Testing Monotone Continuous Distributions on High-dimensional Real Cubes*

Michał Adamaszek † Artur Czumaj ‡ Christian Sohler §

Abstract

We study the task of testing properties of probability distributions. We consider a scenario in which we have access to independent samples of an unknown distribution $\mathfrak{D}$ with infinite (perhaps even uncountable) support. Our goal is to test whether $\mathfrak{D}$ has a given property or it is ε-far from it (in the statistical distance, with the L_1-distance measure).

It is not difficult to see that for many natural distributions on infinite or uncountable domains, no testing algorithm can exist and the central objective of our study is to understand if there are any nontrivial distributions that can be efficiently tested. For example, it is easy to see that there is no testing algorithm that tests if a given probability distribution on $[0,1]$ is uniform. We show however, that if some additional information about the input distribution is known, testing uniform distribution is possible. We extend the recent result about testing uniformity for monotone distributions on Boolean n-dimensional cubes by Rubinfeld and Servedio (STOC'2005) to the case of *continuous* $[0,1]^n$ cubes. We show that if a distribution $\mathfrak{D}$ on $[0,1]^n$ is monotone, then one can test if $\mathfrak{D}$ is uniform with the sample complexity $\mathcal{O}(n/\varepsilon^2)$. This result is optimal up to a polylogarithmic factor.

1 Introduction.

The topic of testing basic properties of probability distributions has been extensively studied for many decades. While the traditional approach used in statistics (and also more modern approach, e.g., in data mining) has led to the development of many high quality techniques and algorithms, until very recently little attention has been paid to the computational complexity of testing in the situations when the distributions are over very large domains. Motivated by these considerations, a number of new studies have emerged that aim at developing efficient testers for various properties of distributions with the special emphasis on the small number of samples used for testing (see, e.g., [1, 3, 4, 5, 6, 10, 12] and the survey [11]). In particular, it has been shown that a number of fundamental properties, such as independence, entropy estimation, and the closeness between two distributions, can be tested in time (with the number of samples) that is sublinear in the domain size.

While these studies lead to very efficient testers for various properties of distributions with finite support, they seem not to apply when the underlying distribution is defined over an infinite or even uncountable domain. The central goal of this paper is to better understand the phenomenon of testability of such distributions.

We assume that there is a probability distribution $\mathfrak{D}$ from which we can receive independent identically distributed samples (see, e.g., [5]). We assume that each sample is of infinite precision and we will not consider the issue of representation of the real numbers. The *complexity of the tester* is measured in terms of the *number of samples* it takes in order to obtain a desired information about the distribution.

We study probability distributions over a domain Ω which will be either finite or infinite; our main focus is on the domain $\Omega = [0,1]^n$, $n \in \mathbb{N}$, that is, (continuous) n-dimensional unit cube.

Recall, by the Radon-Nikodym theorem, that every distribution on Ω has a Lebesgue decomposition (see, e.g., [9, Section 32, Theorem C]) into a sum of two parts:

- continuous (with respect to the standard Lebesgue measure), that is, given by a measurable density function $\mathfrak{f}$,
- singular (concentrated on a set of Lebesgue measure 0).

1.1 Similarity and dissimilarity between distributions: ε-farness and ε-closeness.

We study the similarity and dissimilarity between various distributions. Following the mainstream research of testing properties of distributions in theoretical computer science [3, 4, 5, 6, 11, 12], we use the *total variation dis-*

*Research supported in part by EPSRC award EP/G064679/1, DFG grant So 514/3-1, and by the Centre for Discrete Mathematics and its Applications (DIMAP), EPSRC award EP/D063191/1.

†Centre for Discrete Mathematics and its Applications (DIMAP) and Warwick Mathematics Institute, University of Warwick. *M.J.Adamaszek@warwick.ac.uk*.

‡Department of Computer Science and Centre for Discrete Mathematics and its Applications (DIMAP), University of Warwick. *A.Czumaj@warwick.ac.uk*.

§Department of Computer Science, Technische Universität Dortmund. *christian.sohler@tu-dortmund.de*.

tance to measure the similarity between distributions (L_1-distance).

Definition 1. *For any two* discrete *distributions* $\mathcal{X}$ *and* $\mathcal{Y}$ *over* Ω*, defined by the probability functions* $\mathbf{Pr}_{\mathcal{X}}$ *and* $\mathbf{Pr}_{\mathcal{Y}}$*, respectively, we say* $\mathcal{Y}$ is ε-far from $\mathcal{X}$ *if*

$$\frac{1}{2} \cdot \sum_{\omega \in \Omega} |\mathbf{Pr}_{\mathcal{X}}[\omega] - \mathbf{Pr}_{\mathcal{Y}}[\omega]| \quad \geq \quad \varepsilon \ .$$

For general distributions, the definition is analogous:

Definition 2. *For any two distributions* $\mathcal{X}$ *and* $\mathcal{Y}$ *over* Ω*, with density functions* $\mathfrak{f}_{\mathcal{X}}$ *and* $\mathfrak{f}_{\mathcal{Y}}$*, respectively, we say* $\mathcal{Y}$ is ε-far from $\mathcal{X}$ *if*

$$\text{(1.1)} \qquad \frac{1}{2} \cdot \int_{\boldsymbol{x} \in \Omega} |\mathfrak{f}_{\mathcal{X}}(\boldsymbol{x}) - \mathfrak{f}_{\mathcal{Y}}(\boldsymbol{x})| \, d\boldsymbol{x} \quad \geq \quad \varepsilon \ .$$

We say $\mathcal{Y}$ is ε-close to $\mathcal{X}$ *if* $\mathcal{Y}$ *is not* ε*-far from* $\mathcal{X}$*.*

Note that inequality (1.1) is equivalent to

$$\int_{\boldsymbol{x} \in \Omega : \mathfrak{f}_{\mathcal{X}}(\boldsymbol{x}) \geq \mathfrak{f}_{\mathcal{Y}}(\boldsymbol{x})} (\mathfrak{f}_{\mathcal{X}}(\boldsymbol{x}) - \mathfrak{f}_{\mathcal{Y}}(\boldsymbol{x})) \, d\boldsymbol{x} \quad \geq \quad \varepsilon.$$

Definition 3. *A distribution* $\mathfrak{D}$ *over* $[0,1]^n$ *with density function* $\mathfrak{f}$ *is* uniform *if* $\mathfrak{f}$ *is identically* 1*.*

For the uniform distribution, Definition 2 can be rephrased as follows.

Definition 4. *Let* $\mathfrak{D}$ *be a distribution over* $[0,1]^n$ *with density function* $\mathfrak{f}$*. We say* $\mathfrak{D}$ *is* ε-far from uniform *if*

$$\frac{1}{2} \cdot \int_{\boldsymbol{x} \in [0,1]^n} |\mathfrak{f}(\boldsymbol{x}) - 1| \, d\boldsymbol{x} \quad \geq \quad \varepsilon \ .$$

1.2 Distributions on infinite domains are usually not testable. In general, when using the total variation distance to measure the similarity between distributions, it is infeasible to investigate interesting properties of distributions on infinite domains without any assumptions on the density function. For example, let us consider the problem of testing if a distribution on $[0,1]$ is uniform. Let us suppose that there is a tester A that for some integer t distinguishes with at most t samples between uniform distribution $\mathfrak{D}_U$ on $[0,1]$ and any distribution that is ε-far from uniform. We show that this is impossible. Let us define a distribution $\mathfrak{D}$ that is $\frac{1}{2}$-far from uniform as follows: Partition $[0,1]$ into t^3 intervals of identical length $1/t^3$ each. Split each interval into two halves and then randomly choose one half. The chosen half gets in $\mathfrak{D}$ the distribution with the density function $\mathfrak{f} = 2$ and the other half has zero probability (density function $\mathfrak{f} = 0$). Now, let us consider the behavior of A on $\mathfrak{D}_U$ and $\mathfrak{D}$. Since A draws according to $\mathfrak{D}$ at most t points from $[0,1]$, no interval of length $1/t^3$ will be chosen more than once, with high probability. As the result, A will not be able to distinguish $\mathfrak{D}_U$ from $\mathfrak{D}$. This observation can be easily generalized to testing a number of natural properties for distributions on infinite domains.

