom{|V|}{2}}.$$

Finally, for any distinct $u, u', v, v' \in V$, we can use the identity $\binom{n}{2}/\binom{n-2}{2} = 1 + \frac{2}{n-2}$ to obtain

$$\Pr_{\pi}[(\pi u, \pi v) \in S \wedge (\pi u', \pi v') \in S] \leq \frac{|S|(|S|-1)}{\binom{|V|}{2} \cdot \binom{|V|-2}{2}} \leq \frac{|S|^2}{\binom{|V|}{2}^2}\Big(1 + \frac{2}{|V|-2}\Big).$$

Therefore,

$$\begin{aligned}\mathrm{Var}[\mu_S(\pi G)^2] &= \mathop{\mathbb{E}}_{\pi}[\mu_S(\pi G)^2] - \mathop{\mathbb{E}}_{\pi}[\mu_S(\pi G)]^2 \\ &\leq |S|^{-2}\Big(|E| \cdot \frac{|S|}{\binom{|V|}{2}} + |E|(|V|-2) \cdot \frac{|S|}{2\binom{|V|}{2}} + |E|^2 \frac{|S|^2}{\binom{|V|}{2}^2}\big(1 + \frac{2}{|V|-2}\big)\Big) - \mu(G)^2 \\ &= \frac{\mu(G)\,|V|}{2|S|} + \frac{2\,\mu(G)^2}{|V|-2} \leq \frac{|V|}{2|S|} + \frac{2}{|V|-2} \leq \frac{8k}{\epsilon(n_0-2)} \leq \frac{\epsilon^2}{128k}.\end{aligned}$$

Then by Chebyshev's inequality, $\Pr_\pi\left[|\mu_S(\pi G) - \mu(G)| > \frac{\epsilon}{8}\right] < \frac{1}{2k}$ and the group Π of vertex permutations satisfies the desired mixing condition. $\square$

3.3 Affine-invariant properties

For an affine transformation $A : \mathbb{F}_q^n \to \mathbb{F}_q^n$ and a function $f \colon \mathbb{F}_q^n \to \{0,1\}$, we define $Af \colon \mathbb{F}_q^n \to \{0,1\}$ to be the function that satisfies $Af(x) = f(Ax)$ for every $x \in \mathbb{F}_q^n$. We define the action of an affine transformation A on a property $\mathcal{P}$ of functions mapping $\mathbb{F}_q^n \to \{0,1\}$ to be $A\mathcal{P} = \{Af : f \in \mathcal{P}\}$. A property $\mathcal{P}$ of functions $f \colon \mathbb{F}_q^n \to \{0,1\}$ is *affine-invariant* if $A\mathcal{P} = \mathcal{P}$ for every affine transformation A. The density of a function $f \colon \mathbb{F}_q^n \to \{0,1\}$ is $\mu(f) = |f^{-1}(1)|/|\mathbb{F}_q^n|$.

Corollary 2 follows immediately from the following result.

Corollary 5. *For every $\epsilon > 0$, any prime power q, and any integer $s \geq 1$, there is a positive integer $n_0 = n_0(\epsilon, q, s)$ such that for every $n \geq n_0$, if $\mathcal{P}$ is an affine-invariant property of functions $f \colon \mathbb{F}_q^n \to \{0,1\}$ that is $\frac{\epsilon}{2}$-testable with s samples, then there is a symmetric property $\mathcal{P}^{\mathrm{sym}}$ such that $\mathcal{P} \subseteq \mathcal{P}^{\mathrm{sym}} \subseteq \mathcal{P}_\epsilon$.*