Another approach that brings negative results uses the existing lower bounds for testing properties of discrete probability distributions. For example, Batu et al. [5] (see also [8]) show that testing if a given distribution on the support of size n is uniform requires $\Omega(\sqrt{n}/\varepsilon^2)$ samples. With that, by taking $n \to \infty$, the lower bound in [5] immediately implies that no algorithm can test if a given distribution on $[0,1]$ is uniform. This approach implies also similar impossibility results for testing if a given distribution is monotone, unimodal, or if two distributions are identical, are independent, and so on (see [1, 3, 4, 5, 6, 10, 12] for more examples).

Once we see these negative result, the natural question is: what properties of distributions on infinite domains can be tested?

1.3 Testing if a distribution is discrete on N points. In order to understand the problem of testing distributions on infinite domains, the very first question should be to test if a given distribution has infinite support. We first briefly consider a dual question: to verify if a given distribution has support of up to a given size. Recall, that a point $\boldsymbol{x}$ is called an *atom* of $\mathfrak{D}$ if $\mathbf{Pr}_{\mathfrak{D}}[\boldsymbol{x}] > 0$. Detection of a single atom is not possible, since its probability may be arbitrarily small, beyond the resolution of any given algorithm. Instead we may try to determine whether a large part of the probability mass is concentrated on the atoms: for a given parameter N, distinguish between distributions that have the entire support on at most N points (*discrete on N points*) and those that are ε-far from discrete on N points.

A related question has been studied recently by Raskhodnikova et al. [10], who were interested in estimating the size of the support of a given distribution under the assumption that every element in the support is an atom (distribution is singular) with the probability at least $\frac{1}{M}$. For such problem, Raskhodnikova et al. [10] (see also [13]) show that one needs at least $\Omega(M^{1-o(1)})$ samples to estimate the size of the support. On the other hand, it is easy to compute (exactly) the size of the support with $\mathcal{O}(M \log M)$ samples (e.g., by using the approach from the coupon collector problem). Our goal is different than that in [10], because on one hand, we do not have any lower bound for the probability of the points in the support (which makes the task of even estimating the size of the support impossible), and on the other hand, we want to test if a given distribution has at most N points in the support (rather than estimate the size of the support). Still, the lower bound result from Raskhodnikova et al. [10] carries over for our problem and gives a lower bound for the sample size of $\Omega(N^{1-o(1)})$ (see Theorem 2.1). It is also not difficult to prove that an algorithm sampling $\mathcal{O}(N/\varepsilon)$ elements and then checking if the sample contains more than N distinct elements is a testing algorithm that distinguishes between a discrete distribution on N points and any distribution that is ε-far from discrete on N points with $\mathcal{O}(N/\varepsilon)$ samples (see Theorem 2.2). Observe that this result immediately implies that we can estimate the smallest number $\mathcal{N}$ of points in the domain of $\mathfrak{D}$ such that $\mathfrak{D}$ has $\mathcal{N}$ points that have the total probability at least $1-\varepsilon$ using $\mathcal{O}(\mathcal{N}/\varepsilon)$ samples.

1.4 Main contribution: Testing if a monotone distribution on a real hypercube is uniform. The main goal of this paper is to investigate if there are any interesting distributions on infinite domains that are testable. One of a very few properties of discrete distributions considered in the Computer Science literature that has only a light dependency on the size of the support (the condition that by our discussion above seems to be necessary to hope for a fast tester) is that of *testing if a monotone distribution on the Boolean cube is uniform.* A probability distribution $\mathfrak{D}$ on the Boolean n-dimensional cube $\{0,1\}^n$ is *monotone* if for any $\boldsymbol{x} = (x_1,\dots,x_n)$, $\boldsymbol{y} = (y_1,\dots,y_n)$, $\mathbf{Pr}_{\mathfrak{D}}[\boldsymbol{x}] \le \mathbf{Pr}_{\mathfrak{D}}[\boldsymbol{y}]$ whenever $x_i \le y_i$ for every i. Rubinfeld and Servedio [12] consider the following problem: given a monotone (discrete) distribution $\mathfrak{D}$ on a Boolean n-dimensional cube $\{0,1\}^n$, test if $\mathfrak{D}$ is uniform. Observe that in this case the domain's size is 2^n and therefore the standard uniformity testing algorithms (cf. [5, 8]) require $\widetilde{\mathcal{O}}(2^{n/2}/\varepsilon^2)$ samples. Furthermore, Rubinfeld and Servedio [12] show that without the assumption about the monotonicity of $\mathfrak{D}$, $2^{\Omega(n)}$ samples are also necessary. However, once we assume $\mathfrak{D}$ is monotone, it is possible to test if $\mathfrak{D}$ is uniform using $\mathcal{O}(n\log(n/\varepsilon)/\varepsilon^2)$ samples [12]. Furthermore, this result is almost optimal in the sense that $\Omega(n/\log^2 n)$ samples are necessary [12].

Our main contribution is the analysis of this problem in the setting when $\mathfrak{D}$ is a monotone distribution on an n-dimensional (*real*) cube $[0,1]^n$. On high-level our approach is similar to that used by Rubinfeld and Servedio [12] in the case of Boolean n-cubes. However, the fact that we have to deal with continuous domain makes our proof of the main result, Theorem 3.1, more complicated. Furthermore, our analysis leads to an algorithm slightly faster than that from [12] (we shave off an $\mathcal{O}(\log(n/\varepsilon))$ factor) for both the Boolean and real hypercube. We observe that since the lower bound from [12] can be directly carried over to the case of real $[0,1]^n$ cubes, our upper bound is almost optimal.

Let us also notice that our tester will work with the same complexity if the input is a monotone distribution on a discrete cube $\{0,1,\dots,k\}^n$. The obtained sample size is independent of k.

2 Warm-up: Testing if distribution is discrete.

We begin our study with a simple analysis of the problem of testing if a given distribution is discrete.

DEFINITION 5. *A* probability distribution $\mathfrak{D}$ over Ω is discrete on N points *if its support is of size N, or equivalently, if there exists a set $X \subseteq \Omega$ of size at most N such that $\mathbf{Pr}_{\mathfrak{D}}[X] = 1$.*

DEFINITION 6. *A probability distribution $\mathfrak{D}$ over Ω is* ε-far from discrete on N points *if for every set $X \subseteq \Omega$ of size N we have $\mathbf{Pr}_{\mathfrak{D}}[X] < 1-\varepsilon$.*

The problem of testing the support size of a given discrete distribution has been investigated before, see, e.g., [10, 13]. The main difference between these papers and our paper is that in the formers, one assumes a lower bound for the probability of any element in the support, and this lower bound is later used in the quantitative bounds for the sample complexity of the algorithms. While this assumption allows us to carry over the lower bound from [10], this seems to be not directly applicable for the upper bound.