Proof. By Theorem 2, it suffices to show that the group of affine transformations on $\mathbb{F}_q^n$ is $(\frac{1}{2k}, \frac{\epsilon}{8}, \frac{\epsilon}{4k})$-mixing. Set n_0 to be the smallest integer that satisfies $q^{n_0} > \frac{512k^2}{\epsilon^3}$, and fix any function $f \colon \mathbb{F}_q^n \to \{0,1\}$ with $n \geq n_0$ and any set $S \subseteq \mathbb{F}_q^n$ of cardinality $|S| \geq \frac{\epsilon}{4k}|\mathbb{F}_q^n|$. When A is drawn uniformly at random from the set of affine transformations on $\mathbb{F}_q^n$, the expected value of the density of Af in S is

$$\mathop{\mathbb{E}}_A[\mu_S(Af)] = |S|^{-1} \sum_{x \in f^{-1}(1)} \Pr_A[Ax \in S] = |S|^{-1} \cdot |f^{-1}(1)| \cdot \frac{|S|}{|\mathbb{F}_q^n|} = \mu(f)$$

and the expected value of the squared density of Af on S satisfies

$$\begin{aligned}
\mathop{\mathbb{E}}_A\left[\mu_S(Af)^2\right] &= |S|^{-2} \mathop{\mathbb{E}}_A\left[|(Af)^{-1}(1) \cap S|^2\right] \\
&= |S|^{-2} \sum_{x,x' \in f^{-1}(1)} \Pr_A[Ax \in S \wedge Ax' \in S] \\
&= |S|^{-2}\Big(\sum_{x \in f^{-1}(1)} \Pr[Ax \in S] + \sum_{x \neq x' \in f^{-1}(1)} \Pr_A[Ax \in S \wedge Ax' \in S]\Big).
\end{aligned}$$

The uniform distribution over the set of all affine transformations has the property that every pair of elements $x \neq x' \in \mathbb{F}_q^n$ satisfies $\Pr_A[Ax \in S] = \frac{|S|}{|\mathbb{F}_q^n|}$ and $\Pr_A[Ax \in S \wedge Ax' \in S] = \left(\frac{|S|}{|\mathbb{F}_q^n|}\right)^2$ so the variance of $\mu_S(Af)$ is bounded above by

$$\begin{aligned}
\mathop{\mathrm{Var}}_A[\mu_S(Af)] &= \mathop{\mathbb{E}}_A\left[\mu_S(Af)^2\right] - \mathop{\mathbb{E}}_A\left[\mu_S(Af)\right]^2 \\
&\leq \frac{1}{|S|^2}\left(|f^{-1}(1)| \cdot \frac{|S|}{|\mathbb{F}_q^n|} + |f^{-1}(1)|^2 \cdot \left(\frac{|S|}{|\mathbb{F}_q^n|}\right)^2\right) - \mu(f)^2 = \frac{\mu(f)}{|S|} \leq \frac{1}{|S|} \leq \frac{4k}{\epsilon q^{n_0}}.
\end{aligned}$$

Then by Chebyshev's inequality, $\Pr_A\left[|\mu_S(Af) - \mu(f)| > \frac{\epsilon}{8}\right] < \frac{1}{2k}$ and the group of affine transformations over $\mathbb{F}_q^n$ satisfies the desired mixing condition. ◻

3.4 Testing monotonicity

With Theorem 1, to show that monotonicity of functions $f \colon [n]^d \to \{0,1\}$ is constant-sample testable, it suffices to identify an $O(1)$-part symmetric function that covers all the monotone functions and does not include any function that is far from monotone. This is what we do below.

Corollary 3 (Restated)**.** *For every constant $d \geq 1$ and constant $\epsilon > 0$, we can ϵ-test monotonicity of functions $f \colon [n]^d \to \{0,1\}$ on the d-dimensional hypergrid with a constant number of samples.*

Proof. For any $\epsilon > 0$, fix $k = \lceil d/\epsilon \rceil$ and let $\mathcal{R}$ be a partition of the space $[n]^d$ into k^d subgrids of side length (at most) $\lfloor \epsilon n \rfloor$ each. We identify the parts in $\mathcal{R}$ with the points in $[k]^d$. For an input $x \in [n]^d$, let $\phi_{\mathcal{R}}(x)$ denote the part of $\mathcal{R}$ that contains x.

Given some function $f \colon [n]^d \to \{0,1\}$, define the $\mathcal{R}$-*granular* representation of f to be the function $f_{\mathcal{R}} \colon [k]^d \to \{0,1,*\}$ defined by

$$f_{\mathcal{R}}(x) = \begin{cases} 0 & \text{if } \forall y \in [n]^d \text{ with } \phi_{\mathcal{R}}(y) = x, f(y) = 0 \\ 1 & \text{if } \forall y \in [n]^d \text{ with } \phi_{\mathcal{R}}(y) = x, f(y) = 1 \\ * & \text{otherwise.} \end{cases}$$

Let $\mathcal{P} = \{f\colon [n]^d \to \{0,1\} : \exists$ monotone g s.t. $f_\mathcal{R} = g_\mathcal{R}\}$ be the property that includes every function whose $\mathcal{R}$-granular representation equals that of a monotone function. By construction $\mathcal{P}$ includes all the monotone functions and is invariant under any permutations within the $O(1)$ parts of $\mathcal{R}$. To complete the proof of the corollary, we want to show that every function in $\mathcal{P}$ is ϵ-close to monotone.

Fix any $f \in \mathcal{P}$ and let $g\colon [n]^d \to \{0,1\}$ be a monotone function for which $f_\mathcal{R} = g_\mathcal{R}$. The distance between f and g is bounded by

$$\mathrm{d}(f,g) \le \frac{|g_\mathcal{R}^{-1}(*)|}{k^d}.$$

Now consider the poset P on $[k]^d$ where $x \prec y$ iff $x_i < y_i$ for every $i \in [d]$. We first observe that the set $g_\mathcal{R}^{-1}(*)$ forms an antichain on this poset, i.e. no two elements $x, y \in g_\mathcal{R}^{-1}(*)$ satisfy $x \prec y$ or $y \prec x$. Indeed, if there exist $x, y \in [k]^d$ with $x \prec y$ and $g(x) = g(y) = *$, then there exist $x', y' \in [n]^d$ such that $\phi_R(x') = x$, $\phi_R(y') = y$, $g(x') = 1$, and $g(y') = 0$. But this contradicts the monotonicity of g because $x' \le y'$ holds from $x \prec y$.

For $x \in [k]^d$, let $x_{\max} = \max_{i\in[d]} x_i$ and $x_{\min} = \min_{i\in[d]} x_i$. Define $\mathbf{1} = (1,1,1,\ldots,1) \in [k]^d$ and $S = \{x \in [k]^d : x_{\min} = 1\}$. We can partition $[k]^d$ into $|S| = k^d - (k-1)^d \le dk^{d-1}$ chains $(x, x+\mathbf{1}, x+2\cdot\mathbf{1}, \ldots, x+(x_{\max}-1)\cdot\mathbf{1})$, one for each $x \in S$. Since every antichain in P can contain at most one element from each of these chains, all the antichains in P have cardinality at most dk^{d-1}. In particular, this bound holds for the antichain $g^{-1}(*)$ so $\mathrm{d}(f,g) \le \frac{dk^{d-1}}{k^d} = \frac{d}{k} \le \epsilon$. □

4 Proof of the weak regularity lemma

4.1 Information theory

The proof we provide for Lemma 1 is information-theoretic. In this section, we will use bold fonts to denote random variables. The Shannon *entropy* of a discrete random variable $\mathbf{x}$ over the finite domain $\mathcal{X}$ is $H(\mathbf{x}) = -\sum_{x\in\mathcal{X}} \Pr[\mathbf{x} = x] \log_2 \Pr[\mathbf{x} = x]$, a value that is always bounded above by $H(\mathbf{x}) \le \log_2 |\mathcal{X}|$. The *conditional entropy* of $\mathbf{x}$ given $\mathbf{y}$ is $H(\mathbf{x} \mid \mathbf{y}) = H(\mathbf{x},\mathbf{y}) - H(\mathbf{y})$. The *mutual information* between $\mathbf{x}$ and $\mathbf{y}$ is $I(\mathbf{x}\,;\mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x} \mid \mathbf{y})$ and the *conditional mutual information* of $\mathbf{x}$ and $\mathbf{y}$ given a third random variable $\mathbf{z}$ is $I(\mathbf{x}\,;\mathbf{y} \mid \mathbf{z}) = H(\mathbf{x} \mid \mathbf{z}) - H(\mathbf{x} \mid \mathbf{y},\mathbf{z})$. The *chain rule* for mutual information states that for any random variables $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ we have $I(\mathbf{x}\,;\mathbf{y},\mathbf{z}) = I(\mathbf{x}\,;\mathbf{y}) + I(\mathbf{x}\,;\mathbf{z} \mid \mathbf{y})$. For a more detailed introduction to information theory, we recommend [10].