The first result follows easily from [10].

THEOREM 2.1. *The problem of distinguishing whether a given distribution $\mathfrak{D}$ on Ω is discrete on at most N points or is $\frac{1}{13}$-far from discrete on N points requires at least $\Omega(N^{1-o(1)})$ queries.*

Proof. Recall that the *Distinct Elements (DE) problem* is the following: given sample access to a sequence (of balls) of length n, estimate the number of distinct elements (ball colors) in the sequence. Raskhodnikova et al. [10] (see also [13] for analogous results) showed that at least $\Omega(n^{1-o(1)})$ queries are necessary to distinguish instances of DE with at most $\frac{n}{100}$ and at least $\frac{n}{11}$ colors.

Observe that any instance of DE of length n with k colors can be transformed to a distribution on k points, each point having probability a multiple of $\frac{1}{n}$. Access to this distribution is given by sampling the DE instance. Any DE instance with at most $\frac{n}{100}$ colors yields a distribution concentrated on $\frac{n}{100}$ points. On the other hand an instance of DE with at least $\frac{n}{11}$ colors yields a distribution whose distance from being concentrated on $\frac{n}{100}$ points is at least $\left(\frac{n}{11} - \frac{n}{100}\right) \cdot \frac{1}{n} > \frac{1}{13}$. Therefore, an algorithm that can distinguish such distributions can be used to separate instances of DE with at most $\frac{n}{100}$ and at least $\frac{n}{11}$ colors. ☐

The lower bound above shows a close correspondence between the problem of testing if a given distribution is discrete on N points and the DE problem. This of course suggests to verify if the upper bound analysis from [10] (see, e.g., Claim 3.5 in [10]) could be carried over for our problem. However, the requirement that in the DE problem we have a lower bound for the probability of any event, does not seem to directly transform the bounds for the DE problem in our setting. Still, we can easily show that the following simple algorithm will do the job (the proof presented here has been shown to us by Oliver Riordan; our own original proof was significantly longer):

Testing discreteness (N)**:**

- Draw a sample (according to the distribution $\mathfrak{D}$) $S = \langle s_1, \ldots, s_\ell \rangle$ from Ω with $\ell = \lceil 2N/\varepsilon \rceil$
- If S has more than N distinct elements then **Reject**
- else **Accept**

Theorem 2.2. *Algorithm* Testing discreteness (N) *is a property tester for testing if a given distribution $\mathfrak{D}$ on Ω is discrete on at most N points. Its sample complexity is $\mathcal{O}(N/\varepsilon)$.*

Proof. It is easy to see that the algorithm above will accept any probability distribution that is discrete on N elements. Therefore, now we only have to prove that if $\mathfrak{D}$ is ε-far from discrete on N points then the algorithm will reject with probability at least $\frac{2}{3}$.

Let us fix a distribution $\mathfrak{D}$ that is ε-far from discrete on N points. Let Q_t be the random variable denoting the number of distinct elements drawn by the algorithm in the first t sampling steps. Our goal is to show that $Q_\ell > N$ with probability at least $\frac{2}{3}$. Observe that $Q_1 = 1$ and conditioned on the fact that $Q_t \le N$, the assumption that $\mathfrak{D}$ is ε-far from discrete on N points implies that with probability at least ε a new point is chosen by the algorithm (that is, $\mathbf{Pr}[Q_{t+1} = Q_t + 1 \mid Q_t \le N] \ge \varepsilon$). Therefore, by Chernoff bound for binomial random variables,

$$\mathbf{Pr}[Q_\ell \le N] \quad \le \quad \mathbf{Pr}[\mathbb{B}(\ell - 1, \varepsilon) \le N] \ \le \ \tfrac{1}{3} \ ,$$

where $\mathbb{B}(n, p)$ is a binomial random variable with parameters n and p. ☐

3 Testing uniformity for monotone distributions.

We use bold letters $\boldsymbol{x}, \boldsymbol{y}$ to denote points in $[0,1]^n$. Given points $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$, we write $\boldsymbol{x} \preccurlyeq \boldsymbol{y}$ (respectively $\boldsymbol{x} \prec \boldsymbol{y}$) if $x_i \le y_i$ (respectively $x_i < y_i$) for all $i = 1, \ldots, n$. Variables range over the entire n-dimensional cube $[0,1]^n$ unless indicated otherwise, e.g., $\int_{\boldsymbol{x}} g(\boldsymbol{x}) \ d\boldsymbol{x}$ denotes the integral over $[0,1]^n$ while $\iint_{\boldsymbol{x} \prec \boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y}) \ d\boldsymbol{x} \ d\boldsymbol{y}$ is the integral over the set of all $(\boldsymbol{x}, \boldsymbol{y}) \in [0,1]^{2n}$ satisfying $\boldsymbol{x} \prec \boldsymbol{y}$.

For simplicity of presentation, we use the handy notation $\int_{\boldsymbol{x}} g(\boldsymbol{x})$ to denote $\int_{\boldsymbol{x} \in [0,1]^n} g(\boldsymbol{x}) \quad d\boldsymbol{x}$, $\iint_{\boldsymbol{x},\boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y})$ to denote $\int_{\boldsymbol{y} \in [0,1]^n} \int_{\boldsymbol{x} \in [0,1]^n} g(\boldsymbol{x}, \boldsymbol{y}) \quad d\boldsymbol{x} \quad d\boldsymbol{y}$, and $\iint_{\boldsymbol{x} \prec \boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y})$ to denote $\int_{\boldsymbol{y} \in [0,1]^n} \int_{\boldsymbol{x} \in [0,1]^n : \boldsymbol{x} \prec \boldsymbol{y}} g(\boldsymbol{x}, \boldsymbol{y}) \ d\boldsymbol{x} \ d\boldsymbol{y}$.

Definition 7. *A function $f : [0,1]^n \to \mathbb{R}$ is called* monotone *if $f(\boldsymbol{x}) \le f(\boldsymbol{y})$ whenever $\boldsymbol{x} \preccurlyeq \boldsymbol{y}$.*

Definition 8. *A* distribution over $[0,1]^n$ is monotone *if its density function is monotone.*

We consider the problem of testing if a monotone distribution over $[0,1]^n$ is uniform, that is, of distinguishing between the uniform distribution and distributions that are ε-far from uniform. The tester can query a random point from $[0,1]^n$ drawn independently at random according to a given distribution.

Before we test uniformity for monotone distributions, we first provide a characterization of monotone distributions that are ε-far from uniform. Our main theorem shows that if a monotone distribution on $[0,1]^n$ is ε-far from uniform then the expected L_1-value of a randomly chosen point from $[0,1]^n$ is at least $\frac{n}{2} + \frac{\varepsilon}{2}$.

Theorem 3.1. *Let $\mathfrak{D}$ be a monotone distribution on $[0,1]^n$ with density function $\mathfrak{f}$. If $\mathfrak{D}$ is ε-far from uniform then*

$$\mathbf{E}_{\mathfrak{f}}[\|\boldsymbol{x}\|_1] \quad = \quad \int_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \mathfrak{f}(\boldsymbol{x}) \ \ge \ \frac{n}{2} + \frac{\varepsilon}{2} \ .$$

This result is rather easy in the basic case when $n = 1$ (cf. Appendix A), but it is significantly more complicated for $n > 1$. Our analysis uses a similar approach to that used by Rubinfeld and Servedio [12] in the case of discrete n-cube. However, the fact that we have to deal with continuous domain makes the proof of the main result, Theorem 3.1 more complicated. The key technical part of the proof is formalized in Section 4, and our proof of Theorem 3.1 follows directly from Theorem 4.1 after substituting there $f(\boldsymbol{x}) = \mathfrak{f}(\boldsymbol{x}) - 1$.