The one non-basic information-theoretic inequality that we use in the proof is an inequality established by Tao [25] and later refined by Ahlswede [1].

Lemma 5 (Tao [25], Ahlswede [1])**.** *Let* $\mathbf{y}$*,* $\mathbf{z}$*, and* $\mathbf{z}'$ *be discrete random variables where* $\mathbf{y} \in [-1,1]$ *and* $\mathbf{z}' = \phi(\mathbf{z})$ *for some function* ϕ*. Then*

$$\mathop{\mathbb{E}}_{\mathbf{z}}\Big[\big|\mathop{\mathbb{E}}_{\mathbf{y}}[\mathbf{y} \mid \mathbf{z}'] - \mathop{\mathbb{E}}_{\mathbf{y}}[\mathbf{y} \mid \mathbf{z}]\big|\Big] \le \sqrt{2\ln 2 \cdot I(\mathbf{y}\,;\mathbf{z} \mid \mathbf{z}')}.$$

Tao originally used his inequality to offer an information-theoretic proof of Szemerédi's (strong) regularity lemma. The proof we offer below follows (a simplified version of) the same approach. The fact that Tao's proof of the strong regularity lemma can also be applied (with simplifications) to prove the Frieze–Kannan weak regularity lemma was observed previously by Trevisan [26].

4.2 Proof of Lemma 1

For any partition $\mathcal{V} = (V_1, \ldots, V_\ell)$ of V, let $\psi_\mathcal{V} \colon V^s \to [\ell]^s$ be the function that identifies the parts that contain each of its arguments. I.e., for every index set $I \in [\ell]^s$, $\psi_\mathcal{V}^{-1}(I) = V_{I_1} \times \cdots \times V_{I_\ell}$.

Define the *irregularity* of a set $S \subseteq V$ with respect to a partition $\mathcal{V}$ of V with ℓ parts and a weighted hypergraph $G = (V, \xi)$ to be

$$\mathrm{irreg}_{G,\mathcal{V}}(S) = \sum_{I \in [\ell]^s} \frac{\prod_{j\in[s]} |S \cap V_{I_j}|}{|V|^s} \big| w_G(S \cap V_I) - w_G(V_I) \big|.$$

Proposition 1. *For any weighted hypergraph $G = (V, \xi)$, any partition $\mathcal{V} = (V_1, \ldots, V_\ell)$ of V, and any set $S \subseteq V$, if $\mathbf{v} \in V^s$ is drawn uniformly at random and $\mathcal{V}' = (V_1 \cap S, V_1 \setminus S, \ldots, V_\ell \cap S, V_\ell \setminus S)$ then*

$$\mathrm{irreg}_{G,\mathcal{V}}(S) \le \sqrt{2 \ln 2\, I\big(\xi(\mathbf{v})\,;\, \psi_{\mathcal{V}'}(\mathbf{v}) \mid \psi_\mathcal{V}(\mathbf{v})\big)}.$$