Using Theorem 3.1, we can show that the following simple algorithm tests if a distribution is uniform:

Testing uniformity:

- **Repeat** $r = 20$ times:

  Draw a sample (according to the distribution $\mathfrak{D}$) $S = \langle \boldsymbol{x}_1, \ldots, \boldsymbol{x}_s \rangle$ from $[0,1]^n$ with $s = \lceil \frac{40n}{\varepsilon^2} \rceil$

  If $\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 \geq s(\frac{n}{2} + \frac{\varepsilon}{4})$ then **Reject** and exit

- **Accept**

Observe that this algorithm is slightly different from that used by Rubinfeld and Servedio [12]. While we could combine Theorem 3.1 with the algorithm from [12], the sample complexity of our algorithm is slightly better than that from [12] (it is $\mathcal{O}(n\varepsilon^{-2})$ compared to $\mathcal{O}(n\varepsilon^{-2}\ln(n/\varepsilon))$ from [12]). In fact, if one applies our algorithm to the Boolean cube $\{0,1\}^n$ then a similar improvement would be achieved as well.

In the following we will study the performance of our sampling algorithm Testing uniformity. We first prove two auxiliary lemmas, one showing that $\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1$ is not too large for the uniform distribution, and the other showing that $\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1$ is not too small for distributions that are ε-far from uniform.

LEMMA 3.1. *If we sample s points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$ independently and uniformly at random from $[0,1]^n$ (that is, with uniform distribution) with $s \geq 40n/\varepsilon^2$, then*

$$\mathbf{Pr}\Big[\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 \geq s\left(\frac{n}{2} + \frac{\varepsilon}{4}\right)\Big] \quad \leq \quad 0.01 \ .$$

Proof. Observe that the sampling is equivalent to selecting independently and uniformly at random ns random variables Y_j from $[0,1]$, after which, in the stochastic sense, $\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 = \sum_{j=1}^{sn} Y_j$. Hence, we can apply Chernoff-Hoeffding inequality [1] to obtain:

$$\begin{aligned} &\mathbf{Pr}\Big[\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 \geq s\left(\frac{n}{2} + \frac{\varepsilon}{4}\right)\Big] \\ &= \quad \mathbf{Pr}\Big[\sum_{j=1}^{sn} Y_j \geq \frac{sn}{2} + \frac{s\varepsilon}{4}\Big] \\ &\leq \quad \exp\left(\frac{-s\varepsilon^2}{8n}\right) \ . \end{aligned}$$

This implies the lemma since $s \geq 40n/\varepsilon^2$ yields $e^{-s\varepsilon^2/(8\cdot n)} \leq 0.01$. □

LEMMA 3.2. *Let $\mathfrak{D}$ be a distribution over $[0,1]^n$ that is ε-far from uniform. If $s \geq \frac{n}{6\,\varepsilon} + \frac{1}{6}$ and we sample s points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s$ chosen independently at random from the distribution $\mathfrak{D}$, then:*

$$\mathbf{Pr}\Big[\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 < s\left(\frac{n}{2} + \frac{\varepsilon}{4}\right)\Big] \quad \leq \quad \frac{12}{13} \ .$$

Proof. By Theorem 3.1, we know that $\mathbf{E}_{\mathfrak{f}}[\|\boldsymbol{x}\|_1] \geq \frac{n}{2} + \frac{\varepsilon}{2}$. Let $Y_1, \ldots, Y_s$ be independent random variables such that for every i, $Y_i = 2 - \frac{2\|\boldsymbol{x}_i\|_1}{n+\varepsilon}$. Then, by the properties of the sequence $\|\boldsymbol{x}_1\|_1, \ldots, \|\boldsymbol{x}_s\|_1$ we have $Y_i \geq 0$ and $\mathbf{E}[Y_i] \leq 1$ for every i. Therefore, we can use Feige's inequality, Theorem 1 from [7] (with $\delta = 1/12$ and $\mu_i = 1$ for every i) to conclude that $\mathbf{Pr}\big[\sum_{i=1}^{s} Y_i < s + \frac{1}{12}\big] \geq \frac{1}{13}$. This is equivalent to

$$\begin{aligned} \frac{1}{13} \quad &\leq \quad \mathbf{Pr}\Big[\sum_{i=1}^{s} Y_i < s + \tfrac{1}{12}\Big] \\ &= \quad \mathbf{Pr}\Big[\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 > s \cdot \left(\frac{n+\varepsilon}{2} - \frac{n+\varepsilon}{24s}\right)\Big] \ . \end{aligned}$$

Next, since $s \geq \frac{n}{6\,\varepsilon} + \frac{1}{6}$, we obtain,

$$\begin{aligned} \frac{1}{13} \quad &\leq \quad \mathbf{Pr}\Big[\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 > s \cdot \left(\frac{n+\varepsilon}{2} - \frac{n+\varepsilon}{24s}\right)\Big] \\ &\leq \quad \mathbf{Pr}\Big[\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 > s \cdot \left(\frac{n}{2} + \frac{\varepsilon}{4}\right)\Big] \ , \end{aligned}$$

what yields the lemma. □

THEOREM 3.2. Testing uniformity *distinguishes between uniform distribution on $[0,1]^n$ and any monotone distribution over $[0,1]^n$ that is ε-far from uniform. Its sample complexity is $\mathcal{O}(n/\varepsilon^2)$ and it errs with the probability at most $\frac{1}{4}$.*

Proof. First, by Lemma 3.1, we have:

$$\mathbf{Pr}[\mathfrak{D} \text{ is rejected} \mid \mathfrak{D} \text{ is uniform}] \quad \leq \quad r \cdot 0.01 \leq \frac{1}{4} \ .$$

Next, by Lemma 3.2, we have

$$\begin{aligned} \mathbf{Pr}[\mathfrak{D} \text{ is accepted} \mid \mathfrak{D} \text{ is } \varepsilon\text{-far from uniform}] \quad &\leq \quad \left(\frac{12}{13}\right)^r \\ &\leq \quad \frac{1}{4} \ . \end{aligned}$$

These two bounds imply the theorem. □

3.1 Lower bound.

One can directly apply the lower bound for the Boolean $\{0,1\}^n$ cube due to Rubinfeld and Servedio [12] to show that our bound is tight up to logarithmic terms.

[1] We use the following form of Chernoff-Hoeffding inequality: $\mathbf{Pr}[\sum_{j=1}^{sn} Y_j - \mathbf{E}[\sum_{j=1}^{sn} Y_j] \geq t] \leq \exp(-2t^2/(ns))$.

Theorem 3.3. *Any algorithm which given access to an unknown monotone distribution $\mathfrak{D}^*$ over $[0,1]^n$ determines correctly (with probability at least $\frac{4}{5}$) whether $\mathfrak{D}^*$ is uniform or is $\frac{1}{2}$-far from uniform must make $\Omega(n/\log^2 n)$ samples.*

Proof. The proof is obtained by considering for any distribution $\mathfrak{D}$ on $\{0,1\}^n$ a naturally defined distribution on $[0,1]^n$ that has identical properties as $\mathfrak{D}$, and then the lower bound due to Rubinfeld and Servedio [12] for testing uniformity of an unknown monotone distribution on $\{0,1\}^n$ will imply the claim.