Proof. The irregularity of S with respect to G and $\mathcal{V}$ is bounded above by

$$\mathrm{irreg}_{G,\mathcal{V}}(S) \le \sum_{b \in \{0,1\}^s} \sum_{I \in [\ell]^s} \frac{\prod_{j\in[s]} |S_{b_j} \cap V_{I_j}|}{|V|^s} \big| w_G(S_b \cap V_I) - w_G(V_I) \big|$$

where we set $S_0 = S$, $S_1 = V \setminus S$, and $S_b = (S_{b_1}, \ldots, S_{b_s})$: the irregularity of S is equal to the inner sum when $b = 0^s$ and all the other terms in the outer sum are non-negative. For any $I \in [\ell]^s$ and $b \in \{0,1\}^s$, $\frac{\prod_{j\in[s]} |S_{b_j} \cap V_{I_j}|}{|V|^s} = \Pr_\mathbf{v}\left[\psi_\mathcal{V}(\mathbf{v}) = I \wedge 1_S(\mathbf{v}) = b\right]$, where $1_S \colon V^s \to \{0,1\}^s$ is the indicator function for S. Similarly, $w_G(S_b \cap V_I) = \mathbb{E}_\mathbf{v}[\xi(\mathbf{v}) \mid \psi_\mathcal{V}(\mathbf{v}) = I, 1_S(\mathbf{v}) = b]$ and $w_G(V_I) = \mathbb{E}_\mathbf{v}[\xi(\mathbf{v}) \mid \psi_\mathcal{V}(\mathbf{v}) = I]$, so

$$\mathrm{irreg}_{G,\mathcal{V}}(S) \le \mathop{\mathbb{E}}_{\psi_\mathcal{V}(\mathbf{v}), 1_S(\mathbf{v})} \left[\left| \mathop{\mathbb{E}}_\mathbf{v}\left[\xi(\mathbf{v}) \mid \psi_\mathcal{V}(\mathbf{v}), 1_S(\mathbf{v})\right] - \mathop{\mathbb{E}}_\mathbf{v}\left[\xi(\mathbf{v}) \mid \psi_\mathcal{V}(\mathbf{v})\right] \right| \right].$$

The proposition follows from the Tao–Ahlswede inequality with $\mathbf{y} = \xi(\mathbf{v})$, $\mathbf{z} = (\psi_\mathcal{V}(\mathbf{v}), 1_S(\mathbf{v}))$, and $\mathbf{z}' = \psi_\mathcal{V}(\mathbf{v})$ along with the observation that the random variable $(\psi_\mathcal{V}(\mathbf{v}), 1_S(\mathbf{v}))$ encodes the same information as $\psi_{\mathcal{V}'}(\mathbf{v})$. □

We are now ready to complete the proof of the regularity lemma.

Proof of Lemma 1. For $\tau > 0$, we say that a weighted hypergraph is *τ-granular* if the weight of each hyperedge is a multiple of τ. When proving Lemma 1, we can assume that the given hypergraph is $\frac{\epsilon}{3}$-granular. To see this, let $G' = (V, \xi')$ be the hypergraph obtained from $G = (V, \xi)$ by rounding the weight of each hyperedge to a multiple of $\frac{\epsilon}{3}$. Then for any set $S \subseteq V$ and partition $\mathcal{V} = (V_1, \ldots, V_k)$ of V,

$$\begin{aligned} \mathrm{irreg}_{\mathcal{V},G}(S) &= \sum_{I \in [k]^s} \frac{\prod_{j\in[s]} |S \cap V_{I_j}|}{|V|^s} \left| w_G(S \cap V_I) - w_G(V_I) \right| \\ &\le \sum_{I \in [k]^s} \frac{\prod_{j\in[s]} |S \cap V_{I_j}|}{|V|^s} \left(\left| w_{G'}(S \cap V_I) - w_{G'}(V_I) \right| + \frac{2\epsilon}{3} \right) \le \mathrm{irreg}_{\mathcal{V},G'}(S) + \frac{2\epsilon}{3}. \end{aligned}$$

To complete the proof of the lemma, it therefore suffices to show that for every $\frac{\epsilon}{3}$-granular graph G', there is a partition $\mathcal{V}$ with $k = 2^{O(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon})}$ parts where every set $S \subseteq V$ has irregularity at most $\frac{\epsilon}{3}$ with respect to G' and $\mathcal{V}$. In the following, let G' be any $\frac{\epsilon}{3}$-regular hypergraph.