Let $\mathfrak{D}$ be any monotone distribution with the state space $\{0,1\}^n$. Define $\psi : [0,1]^n \to \{0,1\}^n$ such that $\psi(x_1, x_2, \ldots, x_n) = (x_1^*, x_2^*, \ldots, x_n^*)$ with $x_i^* = \min\{\lfloor 2x_i \rfloor, 1\}$. Observe that $\boldsymbol{x} \preccurlyeq \boldsymbol{y}$ if and only if $\psi(\boldsymbol{x}) \preccurlyeq \psi(\boldsymbol{y})$. Now, we define a distribution $\mathfrak{D}^*$ on $[0,1]^n$ to have density function $\mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{x}) = 2^n \cdot \mathbf{Pr}_{\mathfrak{D}}[\psi(\boldsymbol{x})]$. We observe that $\mathfrak{D}$ is monotone if and only if $\mathfrak{D}^*$ is monotone, and $\mathfrak{D}$ is ε-far from uniform if and only if $\mathfrak{D}^*$ is ε-far from uniform. Therefore, the lower bound of $\Omega(n/\log^2 n)$ samples for testing uniformity of an an unknown monotone distribution $\mathfrak{D}$ over $\{0,1\}^n$ due to Rubinfeld and Servedio [12] implies the same lower bound for testing uniformity of an unknown monotone distribution $\mathfrak{D}^*$ on $[0,1]^n$. □

4 Key technical theorem and the proof of Theorem 3.1.

A proof of Theorem 3.1 can be deduced from the following theorem, by substituting $f(\boldsymbol{x}) = \mathfrak{f}(\boldsymbol{x}) - 1$.

Theorem 4.1. *Let $f : [0,1]^n \to \mathbb{R}$ be a monotone function with $\int_{\boldsymbol{x}} f(\boldsymbol{x}) = 0$. Then*

$$\int_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 f(\boldsymbol{x}) \quad \geq \quad \frac{1}{4} \int_{\boldsymbol{x}} |f(\boldsymbol{x})| \ .$$

The proof relies on two lemmas concerning *discrete cubes*. If $I_n = \{0,1\}^n$ is the standard discrete cube, then let $D(I_n)$ and $E_i(I_n)$ denote, respectively, the set of all 2^n (directed) main diagonals and the set of all directed edges in the i-th direction. Formally:

$$\begin{aligned} D(I_n) &= \{(\boldsymbol{u},\boldsymbol{v}) \in I_n \times I_n \ : \\ &\qquad u_j = 1 - v_j \text{ for } j = 1,\ldots,n\} \ , \\ E_i(I_n) &= \{(\boldsymbol{u},\boldsymbol{v}) \in I_n \times I_n \ : \\ &\qquad u_j = v_j \text{ for } j \neq i \text{ and } u_i = 1 - v_i\} \ . \end{aligned}$$

Set $E(I_n) = \bigcup_i E_i(I_n)$. We have the following lemma.

Lemma 4.1. *For any function $f : I_n \to \mathbb{R}$*

$$\sum_{(\boldsymbol{u},\boldsymbol{v}) \in D(I_n)} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \leq \sum_{(\boldsymbol{u},\boldsymbol{v}) \in E(I_n)} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \ .$$

Proof. The proof is by induction on n. For $n = 1$ both sides equal $2|f(0) - f(1)|$. For the inductive step, note that by the triangle inequality the "length" of any diagonal $(\boldsymbol{u},\boldsymbol{v})$ (understood as the value of $|f(\boldsymbol{u}) - f(\boldsymbol{v})|$) is less than or equal to the length of some edge in the n-th direction plus the length of some diagonal in one of the two $(n-1)$-dimensional sub-cubes. One can easily check that as we add these inequalities over all diagonals, every such edge and every lower-dimensional diagonal appear exactly once, thus the result is obtained by induction. We omit the details. □

Every two points $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^n$ with $\boldsymbol{x} \prec \boldsymbol{y}$ give rise to a discrete cube $K_{\boldsymbol{x},\boldsymbol{y}}$, which is the image of I_n under the obvious affine transformation sending 0^n to $\boldsymbol{x}$ and 1^n to $\boldsymbol{y}$. Let $D(\boldsymbol{x},\boldsymbol{y})$, $E_i(\boldsymbol{x},\boldsymbol{y})$ and $E(\boldsymbol{x},\boldsymbol{y})$ denote the diagonals and edges of $K_{\boldsymbol{x},\boldsymbol{y}}$, i.e., the coordinate-wise images of the respective sets associated to I_n. Lemma 4.1 immediately yields

Corollary 4.1. *For any function $f : [0,1]^n \to \mathbb{R}$ and any $\boldsymbol{x} \prec \boldsymbol{y}$*

$$\sum_{(\boldsymbol{u},\boldsymbol{v}) \in D(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \leq \sum_{(\boldsymbol{u},\boldsymbol{v}) \in E(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \ .$$

Next, we prove the following lemma.

Lemma 4.2. *For any monotone function $f : [0,1]^n \to \mathbb{R}$ and any $i = 1,\ldots,n$ we have*

$$\begin{aligned} &\iint_{\boldsymbol{x} \prec \boldsymbol{y}} \left(\sum_{(\boldsymbol{u},\boldsymbol{v}) \in E_i(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \right) \\ &\qquad = \quad 2 \int_{\boldsymbol{x}} (2x_i - 1) f(\boldsymbol{x}) \ . \end{aligned}$$

Proof. To simplify notation we only consider $i = 1$. The proof is by induction on n. For $n = 1$ we have:

$$\begin{aligned} \iint_{x<y} 2\,|f(x) - f(y)| &= 2 \iint_{x<y} (f(y) - f(x)) \\ &= 2 \left(\int_y y f(y) - \int_x (1-x) f(x) \right) \\ &= 2 \int_x (2x - 1) f(x) \ . \end{aligned}$$

Now let $n \geq 2$. In what follows $\boldsymbol{x}'$ denotes the $(n-1)$-dimensional prefix of $\boldsymbol{x}$ and we write $\boldsymbol{x} = (\boldsymbol{x}', x_n)$. For $\boldsymbol{x} = (\boldsymbol{x}', x_n)$ and $\boldsymbol{y} = (\boldsymbol{y}', y_n)$, we have:

$$\begin{aligned} &\sum_{(\boldsymbol{u},\boldsymbol{v}) \in E_1(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \\ &\quad = \sum_{(\boldsymbol{u}',\boldsymbol{v}') \in E_1(\boldsymbol{x}',\boldsymbol{y}')} \big(|f(\boldsymbol{u}', x_n) - f(\boldsymbol{v}', x_n)| + \\ &\qquad\qquad |f(\boldsymbol{u}', y_n) - f(\boldsymbol{v}', y_n)| \big) \ . \end{aligned}$$

For every fixed z, the function $f(\cdot, z)$ is monotone over the $(n-1)$-dimensional cube. Hence, we may apply the induction hypotheses to obtain

$$\begin{aligned}
&\iint_{\boldsymbol{x}\prec\boldsymbol{y}} \left(\sum_{(\boldsymbol{u},\boldsymbol{v})\in E_1(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \right) \\
&= \iint_{x_n<y_n} \iint_{\boldsymbol{x}'\prec\boldsymbol{y}'} \sum_{(\boldsymbol{u}',\boldsymbol{v}')\in E_1(\boldsymbol{x}',\boldsymbol{y}')} \\
&\quad (|f(\boldsymbol{u}', x_n) - f(\boldsymbol{v}', x_n)| + |f(\boldsymbol{u}', y_n) - f(\boldsymbol{v}', y_n)|) \\
&= 2 \iint_{x_n<y_n} \int_{\boldsymbol{x}'} (2x_1 - 1)(f(\boldsymbol{x}', x_n) + f(\boldsymbol{x}', y_n)) \\
&= 2 \int_{\boldsymbol{x}'} (2x_1 - 1) \iint_{x_n<y_n} (f(\boldsymbol{x}', x_n) + f(\boldsymbol{x}', y_n)) \\
&= 2 \int_{\boldsymbol{x}'} (2x_1 - 1) \int_{x_n} f(\boldsymbol{x}', x_n) \\
&= 2 \int_{\boldsymbol{x}} (2x_1 - 1) f(\boldsymbol{x}) \ ,
\end{aligned}$$

where we used the identity $\iint_{x<y}(h(x)+h(y)) = \int_x h(x)$, true for any $h : [0,1] \to \mathbb{R}$, with $h = h_{\boldsymbol{x}'} = f(\boldsymbol{x}', \cdot)$. This ends the proof. □

Proof of Theorem 4.1. Let $P = \{\boldsymbol{x} : f(\boldsymbol{x}) \geq 0\}$ and $N = \{\boldsymbol{x} : f(\boldsymbol{x}) < 0\}$. Consider the double integral

$$\int_{\boldsymbol{x}\in N} \int_{\boldsymbol{y}\in P} |f(\boldsymbol{x}) - f(\boldsymbol{y})| \ .$$

For $f(\boldsymbol{x}) < 0 \leq f(\boldsymbol{y})$, we have $|f(\boldsymbol{x}) - f(\boldsymbol{y})| = |f(\boldsymbol{x})| + |f(\boldsymbol{y})|$. Moreover $\int_{\boldsymbol{x}\in N} |f(\boldsymbol{x})| = \int_{\boldsymbol{y}\in P} |f(\boldsymbol{y})| = \frac{1}{2}\int_{\boldsymbol{x}} |f(\boldsymbol{x})|$. Combining this, we can express the double integral as follows:

$$\begin{aligned}
&\int_{\boldsymbol{x}\in N} \int_{\boldsymbol{y}\in P} (|f(\boldsymbol{x})| + |f(\boldsymbol{y})|) \\
&\quad = \int_{\boldsymbol{y}\in P} \int_{\boldsymbol{x}\in N} |f(\boldsymbol{x})| + \int_{\boldsymbol{x}\in N} \int_{\boldsymbol{y}\in P} |f(\boldsymbol{y})| \\
&\quad = \frac{1}{2} \int_{\boldsymbol{y}\in P} \int_{\boldsymbol{x}} |f(\boldsymbol{x})| + \frac{1}{2} \int_{\boldsymbol{x}\in N} \int_{\boldsymbol{y}} |f(\boldsymbol{y})| \\
&\quad = \frac{1}{2} \int_{\boldsymbol{y}} \int_{\boldsymbol{x}} |f(\boldsymbol{x})| \\
&\quad = \frac{1}{2} \int_{\boldsymbol{x}} |f(\boldsymbol{x})| \ .
\end{aligned}$$

At the same time, since every pair $(\boldsymbol{x}, \boldsymbol{y})$ can satisfy at most one of the conditions $(\boldsymbol{x}, \boldsymbol{y}) \in P \times N$ and $(\boldsymbol{x}, \boldsymbol{y}) \in N \times P$, we obtain the inequality

$$\int_{\boldsymbol{x}\in N} \int_{\boldsymbol{y}\in P} |f(\boldsymbol{x}) - f(\boldsymbol{y})| \quad \leq \quad \frac{1}{2} \iint_{\boldsymbol{x},\boldsymbol{y}} |f(\boldsymbol{x}) - f(\boldsymbol{y})| \ .$$

The set of points $(\boldsymbol{x}, \boldsymbol{y}) \in [0,1]^n \times [0,1]^n$ with at least one common coordinate is of measure zero in $[0,1]^n \times [0,1]^n$, while every other pair $(\boldsymbol{x}, \boldsymbol{y})$ determines a unique pair $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}})$ such that $\overline{\boldsymbol{x}} \prec \overline{\boldsymbol{y}}$ and $(\boldsymbol{x}, \boldsymbol{y})$ is a main diagonal in $K_{\overline{\boldsymbol{x}},\overline{\boldsymbol{y}}}$. By considering all the possible relative placements of $\boldsymbol{x}$ and $\boldsymbol{y}$ within $[0,1]^n$ and splitting the domain accordingly, one can see that

$$\begin{aligned}
&\iint_{\boldsymbol{x},\boldsymbol{y}} |f(\boldsymbol{x}) - f(\boldsymbol{y})| \\
&\quad = \iint_{\boldsymbol{x}\prec\boldsymbol{y}} \left(\sum_{(\boldsymbol{u},\boldsymbol{v})\in D(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \right) \ .
\end{aligned}$$

We can now combine the statements above to obtain the following sequence of inequalities

$$\begin{aligned}
&\frac{1}{2} \int_{\boldsymbol{x}} |f(\boldsymbol{x})| = \\
&\quad = \int_{\boldsymbol{x}\in N} \int_{\boldsymbol{y}\in P} |f(\boldsymbol{x}) - f(\boldsymbol{y})| \\
&\quad \leq \frac{1}{2} \iint_{\boldsymbol{x},\boldsymbol{y}} |f(\boldsymbol{x}) - f(\boldsymbol{y})| \\
&\quad = \frac{1}{2} \iint_{\boldsymbol{x}\prec\boldsymbol{y}} \left(\sum_{(\boldsymbol{u},\boldsymbol{v})\in D(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \right) \\
&\leq \frac{1}{2} \iint_{\boldsymbol{x}\prec\boldsymbol{y}} \left(\sum_{(\boldsymbol{u},\boldsymbol{v})\in E(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \iint_{\boldsymbol{x}\prec\boldsymbol{y}} \left(\sum_{(\boldsymbol{u},\boldsymbol{v})\in E_i(\boldsymbol{x},\boldsymbol{y})} |f(\boldsymbol{u}) - f(\boldsymbol{v})| \right) \\
&= \sum_{i=1}^{n} \int_{\boldsymbol{x}} (2x_i - 1) f(\boldsymbol{x}) \\
&= 2 \int_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 f(\boldsymbol{x}) \ ,
\end{aligned}$$

where the fourth bound follows from Corollary 4.1, the sixth one follows from Lemma 4.2, and we used $\int_{\boldsymbol{x}} f(\boldsymbol{x}) = 0$ in the last step. This is exactly the required inequality.

Let us observe that the constant $\frac{1}{4}$ in Theorem 4.1 is best possible since the equality holds for $f(\boldsymbol{x}) = \mathrm{sgn}(x_1 - \frac{1}{2})$.

5 Extension to discrete n-dimensional grids.

Our result in Theorem 3.2 can be easily extended to discrete n-dimensional grids, with the sample complexity independent of the size of the grid (depending only on the dimension).