Let $\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_a$ be a sequence of partitions of V defined by the following process. We first set $\mathcal{V}_0 = (V)$ to be the trivial partition of V. This partition has $1 = 2^0$ parts. When we have generated $\mathcal{V}_0, \ldots, \mathcal{V}_{i-1}$, with $\mathcal{V}_{i-1} = (V_1, \ldots, V_{2^{i-1}})$, we proceed as follows. If every set $S \subseteq V$ has irregularity $\mathrm{irreg}_{\mathcal{V}_{i-1},G'}(S) \le \frac{\epsilon}{3}$, then we terminate the process and set $a = i - 1$. Otherwise, we choose any set $T_i \subseteq V$ with irregularity $\mathrm{irreg}_{\mathcal{V}_{i-1},G'}(T_i) > \frac{\epsilon}{3}$ and define $\mathcal{V}_i = (V_1 \cap T_i, V_1 \setminus T_i, \ldots, V_{2^{i-1}} \cap T_i, V_{2^{i-1}} \setminus T_i)$. The resulting partition has 2^i parts. The final partition $\mathcal{V}_a$ in this process has 2^a parts and satisfies the regularity condition of the lemma; to complete the proof we want to show that the process always terminates with $a = O(\frac{1}{\epsilon^2}\log\frac{1}{\epsilon})$.

Define the *information value* of a partition $\mathcal{V}$ with respect G' to be $\mathrm{info}_{G'}(\mathcal{V}) = I\big(\xi(\mathbf{v})\,;\,\psi_{\mathcal{V}}(\mathbf{v})\big)$ where $\mathbf{v}$ is drawn uniformly at random from V^s. The information value of $\mathcal{V}_0$ is $\mathrm{info}_{G'}(\mathcal{V}_0) = 0$. And by the chain rule for mutual information, for every $0 < i \le a$, the information value of $\mathcal{V}_i$ satisfies

$$\begin{aligned}\mathrm{info}_{G'}(\mathcal{V}_i) &= I\big(\xi(\mathbf{v})\,;\,\psi_{\mathcal{V}_i}(\mathbf{v})\big) = I\big(\xi(\mathbf{v})\,;\,\psi_{\mathcal{V}_{i-1}}(\mathbf{v}), 1_{T_i}(\mathbf{v})\big)\\ &= I\big(\xi(\mathbf{v})\,;\,\psi_{\mathcal{V}_{i-1}}(\mathbf{v})\big) + I\big(\xi(\mathbf{v})\,;\,1_{T_i}(\mathbf{v}) \mid \psi_{\mathcal{V}_{i-1}}(\mathbf{v})\big)\\ &= I\big(\xi(\mathbf{v})\,;\,\psi_{\mathcal{V}_{i-1}}(\mathbf{v})\big) + I\big(\xi(\mathbf{v})\,;\,\psi_{\mathcal{V}_i}(\mathbf{v}) \mid \psi_{\mathcal{V}_{i-1}}(\mathbf{v})\big).\end{aligned}$$

By Proposition 1 and the definition of $\mathcal{V}_i$, $I\big(\xi(\mathbf{v})\,;\,\psi_{\mathcal{V}_i}(\mathbf{v}) \mid \psi_{\mathcal{V}_{i-1}}(\mathbf{v})\big) \ge \frac{\mathrm{irreg}_{G,\mathcal{V}_{i-1}}(T_i)^2}{2\ln 2} > \frac{\epsilon^2}{18\ln 2}$ and

$$\mathrm{info}_{G'}(\mathcal{V}_i) > \mathrm{info}_{G'}(\mathcal{V}_{i-1}) + \frac{\epsilon^2}{18\ln 2} \ge i \cdot \frac{\epsilon^2}{18\ln 2}.$$

Since G' is $\frac{\epsilon}{3}$-granular, every partition $\mathcal{V}$ of V has information value $\mathrm{info}_{G'}(\mathcal{V}) \le H(\xi(\mathbf{v})) \le \log(\frac{3}{\epsilon})$ so the partition refinement process must stop with $a \le \frac{18\ln 2}{\epsilon^2}\log(\frac{3}{\epsilon})$, as we wanted to show. $\square$

References

[1] Rudolf Ahlswede. The final form of Tao's inequality relating conditional expectation and conditional mutual information. *Advances in Mathematics of Communications*, 1(2):239–242, 2007.

[2] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: It's all about regularity. *SIAM Journal on Computing*, 39(1):143–167, 2009.

[3] Noga Alon, Rani Hod, and Amit Weinstein. On active and passive testing. *Combinatorics, Probability and Computing*, 25(1):1–20, 2016.

[4] Maria-Florina Balcan, Eric Blais, Avrim Blum, and Liu Yang. Active property testing. *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 21–30, 2012.

[5] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Constant-time testing and learning of image properties. *CoRR*, abs/1503.01363, 2015.

[6] Arnab Bhattacharyya, Elena Grigorescu, and Asaf Shapira. A unified framework for testing linear-invariant properties. *Random Structures & Algorithms*, 46(2):232–260, 2015.

[7] Arnab Bhattacharyya and Yuichi Yoshida. An algebraic characterization of testable Boolean CSPs. In *Proceedings of the 40th International Colloquium Conference on Automata, Languages, and Programming (ICALP)*, pages 123–134, 2013.

[8] Eric Blais, Amit Weinstein, and Yuichi Yoshida. Partially symmetric functions are efficiently isomorphism testable. *SIAM Journal on Computing*, 44(2):411–432, 2015.

[9] Xi Chen, Rocco A Servedio, and Li-Yang Tan. New algorithms and lower bounds for monotonicity testing. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 286–295, 2014.

[10] Thomas M Cover and Joy A Thomas. *Elements of information theory.* John Wiley & Sons, 2012.

[11] Eldar Fischer, Yonatan Goldhirsh, and Oded Lachish. Partial tests, universal tests and decomposability. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science (ITCS)*, pages 483–500, 2014.

[12] Eldar Fischer, Oded Lachish, and Yadu Vasudev. Trading query complexity for sample-based testing and multi-testing scalability. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1163–1182, 2015.

[13] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, Ronitt Rubinfeld, and Alex Samorodnitsky. Monotonicity testing over general poset domains. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 474–483, 2002.

[14] Alan M. Frieze and Ravi Kannan. The regularity lemma and approximation schemes for dense problems. In *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 12–20, 1996.

[15] Oded Goldreich, editor. *Property Testing: Current Research and Surveys*, volume 6390 of *LNCS*. Springer, 2010.

[16] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000.

[17] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.

[18] Oded Goldreich and Dana Ron. On sample-based testers. In *Proceedings of the 6th Conference on Innovations in Theoretical Computer Science (ITCS)*, pages 337–345, 2015.

[19] Oded Goldreich and Luca Trevisan. Three theorems regarding testing graph properties. *Random Structures & Algorithms*, 23(1):23–57, 2003.

[20] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 403–412, 2008.

[21] Michael Kearns and Dana Ron. Testing problems with sublearning sample complexity. *Journal of Computer and System Sciences*, 61(3):428–456, 2000.

[22] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[23] Claude Shannon. The synthesis of two-terminal switching circuits. *Bell System Technical Journal*, 28(1):59–98, 1949.

[24] Madhu Sudan. Invariance in property testing. In Oded Goldreich, editor, *Property Testing: Current Research and Surveys*, pages 211–227. Springer, 2010.

[25] Terence Tao. Szemerédi's regularity lemma revisited. *Contributions to Discrete Mathematics*, 1(1), 2006.

[26] Luca Trevisan. Entropy and the weak regularity lemma. https://lucatrevisan.wordpress.com/2006/09/10/entropy-and-the-weak-regularity-lemma/, 2006.

[27] Yuichi Yoshida. A characterization of locally testable affine-invariant properties via decomposition theorems. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 154–163, 2014.