Definition 9. *A distribution $\mathfrak{D}$ over $\{0,1,2,\ldots,k\}^n$ is called* monotone *if* $\mathbf{Pr}_{\mathfrak{D}}[\boldsymbol{x}] \le \mathbf{Pr}_{\mathfrak{D}}[\boldsymbol{y}]$ *whenever* $\boldsymbol{x} \preccurlyeq \boldsymbol{y}$*, where* $(x_1,\ldots,x_n) \preccurlyeq (y_1,\ldots,y_n)$ *if* $x_i \le y_i$ *for all* i*.*

Theorem 5.1. *Let k be any positive integer and consider any n-dimensional finite grid $\{0,1,2,\ldots,k\}^n$. One can test if a given monotone distribution $\mathfrak{D}$ over $\{0,1,2,\ldots,k\}^n$ is uniform with $\mathcal{O}(n/\varepsilon^2)$ samples.*

Proof. We first show that a result analogous to that in Theorem 3.1 holds and then we apply the same arguments as in Theorem 3.2 to argue that our Testing uniformity algorithm (modified accordingly to deal with values in $\{0,1,2,\ldots,k\}^n$) will work in the discrete case.

Let $\mathfrak{D}$ be any monotone distribution on $\{0,1,\ldots,k\}^n$. It is not difficult to see that if $\mathfrak{D}$ is uniform then $\mathbf{E}_{\mathfrak{D}}[\|\boldsymbol{x}\|_1] = \frac{k\cdot n}{2}$. On the other hand, one can apply Theorem 3.1 to prove that if $\mathfrak{D}$ is ε-far from uniform then

$$\begin{aligned}\mathbf{E}_{\mathfrak{D}}[\|\boldsymbol{x}\|_1] &= \sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \|\boldsymbol{x}\|_1 \cdot \mathbf{Pr}_{\mathfrak{D}}[\boldsymbol{x}]\\ &\ge \frac{kn}{2} + \frac{(k+1)\cdot\varepsilon}{2} \ . \qquad (5.2)\end{aligned}$$

Once we have (5.2), we can proceed as in Theorem 3.2 to show that the following algorithm is a tester:

Testing uniformity on a grid:

- **Repeat** $r = 20$ times:

  Draw a sample (according to the distribution $\mathfrak{D}$) $S = \langle \boldsymbol{x}_1,\ldots,\boldsymbol{x}_s\rangle$ from $\{0,1,\ldots,k\}^n$ with $s = \lceil \frac{40n}{\varepsilon^2}\rceil$

  If $\sum_{i=1}^{s} \|\boldsymbol{x}_i\|_1 \ge sk(\frac{n}{2} + \frac{\varepsilon}{4})$ then **Reject** and exit
- **Accept**

In view of the arguments above, all what we have to prove is to show inequality (5.2).

Let $\mathfrak{D}$ be any probability distribution on $\{0,1,\ldots,k\}^n$. Similarly as in the proof of Theorem 3.3, we define a coupling distribution $\mathfrak{D}^*$ on $[0,1]^n$ by partitioning $[0,1]^n$ into $(k+1)^n$ n-dimensional subcubes of the same size, assigning each subcube to the corresponding point from the grid $\{0,1,\ldots,k\}^n$ in which the density corresponds to the probability of choosing given point from the grid. More formally, let $\psi_k : [0,1]^n \to \{0,1,\ldots,k\}^n$ such that $\psi(y_1,y_2,\ldots,y_n) = (x_1,x_2,\ldots,x_n)$ with $x_i = \min\{\lfloor (k+1)\cdot y_i\rfloor, k\}$. We define a distribution $\mathfrak{D}^*$ on $[0,1]^n$ to have density function $\mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y}) = (k+1)^n \cdot \mathbf{Pr}_{\mathfrak{D}}[\psi_k(\boldsymbol{y})]$.

Now, let us observe some basic properties of $\mathfrak{D}^*$. Notice first that $\boldsymbol{x} \preccurlyeq \boldsymbol{y}$ if and only if $\psi_k(\boldsymbol{x}) \preccurlyeq \psi_k(\boldsymbol{y})$. Next, the definition of $\mathfrak{f}_{\mathfrak{D}^*}$ immediately yields $\mathbf{Pr}_{\mathfrak{D}}[\boldsymbol{x}] = \int_{\boldsymbol{y}\in[0,1]^n:\psi_k(\boldsymbol{y})=\boldsymbol{x}} \mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y})\ d\boldsymbol{y}$. With this in hand, it is not difficult to check that $\mathfrak{D}$ is monotone if and only if $\mathfrak{D}^*$ is monotone, $\mathfrak{D}$ is uniform if and only if $\mathfrak{D}^*$ is uniform, and $\mathfrak{D}$ is ε-far from uniform if and only if $\mathfrak{D}^*$ is ε-far from uniform

Now, we consider a distribution $\mathfrak{D}$ on $\{0,1,\ldots,k\}^n$ that is monotone and is ε-far from uniform. Then, since the distribution $\mathfrak{D}^*$ is also monotone and ε-far from uniform, Theorem 3.1 yields the following:

$$\begin{aligned}\sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \int_{\boldsymbol{y}\in[0,1]^n:\psi_k(\boldsymbol{y})=\boldsymbol{x}} \|\boldsymbol{y}\|_1 \cdot \mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y})\ d\boldsymbol{y}\\ = \int_{\boldsymbol{y}\in[0,1]^n} \|\boldsymbol{y}\|_1 \cdot \mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y})\ d\boldsymbol{y}\\ \ge \frac{n+\varepsilon}{2} \ . \qquad (5.3)\end{aligned}$$

Observe that the set of all $\boldsymbol{y}$ with $\psi_k(\boldsymbol{y}) = \boldsymbol{x}$ is the set of points from the cube $[0,\frac{1}{k+1})^n$ shifted by $\frac{\boldsymbol{x}}{k+1}$ (with an exception of a set of measure 0, when $x_i = k$ for some i, in which case we may have closed intervals). In that case $\|\boldsymbol{x}\|_1 = (k+1)\|\boldsymbol{y}\|_1 - \|(k+1)\cdot\boldsymbol{y} - \boldsymbol{x}\|_1$, where the last term correspond to the difference between $\boldsymbol{y}$ and the shift, and $(k+1)\cdot\boldsymbol{y} - \boldsymbol{x}$ is a point from $[0,\frac{1}{k+1})^n$. With this, we compute $\mathbf{E}_{\mathfrak{D}}[\|\boldsymbol{x}\|_1]$ under the assumption that $\mathfrak{D}$ is monotone and ε-far from uniform:

$$\begin{aligned}&\mathbf{E}_{\mathfrak{D}}[\|\boldsymbol{x}\|_1] =\\ &= \sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \|\boldsymbol{x}\|_1 \cdot \mathbf{Pr}_{\mathfrak{D}}[\boldsymbol{x}]\\ &= \sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \int_{\boldsymbol{y}\in[0,1]^n:\psi_k(\boldsymbol{y})=\boldsymbol{x}} \|\boldsymbol{x}\|_1 \cdot \mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y})\\ &= \sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \int_{\boldsymbol{y}\in[0,1]^n:\psi_k(\boldsymbol{y})=\boldsymbol{x}} ((k+1)\|\boldsymbol{y}\|_1 - \|(k+1)\boldsymbol{y}-\boldsymbol{x}\|_1)\, \mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y})\\ &= (k+1)\cdot \sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \int_{\boldsymbol{y}\in[0,1]^n:\psi_k(\boldsymbol{y})=\boldsymbol{x}} \|\boldsymbol{y}\|_1 \cdot \mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y}) -\\ &\quad - \sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \int_{\boldsymbol{y}\in[0,1]^n:\psi_k(\boldsymbol{y})=\boldsymbol{x}} \|(k+1)\cdot\boldsymbol{y}-\boldsymbol{x}\|_1 \cdot\\ &\qquad ((k+1)^n \cdot \mathbf{Pr}_{\mathfrak{D}}[\psi_k(\boldsymbol{x})])\\ &= (k+1)\cdot \int_{\boldsymbol{y}\in[0,1]^n} \|\boldsymbol{y}\|_1 \cdot \mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y})\\ &\quad -(k+1)^n \cdot \sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \mathbf{Pr}_{\mathfrak{D}}[\psi_k(\boldsymbol{x})] \int_{\boldsymbol{z}\in[0,\frac{1}{k+1})^n} \|\boldsymbol{z}\|_1\\ &= (k+1)\cdot \int_{\boldsymbol{y}\in[0,1]^n} \|\boldsymbol{y}\|_1 \cdot \mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y})\\ &\quad -(k+1)^n \cdot \sum_{\boldsymbol{x}\in\{0,1,\ldots,k\}^n} \mathbf{Pr}_{\mathfrak{D}}[\psi_k(\boldsymbol{x})] \cdot \frac{n}{2\cdot(k+1)^n}\end{aligned}$$

$$
\begin{aligned}
&= (k+1)\cdot\int_{\boldsymbol{y}\in[0,1]^n}\|\boldsymbol{y}\|_1\cdot\mathfrak{f}_{\mathfrak{D}^*}(\boldsymbol{y})-\frac{n}{2}\\
&\geq (k+1)\cdot\frac{n+\varepsilon}{2}-\frac{n}{2}\\
&= \frac{kn}{2}+\frac{(k+1)\cdot\varepsilon}{2}\ ,
\end{aligned}
$$

where the last inequality was was using (5.3). □

The result above in Theorem 5.1 can be also easily extended to the case when $\mathfrak{D}$ is defined over an n-dimensional grid of size $k_1 \times k_2 \cdots \times k_n$ for possibly distinct values of k_i.

6 Conclusions and final remarks.

In this paper we made a first attempt to understand the testability of distributions on continuous domains for basic distribution properties. We believe that understanding, which properties of continuous distributions are testable, is an interesting and challenging direction of research.

Another interesting direction to study testability of distributions on continuous domains for basic distribution properties has been proposed in a recent paper by Ba et al. [2]. Similarly as in our paper, Ba et al. observed that what makes testing distributions on continuous domains difficult is the use of the L_1-distance (or the total variation distance) to measure the similarity between the distributions that defines the notion of being ε-far from a given distribution. In view of that, Ba et al. consider another measure of the distance between two distributions, the earth mover's distance. Among other, they show that in such setting, it is possible to test if a given distribution over $[0, \Delta]^n$ is uniform, if two such distributions are identical, etc., in time independent of the input size.

Acknowledgements. We thank Oliver Riordan, Noga Alon, Mark Jerrum, Ronitt Rubinfeld, and Sofya Rashkhodnikova for their helpful comments.

References

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

[2] K. D. Ba, H. L. Nguyen, H. N. Nguyen, R. Rubinfeld. Sublinear time algorithms for earth mover's distance. Manuscript, 2009, arXiv:0904.0292v1.

[3] T. Batu, S. Dasgupta, R. Kumar, and R. Rubinfeld. The complexity of approximating the entropy. *In Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 678–687, 2002.

[4] T. Batu, E. Fischer, L. Fortnow, R. Kumar, R. Rubinfeld, and P. White. Testing random variables for independence and identity. *In Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 442–415, 2001.

[5] T. Batu, L. Fortnow, R. Rubinfeld, W. D. Smith, and P. White. Testing that distributions are close. *In Proceedings of the 41st IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 259–269, 2000.

[6] T. Batu, R. Kumar, and R. Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. *In Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 381–390, 2004.

[7] U. Feige. On sums of independent random variables with unbounded variance and estimating the average degree in a graph. *SIAM Journal on Computing*, 35(4): 964–984, 2006.

[8] O. Goldreich and D. Ron. On testing expansion in bounded-degree graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, Report No. 7, 2000.

[9] P. R. Halmos. *Measure Theory*. Springer 1974.

[10] S. Raskhodnikova, D. Ron, A. Shpilka, and A. Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM Journal on Computing*, 39(3): 813–842, 2009.

[11] R. Rubinfeld. Sublinear time algorithms. *In Proceedings of the International Congress of Mathematicians*, Madrid, Spain, August 22–30, 2006.

[12] R. Rubinfeld and R. A. Servedio. Testing monotone high-dimensional distributions. *In Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 147–156, 2005.

[13] P. Valiant. Testing symmetric properties of distributions. *In Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 383–392, 2008.

Appendix

A Warm-up: Proof of Theorem 3.1 for $n = 1$.

While the proof of Theorem 3.1 is rather long, the very simple case when $n = 1$ can be proven easily using elementary tools.

THEOREM A.1. *Let $\mathfrak{D}$ be a monotone distribution on $[0,1]$ with density function $\mathfrak{f}$. If $\mathfrak{D}$ is ε-far from uniform (that is, $\frac{1}{2}\int_0^1 |\mathfrak{f}(x) - 1| dx \geq \varepsilon$) then*

$$\mathbf{E}_{\mathfrak{f}}[x] = \int_0^1 x\ \mathfrak{f}(x)\ dx \geq \frac{1}{2} + \frac{\varepsilon}{2} .$$

Proof. Let $\sigma = \inf\{x \in [0,1] : \mathfrak{f}(x) \geq 1\}$ and let Φ be such that $\int_\sigma^1 \mathfrak{f}(x)\ dx = 1 - \sigma + \Phi$. Then also $\int_0^\sigma \mathfrak{f}(x)\ dx = \sigma - \Phi$. Furthermore, the assumption that $\mathfrak{f}$ is ε-far from uniform is now equivalent to $\Phi \geq \varepsilon$.

Then, we have

$$\begin{aligned}
\int_0^1 x\ \mathfrak{f}(x)\ dx &= \int_0^\sigma x\ \mathfrak{f}(x)\ dx + \int_\sigma^1 x\ \mathfrak{f}(x)\ dx \\
&\geq \frac{1}{\sigma} \cdot \int_0^\sigma x\ dx \cdot \int_0^\sigma \mathfrak{f}(x)\ dx + \\
&\quad \frac{1}{1-\sigma} \cdot \int_\sigma^1 x\ dx \cdot \int_\sigma^1 \mathfrak{f}(x)\ dx \\
&= \frac{1}{\sigma} \cdot \frac{\sigma^2}{2} \cdot (\sigma - \Phi) + \\
&\quad \frac{1}{1-\sigma} \cdot \frac{(1+\sigma)(1-\sigma)}{2} \cdot (1 - \sigma + \Phi) \\
&= \frac{\sigma \cdot (\sigma - \Phi)}{2} + \frac{(1+\sigma) \cdot (1 - \sigma + \Phi)}{2} \\
&= \frac{1 + \Phi}{2} \geq \frac{1 + \varepsilon}{2} ,
\end{aligned}$$

where in the first inequality above we used Chebyshev integral inequality. ☐

While it is tempting to apply a similar approach to extend the proof of Theorem A.1 to larger n, we do not know how to extend it even for small values of $n > 1$. Instead, in general case, we use a different approach which works via comparing the values of $\int_{\boldsymbol{x} \in [0,1]^n} |f(\boldsymbol{x}) - 1| d\boldsymbol{x}$ with $\int_{\boldsymbol{x} \in [0,1]^n} \|\boldsymbol{x}\|_1 (f(\boldsymbol{x}) - 1) d\boldsymbol{x}$ for any monotone function $f : [0,1]^n \to \mathbb{R}$.